



Delivering Better Policies Through Behavioural Insights

NEW APPROACHES



Delivering Better Policies Through Behavioural Insights

NEW APPROACHES

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Please cite this publication as:

OECD (2019), *Delivering Better Policies Through Behavioural Insights: New Approaches*, OECD Publishing, Paris.

<https://doi.org/10.1787/6c9291e2-en>

ISBN 978-92-64-55573-0 (print)

ISBN 978-92-64-57741-1 (pdf)

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Photo credits: Cover © Hendrasu/Shutterstock.com.

Corrigenda to OECD publications may be found on line at: www.oecd.org/about/publishing/corrigenda.htm.

© OECD 2019

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgement of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.

Foreword

Behavioural insights (BI) has greatly expanded our understanding of how the psychological, social and cultural factors governing human behaviour affect policy outcomes. Governments increasingly recognise the value of BI: OECD research has mapped over 200 government units, initiatives and partnerships across the world applying BI to public policy. As BI has become more prevalent, countries are using it to address new and more complex policy problems.

The OECD has been at the forefront of documenting and researching the use of BI in public policy in general and in policy fields such as consumer protection, environment, financial education, firm behaviour, public sector integrity, obesity, regulatory policy and taxation. This report builds on this work by tackling new problems related to individual behaviour as well as organisational behaviour.

This work brings together four OECD Directorates – Environment; Financial and Enterprise Affairs; Public Governance; and Science, Technology and Innovation – to test behaviourally informed solutions to complex policy problems in the fields of energy consumption, competition, safety and consumer protection respectively. It uses a variety of methods, including literature reviews and modelling to expand the theoretical basis for applying BI to policy problems. Where feasible, these solutions were tested using various experimental methods. Research institutions and leading experts across the world also participated in the development of this report.

The report highlights promising results in applying BI to complex policy problems as well as areas for further research and action. For example, it discusses ways to improve individual behaviour with regard to energy consumption, as well as an experimental approach to examining online advertising, disclosure agreements and personalised pricing. In terms of changing the behaviour of organisations, the report looks at ways to deter cartels through antitrust regimes and foster a culture of safety in the energy sector.

The report also identifies some broader lessons for using BI, such as the importance of scoping the policy problem, studying ways to change organisational behaviour, encouraging cross-national and cross-cultural experimentation, using a variety of tools and experimental methods and translating results so that policymakers can implement the findings at scale. It is also important to include BI in every stage of the policy cycle to maximise the potential impacts to policy outcomes. Behavioural practitioners and policymakers must also adhere to ethical standards to ensure they are applying BI responsibly.

Acknowledgements

This report is the product of a joint effort across the Organisation for Economic Co-operation and Development (OECD), bringing together contributions from the OECD Directorates for Environment (ENV), Financial and Enterprise Affairs (DAF), Public Governance (GOV), and Science, Technology and Innovation (STI). This report and the work underlying the analysis would not have been possible without the support of the OECD Central Priority Fund managed by the Office of the Secretary-General.

The overall report was prepared by Francesca Papa, James Drummond and Jun Nakagawa with co-ordination and input from Filippo Cavassini and Anna Pietikainen. Work underlying the report was led by Filippo Cavassini and Faisal Naru and received the encouragement and support of Marcos Bonturi, Director, Irène Hors, Deputy Director, and Nick Malyshev, Head of the Regulatory Policy Division, Public Governance Directorate. Peer review comments to the overall report were provided by Martin Forst and Alessandro Bellantoni, Head and Deputy Head of the Governance Reviews and Partnership Division, and Edwin Lau, Head of the Budgeting and Public Expenditures Division, Public Governance Directorate. Input and comments on the overall report were provided by Walid Oueslati (ENV), Chris Pike (DAF) and Michael Donohue (STI). Eleonore Morena edited the report and Jennifer Stein co-ordinated the editorial process, with editorial support provided by Andrea Uhrhammer and Meral Gedik.

Each chapter was prepared and reviewed by dedicated teams whose work the Secretariat would like to acknowledge:

- Chapter 2: *Real-time information and consumer decisions on energy consumption* was prepared by Dr Nicholas Rivers, University of Ottawa, with input and support from Elisabetta Cornago, Alexandros Dimitropoulos and Walid Oueslati, Environment and Economy Integration Division, ENV. The work underlying the chapter was discussed in the OECD Environmental Policy Committee in February 2018.
- Chapter 3: *Protecting digital consumers* was prepared by Richard Bates, with input and support from Anna Barker and Michael Donohue, Division for Digital Economy Policy, STI. The work underlying the chapter was discussed in the OECD Committee on Consumer Policy in November 2018.
- Chapter 4: *Cartel deterrence and the labour market for managers* was prepared by Dr Ricardo Gonçalves, Católica Porto Business School, Dr Miguel A. Fonseca, University of Exeter Business School, Dr Joana Pinho, Católica Porto Business School, and Dr Giovanni Tabacco, Swansea University, with input and support from Sean Ennis and Chris Pike, Competition Division, DAF. The work underlying the chapter was circulated in the OECD Competition Committee in December 2018.

- Chapter 5: *Fostering a safety culture in the energy sector* was prepared by Mary MacLennan, behavioural science consultant and former member of the Impact and Innovation Unit, Canada, supported by Dr Tom Reader, London School of Economics, and Dr Morgan Tear, Research Fellow, London School of Economics and BehaviourWorks Australia at Monash University, with input and support from Filippo Cavassini, Shelly Hsieh and Anna Pietikainen, GOV. Comments were provided on the country analysis by Claudine Bradley, Melissa Mathieson and Peter Watson, National Energy Board (NEB), Canada; Ann McGarry and Ashleigh Shaheen, the Commission for Regulation of Utilities (CRU), Ireland; Daniela Martínez Ramírez, Patricia Angelica Quiles Martinez and Fernanda Garcia Rodriguez, Agency for Safety, Energy and Environment (*Agencia de Seguridad, Energía y Ambiente*, ASEA), Mexico; and Qais Al Zakwani, Authority for Electricity Regulation (AER), Oman. The work underlying the chapter was discussed in the OECD Network of Economic Regulators in November 2018.

Table of contents

Abbreviations and acronyms.....	11
Executive summary	13
Chapter 1. Overview and key lessons	17
Behavioural insights and public policy: Where are we now	18
Applying behavioural insights to current policy issues: New insights	20
Guidance and lessons for policymakers.....	24
References.....	26
Part I. Individual decision-making	29
Chapter 2. Real-time information and consumer decisions on energy consumption.....	31
Introduction.....	32
Context and problem setting	34
Literature review.....	36
Methodology	39
Results and discussion	41
Conclusion	44
Notes	44
References.....	46
Chapter 3. Protecting digital consumers.....	49
Introduction.....	50
Context and problem setting	52
Methodology	64
Conclusion	66
Notes	66
References.....	67
Part II. Organisational decision-making.....	71
Chapter 4. Cartel deterrence and the labour market for managers.....	73
Introduction.....	74
Context and problem-setting.....	77
Methodology.....	86
Results and discussion	95
Conclusion	110
Notes	111
References.....	112

Chapter 5. Fostering a safety culture in the energy sector	117
Introduction.....	118
Context and problem setting	119
Methodology.....	125
Results and discussion	129
Conclusions.....	139
References.....	140
Annex 5.A. Additional information and sample survey	145

Tables

Table 4.1. Summary of hypotheses	95
Table 4.2. Determinants of cartel formation coefficient and statistical significance (standard errors) .	99
Table 4.3. Determinants of individual and aggregate votes for cartel formation coefficient and statistical significance (standard errors)	100
Table 4.4. Estimates on the propensity of managers to co-ordinate on the same price when a cartel is active.....	102
Table 4.5. Estimates of the determinants of selling price.....	104
Table 4.6. Relative frequency (%) of self-reported motives for contract choice NoChat	106
Annex Table 5.A.1. Additional information on regulators by country	145
Annex Table 5.A.2. Canada National Energy Board	147
Annex Table 5.A.3. Mexico: Agency for Safety, Energy and Environment (ASEA).....	147
Annex Table 5.A.4. Ireland: Commission for Regulation of Utilities.....	148
Annex Table 5.A.5. Behavioural vignette items	153
Annex Table 5.A.6. Sample behavioural vignette.....	154

Figures

Figure 4.1. Cartel formation frequency conditional on treatment and contract type.....	98
Figure 4.2. Distribution of votes in favour of forming a cartel	100
Figure 4.3. Proportion of cases in which cartels were active and firms co-ordinated on prices.....	101
Figure 4.4. Average selling prices conditional on contract type and a cartel being active.....	103
Figure 4.5. Proportion of low-powered contracts offered	105
Figure 5.1. National differences in safety culture perception.....	131
Figure 5.2. Role differences in safety culture perception.....	132
Figure 5.3. Comparing behavioural principles	133
Figure 5.4. Messenger effects.....	134
Figure 5.5. Occupational response to the messenger effect	136
Figure 5.6. Regulator vs. entity responses to messenger vignettes	137

Boxes

Box 3.1. Selected OECD e-commerce recommendation principles on online advertising	53
Box 3.2. Examples of behavioural biases in consumer policy	55
Box 3.3. Do consumers read online terms and conditions?.....	60
Box 4.1. Treaty on the Functioning of the European Union, Section 101	77
Box 4.2. Example 1	87
Box 4.3. Example 2	87
Box 4.4. Example 3	90
Box 4.5. Example 4	92
Box 5.1. Safety culture survey	151

Follow OECD Publications on:



http://twitter.com/CECE_Pubs



<http://www.facebook.com/CECEPublications>



<http://www.linkedin.com/groups/CECE-Publications-4645871>



<http://www.youtube.com/oecdlibrary>



<http://www.oecd.org/oecdirect/>

Abbreviations and acronyms

ARC	Act Against Restrictions on Competition
AER	Authority for Electricity Regulation
ASEA	Agency for Safety, Energy and Environment (<i>Agencia de Seguridad, Energía y Ambiente</i>)
BI	Behavioural Insights
CCP	OECD Committee on Consumer Policy
CRU	Commission for Regulation of Utilities
DAF	OECD Directorate for Financial and Enterprise Affairs
DOJ	Department of Justice, United States
ECJ	European Court of Justice
EEA	European Economic Area
EDC	Electric distribution company
ENV	OECD Environment Directorate
EU	European Union
FCO	Federal Cartel Office
FTC	Federal Trade Commission, United States
GOV	OECD Public Governance Directorate
IHD	In-home displays
NEB	National Energy Board of Canada
OBA	Online Behavioural Advertising
OEB	Ontario Energy Board
OECD	Organisation for Economic Co-operation and Development
OFT	Office of Fair Trading, United Kingdom
PPE	Personal protective equipment
RCT	Randomised controlled trial
STI	OECD Directorate for Science, Technology and Innovation
T&Cs	Terms and conditions
TFEU	Treaty on the Functioning of the European Union
TOU	Time-of-use pricing
TRB	Transportation Research Board

Executive summary

Behavioural insights (BI) provides policymakers with a new set of tools for designing and implementing effective public policies. BI complements conventional economic theory with extensive evidence on how humans depart from rational decision-making and misperceive risk. Moreover, BI can offer guidance on tailoring policies to address these deviations, by designing interventions that provide individuals with the contextual information that they are otherwise lacking to make a rational decision. An increasing number of public bodies around the globe have recognised the value of applying BI to public policy.

As BI and evidence-based methodologies gain popularity with policymakers, the scope of their use is also expanding. While most applications of BI to date have focused on improving individual choices, new frontiers of applying BI include both complex individual behaviour problems and organisational behavioural change.

Furthermore, several open questions remain, such as whether behavioural interventions can generate lasting effects over time, the extent to which individuals deviate from traditional assumptions of rationality and, in particular, how different countries or groups (e.g. professional or social categories) respond to equivalent behavioural policies.

Applying BI to individual and organisational behaviour

This report contains four pieces of OECD research that broaden the application of BI to public policy and tackle new questions related to both individual behaviour and the behaviour of organisations. The work was developed by the OECD Directorates for Environment, Financial and Enterprise Affairs, Public Governance, and Science, Technology and Innovation. It encompasses a diverse set of countries and research topics in the policy areas of competition, consumer protection, energy consumption and safety.

In terms of individual behaviour, this report investigates ways that BI can be used to improve the effectiveness of smart meters and induce energy savings, through an experiment in Canada. Results show that the provision of real-time feedback on electricity consumption through in-home displays causes households to reduce consumption by about 3%. The effect lasts for at least five months and is mainly the result of one-time actions, such as adjusting devices' settings to less energy consuming modes or investing in more energy efficient appliances.

This report also examines new ways BI can be applied to consumer policy, continuing the long history of applying BI to the protection of consumers. The report explains how BI can help understand and address the impact of online advertising on consumers. It also proposes practical next steps for policymakers to improve consumer understanding of online disclosures. It further explores how experimental approaches can help tailor disclosures to consumers regarding personalised pricing.

In terms of changing the behaviour of organisations, this report examines the impact of different antitrust regimes on deterring cartel behaviour. The results, which were produced from a theoretical model and tested empirically through a laboratory experiment, shed light on how competing regulatory frameworks may affect managers' incentives to collude and, in turn, managers' labour markets. Moreover, the research highlights how differences in manager-specific attributes, such as risk aversion and strategy choice, can determine the rise or deterrence of cartels. In doing so, the study also discusses how BI can serve as a tool to explain theory-evidence gaps that arise from empirical observations.

Finally, the report investigates how regulatory policymakers can foster a culture of safety in the energy sector. The results of an online experiment with regulators and regulated entities in Canada, Ireland, Mexico and Oman show the potential importance of messengers and feedback in improving safety but were inconclusive for the effect of social norms on changing behaviours around safety. However, these effects vary considerably amongst the application of each behavioural insight, and also by country and type of respondent.

Key lessons: How can policymakers apply BI to complex problems

Together, the four policy areas examined in the report provide a set of lessons for using BI to deliver better policies and offer pathways for using the tool to contribute meaningfully to solving complex policy problems. These lessons are described below:

- **Investing time and resources in scoping policy problems** to understand whether a behavioural intervention is required and how it can then improve outcomes. This should be the fundamental point of departure for evaluating the use of BI in public policy, and more broadly, getting the solutions right.
- **Further research is needed on the effectiveness of BI in changing the behaviour of organisations**, as many complex policy problems involve addressing the actions of organisations. This can be of interest to businesses and regulated entities, as well as within government bodies to improve the effectiveness of policy delivery.
- **Cross-national and cross-cultural experimentation** should be encouraged as a way to study various behavioural problems in different contexts and assess the potential benefits of behaviourally informed policies. International differences in culture and context also highlight the value-added of evidence-based policies for understanding “what really works” in targeted environments.
- **There is a wide range of robust and cost-effective tools** available for behavioural policymaking. Randomised controlled trials (RCTs), quasi-experiments, theoretical models and laboratory research should be fully exploited by practitioners and policymakers. It is crucial to be rigorous when designing behavioural interventions while capitalising on available resources and opportunities.
- **Findings must be scalable and translated into a language that is practical and relevant to policymakers**; otherwise, BI studies may be little more than exploratory experiences. Meaningfully interpreting and sharing results, even when null or statistically limited, can produce important lessons on how to design future studies and public policies.

- **Equally important is to pay special attention to ethical considerations at all stages of the research and policy-making process**, especially regarding experimentation and testing. Behavioural practitioners and policymakers can rely on available ethical guidelines to ensure they are applying BI responsibly.

Chapter 1. Overview and key lessons

This chapter provides an overview and key lessons of how behavioural insights (BI) is being used to improve public policy and the new insights gained from the application of BI to complex policy problems in the fields of energy consumption, competition, safety and consumer protection. The chapter concludes with guidance for policymakers to consider when applying BI to public policy.

Behavioural insights and public policy: Where are we now

The field of behavioural insights (BI) is based on the idea that context and cognitive abilities shape our decisions. It acknowledges that human behaviour is shaped by biases in decision-making and that our environment and available information can influence our ability to act “rationally” in systematic ways. On a personal level, everyone may be familiar with these ideas from daily experiences, such as being overconfident about meeting deadlines, remembering a doctor’s appointment only thanks to a reminder or picking up an unhealthy snack because it is at eye-level and easy to see. However, in the policymaking context, these notions are not often explicitly addressed. Rather, models that are used to build policy often assume individuals are “rational” enough to avoid these biases.

BI helps policymakers by providing them with a clear methodology that generates evidence on how people “actually” behave and enhances the analysis, design and delivery of public policies. The OECD (2017) provided a formal definition of the field of BI as lessons derived from behavioural and social sciences, including decision-making, psychology, cognitive science, neuroscience, organisational and group behaviour. The key feature of BI is an inductive approach to the design and delivery of policies, that is driven by experimentation and piloting, seeking to understand the actual behaviour of the beneficiaries of policies and testing possible solutions before implementation. BI enables policymakers to develop innovative approaches to designing and implementing policies, while not substituting their role or competency in making decisions.

Likewise, through experimentation and trialling, BI offers a cost-effective way of testing multiple policy responses at once and on a smaller scale to determine the best course of action (Benartzi et al., 2017). This approach limits the risk of committing resources to the full implementation of a given policy solution, which may have to be revisited at a later date. Evidence shows that this approach is having a real impact by providing countries with the resources necessary to learn, iterate and implement innovative policies.

Applying behavioural insights in policy settings

A key aspect of BI is that it involves isolating a specific behaviour that can affect a policy outcome. Isolating this behaviour and breaking down a problem into smaller manageable and impactful ones, policymakers can get to the “crux of the matter” and enhance the impact of interventions. For example, when trying to tackle a complex problem such as encouraging a company to pollute less, BI could be used to tackle a specific action – such as reporting on pollution leaks. The default tends to be that companies have to report pollution leaks only when a problem occurs. Changing the default to reporting every week – even when there is not a problem – is one way that BI can be leveraged to encourage more consistent reporting of pollution leaks, and thus the overall policy aim of reducing overall pollution.

Once a certain behaviour is prioritised as the focus of the BI intervention, it is then important to think about the barriers and enablers of this behaviour. This is often known as behavioural mapping and is the crucial step for identifying specific behavioural levers that can be harnessed to achieve behaviour change and, in turn, the desired policy outcome.

How is BI being used to develop and implement policies?

Governments and organisations are increasingly using BI as a tool to design and deliver better policies and services. This has resulted in a wide application of BI across the globe and has fostered a culture of experimentation to better understand how individuals actually behave. This is reflected in the OECD report on BI case studies, which contains over 150 examples in 11 policy sectors (OECD, 2017). The growing demand of BI has led to governments establishing their own BI units, consultancies providing BI-informed strategies and/or running randomised controlled trials (RCTs) and groups within academic institutions lending their behavioural science expertise to governments.

Thinking about behaviour and acknowledging biases can seem intuitive but the application of BI has proved that it is often not the case. The field of BI is based on decades of academic research across a number of disciplines and often benefits from a nuanced understanding of the research in order to apply it in the right way and in the right context.

Additionally, key elements of the application of BI are the use of rigorous research methods and the generation of evidence-based policies. Evidence-based insights are required because BI is fundamentally based on actual rather than expected “rational” behaviour. To effectively complete this work in a policy setting – whether inside government or when governments work with external partners to test policy solutions – a commitment to and appreciation for a scientific approach to policymaking are key. When it comes to knowledge about academic disciplines and rigorous methods, knowledge sharing between policymakers and academics can play an important role (Lunn, 2014). Academics may provide technical expertise when conducting BI experiments and applications, including helping governments implement rigorous experimental approaches, from randomised controlled trials (RCTs) to quasi-experiments.

The need for rigorous experimentation processes to inform BI has been all the more emphasised in light of questions related to the extent to which experimental results can be replicated and scaled up (Yong, 2018; Schooler 2014; Munafò et al., 2018). As the academic community debates that the findings of a number of scientific studies, including foundational work on judgement and decision-making, are hard or impossible to replicate in subsequent investigations, policymakers need to be aware of this debate and ensure that experiments are conducted with the necessary rigour to ensure robust results.

While methodological debates still exist, BI principles continue to be applied successfully to a wide variety of policy domains, ranging from energy and environmental behaviour (Goldstein et al., 2008; Lehner et al., 2016) to financial practices (Thaler, 2005; Schoar, 2014), health (Milkman et al., 2011) and, more recently, topics on development economics (World Bank, 2015; Kremer et al., 2018). Nascent streams of literature are also exploring domains that were traditionally outside the scope of BI, such as organisational behaviour, digital transformation and macroeconomics (De Grawe, 2012).

Ethical considerations, especially around experimentation and testing, also need to be taken into consideration. For behavioural practitioners and policymakers, this means paying special attention to ethical considerations to ensure they are applying BI responsibly. To help address these concerns, practitioners may consider using academic partners who work in institutions with established codes of ethics and the use of a broader ethical framework already in place in public bodies. The behavioural community has also identified the need to establish a code of ethics for behavioural practitioners that promotes the responsible application of behavioural tools and ensure those working in the field

adhere to certain standards when designing and running experiments in a public sector context and reporting on experiments by governments (OECD, 2018). To this end, the OECD has developed a toolkit and ethical framework that gives policymakers a step-by-step process for analysing a policy problem, building strategies and developing behaviourally-informed interventions with a set of ethical guidelines for each step of the process (OECD, forthcoming).

As BI continues to evolve, there are several promising areas in which further collaboration and efforts could benefit the field. This report aims to address a number of them by discussing evidence from recent OECD work across four policy sectors: electricity consumption; digital consumer protection; cartel deterrence; and safety culture. Respectively, the studies are the result of the effort of four OECD directorates: Environment (ENV), Financial and Enterprise Affairs (DAF), Public Governance (GOV) and Science, Technology and Innovation (STI).

The chapter is structured as follows: first, it presents an overview of the four recent OECD studies, divided in accordance to thematic order, starting with two studies on individual decision-making and continuing with two studies on organisational behaviour. Then the chapter aims to draw the overall lessons and guidance from the studies and presents avenues for future applications of BI. Finally, an overview of the context, key findings, general lessons and policy implications of each project is presented in the chapters that follow.

Applying behavioural insights to current policy issues: New insights

Individual decision-making

Part I of this report first looks at decision-making of individuals, specifically consumers, in electricity consumption and online consumer engagement. As technology continues to develop in both electricity and e-commerce markets, it is increasingly valuable for governments to understand how electricity consumers actually behave and interact with changes in their physical and digital environment. Part I contributes to the research and implementation of energy and consumer protection policies by challenging assumptions on human behaviour and related decision models (i.e. utility maximisation and consistent preferences).

The first two studies demonstrate the advantages of BI when used with traditional policy approaches that focus on technical changes such as energy-efficient technologies and online price mechanisms. In the first study, which looks at the impact of smart meters on energy consumption, researchers implement a field experiment to test traditional theories of optimal consumption decisions. In the second study, researchers first analyse the behavioural biases of online consumer protection and then develop policy recommendations and potential future experiments.

Both studies illustrate how policymakers can benefit from BI, especially through experimentation and behavioural literature to explain discrepancies between theory and practice.

Smart meters and electricity consumption

Chapter 2 examines the increasing popularity of smart meters. In the past decade, governments have rolled out smart meters to replace traditional analogue meters in many regions of the world. In contrast to analogue meters, smart meters track real-time energy

use and automatically send data to energy suppliers. When coupled with feedback technologies such as in-home displays (IHDs), consumers can easily access real-time electricity consumption and time-use pricing inside their home. Smart meters linked to IHDs offer the promise of reducing energy usage by giving consumers control over their usage levels and supporting time-of-use tariffs that can help to spread the demand for electricity more evenly throughout the day.

Although the features of smart meters seem promising for energy conservation, the empirical literature reveals mixed results ranging from no change to a 17% reduction in energy consumption. There have been a considerable number of studies measuring the impact of smart meters, but the earlier studies lacked rigour. More recent studies using (quasi-)experimental approaches with meaningful sample sizes have found that results are context-specific with limited external validity. There remains a lack of clarity on how real-time feedback affects electricity consumption.

To contribute to the gap in the literature, the study shares findings from a robust quasi-experimental field experiment in Ontario, Canada. With an exemplary sample size of 7 000 households, researchers found that during the two-year period of the study, real-time feedback led to an overall 3% decrease in consumption but no significant shifts in consumption patterns. In addition, the study shows adjustments were made as one-time decisions rather than on a continuous basis as reactions to time-of-use prices. These results challenged their theoretical model, which was based on a “rational” consumer who made more optimal energy consumption decisions when given higher quality and quantity of information.

These findings reflect the advantages of experimentation and need for further long-term monitoring. Although in theory, smart meters should encourage optimal decision-making for consumers, the field experiment shows the value of testing assumptions. Only through experimentation did researchers learn that consumption patterns of Ontario households deviate from optimal electricity use when receiving real-time feedback. Further, although researchers found a reduction in electricity usage, they question the sustainability after five months. More longitudinal field experiments are necessary to understand the long-term effects of real-time feedback on energy use.

Digital consumer policy

Chapter 3 analyses the challenges of online consumer protection through a behavioural lens. In this digital age, companies can provide more timely and relevant content for online consumers and, at the same time, easily mislead or misinform when consumers purchase products or agree to terms on line. This chapter uses BI as an analytical tool and provides practical implications for consumer authorities in three areas: online advertising, disclosure agreements and personalised pricing.

The main contribution of this study is that it analyses each policy problem through a behavioural angle from the start. Through this approach, the study leverages relevant scientific relevant literature to identify potential biases that could apply to the context of the specific policy question under investigation. For online advertising and disclosure agreements, the chapter points to behavioural biases such as anchoring and defaults that put online consumers at risk.

Subsequently, the chapter provides practical next steps for policymakers. For online advertising, they recommend potential behavioural experiments to understand the most effective ways to protect consumers from online advertising that may leverage

behavioural biases. For disclosure agreements, in addition to experimentation, the chapter provides a list of behaviourally-informed recommendations that prioritise simple, clear and timely ways to improve consumer understanding of online disclosures.

Finally, the chapter explores new territory by analysing the behavioural biases and potential implications of personalised pricing. Although there is still limited empirical evidence, there is growing interest in how online vendors tailor prices to individuals based on personal data. After exploring potential biases relevant to personalised pricing such as framing and overconfidence, the chapter poses two hypotheses to test these biases through e-commerce simulations for a future lab experiment. This behaviourally-driven angle on problem scoping and emphasis on experimentation is a model approach to understand how best to design consumer protection policies for the online market.

Organisational decision-making

Part II examines the application of BI to organisational behaviour, namely in organisational safety culture and cartel deterrence. These applications respond to the widespread perception that BI can and should go beyond the study of individual-level decision processes for higher impact. As first-of-a-kind studies, these chapters seek to fill unanswered gaps in the literature about applying BI to these domains.

In addition to the novelty of their domains, both applications present a uniquely international perspective, with the safety culture study including respondents from Canada, Ireland, Mexico and Oman, and the cartel deterrence study conducting a comparison between North American and European systems of anti-trust and competition law. Within a wider framework of growing interconnectedness of individuals and organisations, understanding how different cultures think and operate could be extremely fruitful for the field of BI and policymaking more broadly.

A further common attribute of the two applications is a close examination of how different actors within organisations behave and interact. For the safety culture research, this translates into a comparison of how different occupational roles – from frontline workers to managers and senior managers – perceive safety culture in their organisation; as well as a comparison of how entities in different hierarchical positions – regulators vs. regulated entities – perceive the safety culture in their field. Similarly, the application to cartel deterrence analyses the experimental behaviour of both firm owners and managers in the same organisations, thus allowing us to observe the competing and interrelated incentives of different moving parts of a company.

Cartel behaviour

Chapter 4 seeks to compare two alternative models of cartel deterrence through a BI lens. On the one hand, a number of antitrust regimes punish detected cartels through fines which are revenue-based and levied on the firm level (i.e. corporate fines). This model has been prevalent in many European countries and is reflected in the formulation of the antitrust regulations of the European Union. On the other hand, North American jurisdictions, and notably the United States system, punish cartels by means of salary-based fines that function at the level of companies' managers (i.e. individual fines). Comparing the two regimes helps examine whether and how manager incentives differ under competing regulatory frameworks. In addition, the comparison sheds light on how different antitrust regimes impact the labour market for managers, in terms of the type and level of contracts offered by firm owners and shareholders.

In particular, the study addresses the two questions of: i) whether antitrust jurisdictions involving individual-level fines have a stronger deterrence effect than the antitrust jurisdictions based on corporate-level fines; and ii) whether different regulatory frameworks generate distortions in the labour market for managers, by making certain categories of contracts more appealing to shareholders (e.g. contracts with fixed salary versus fixed plus a variable component).

The chapter relies on a unique combination of both theoretical and experimental approaches. Interestingly, the experimental findings deviate in significant ways from the theoretical predictions. While the theoretical model predicts that the type of fines on collusion should impact managers' decisions on whether to form a cartel, in observed experimental behaviour, there is no statistical difference in the frequency (or duration) of cartels between the EU and US antitrust frameworks. Experimental results not only depart from the theoretical model with regard to cartel prevalence but also with regard to pricing behaviour and contractual choices by shareholders.

One of the key contributions of the study is precisely to indicate a set of potential behavioural explanations for why we observe a discrepancy between the way subjects actually behave in the lab and the “as if” theoretical model, in which actors are assumed to have somewhat homogeneous preferences. Conversely, the heterogeneity in manager-specific parameters, such as risk aversion and strategies chosen, can help understand the theory-evidence gap, especially in a context in which the decision of even one single company's manager can prevent the formation of a cartel. While these results leave scope for further discussion, they markedly signal one of the key lessons of this report, which is the value-added of incorporating BI as an *ex ante* rather than *ex post* tool.

Safety behaviour

Chapter 5 provides experimental evidence on the potential application of different behavioural principles (social norms, messenger effect and feedback/benchmarking) to strengthen different dimensions of safety culture. The study involved the participation of both regulators and regulated entities in high-risk industries and explored their perception regarding awareness of safety culture in their field as well as potential responsiveness of different organisational actors to the selected behavioural principles.

The research introduced a new paradigm in the study of safety culture by leveraging a set of behavioural scenarios and vignettes. The vignettes simulated the application of BI to common safety topics, such as the introduction of a new safety regulation and reports on lost-time injury rates. The rationale behind the selection of these scenarios was that they would be generalisable across the four participating countries – Canada, Ireland, Mexico and Oman – and across their sectors, which ranged from oil and gas to electricity regulation.

The findings indicated interesting asymmetries between how regulators and regulated entities perceived safety culture, which suggests that it is essential for regulators to take into account different views around safety when designing new policies. Equally, the results displayed asymmetries between managers and frontline workers, with the latter reporting a more negative perception of safety culture (Tear et al., forthcoming; Parand et al., 2010). This result is possibly due to frontline workers' more tangible contact with the safety risks in the organisation and it corroborates the importance of factoring in perception and information differences to avoid unintended policy effects and to ensure policies are targeted for different audiences.

With regard to the relative effectiveness of different behavioural principles, the study suggests that feedback has a more powerful effect on safety culture than social norms and messenger effects. Interestingly, in the experiment, this is true regardless of the modality through which feedback is provided. Future policies can incorporate this finding by ensuring that workers are provided with some form of data-driven performance feedback that is aimed at correcting and preventing unsafe practices. In particular, given asymmetries between workers and managers, it would be useful to ensure that the feedback is empirical/data-driven rather than perception-based. While messenger effects and social norming were overall less effective than feedback, they still indicated a number of behavioural differences, notably when looking at the data through an international lens. In particular, social norms seemed to have their most powerful effects on the Mexican sample, while messenger effects seemed to mostly only matter in Ireland and Oman. These international differences are further discussed in the relative chapter and represent one of the main value-added of the study.

Overall, the research demonstrates that not all behavioural principles are equally effective in motivating safe behaviour and that different nationalities can report different perceptions of safety culture as well as different levels of responsiveness to behavioural principles. The guidance provided by the research is, therefore, a promising tool to further explore the framework of the multiple “unknown knowns” and “known unknowns” which characterises safety risks.

Guidance and lessons for policymakers

The four applications present distinct methodologies and findings that are of value for the specific policy areas they address. However, they also contribute to identifying common principles that can be taken into consideration when applying BI to public policy. Overall, the report:

- **Points at new policy tools for behaviour change**, as with the adoption of smart meters and new digital technologies.
- **Indicates new insights for more effective policies** such as factoring in how different actors in high-risk industries view safety.
- **Expands the sets of outcomes that can potentially be reached through behavioural public policy**, by exploring how BI could be applied to novel domains, including the study of cartels deterrence as well as digital consumer policy.

The applications made an effort to **embed BI from the start**. In particular, those related to consumer protection and safety culture invested time to analyse the policy problem from a behavioural perspective and use behavioural principles to inform their experimental designs.

Moreover, the applications **push the frontier of BI beyond the usual applications to individuals and explore the domain of organisational behaviour**. Specifically, the report investigates how policymakers can foster components of a strong safety culture with regulators and regulated entities in high-risk industries, and then examines strategies for cartel deterrence by analysing the experimental behaviour of firm owners and managers within the same organisation.

The applications stress the importance of **monitoring long-term effects**, as exemplified in the study on the impact of smart meters on electricity consumptions from a two-year quasi-experimental study in Canada. They also apply **BI in different contexts**, as the report investigates differences in cartel behaviour between North American and European systems, and compares safety culture among Canadian, Irish, Mexican and Omani regulators and regulated entities.

Furthermore, the study on safety culture is exemplary for its investment of time and resources in **scoping the most relevant policy problems**. Not only did the researchers hold multiple formal and informal discussions with national regulators and focal points to identify the most urgent themes around safety but they also distributed a survey on the most important behavioural assets and needs in relevant entities before running the experimentation. This process ensured that the questions asked in the research were of direct interests for the parties involved in both the regulation and the practice of safety, thus allowing for future scaling up of the study's recommendations. Even more importantly, the process ensured that a significant amount of time was spent identifying the most significant behavioural barriers and levers to be tested in the study.

Finally, with each chapter involving a different research methodology, the report as a whole illustrates the **variety of tools and areas of application available for BI**. This includes literature reviews and modelling to expand the theoretical basis for applying BI to policy problems and, where feasible, testing these solutions using various experimental methods, from laboratory experiments (Chapter 4) to large-scale quasi-random experiments (Chapter 1).

Insights from the four applications point to general lessons for pushing the frontier of BI applications to policymaking:

- **Embed BI throughout the policy cycle** as a part of the *ex ante* evaluation and *ex post* review. BI has mostly been applied at the late-design and implementation phase of the policy cycle, mostly to fine-tune and improve implementation or compliance when a policy is already in place (OECD, 2018). Applying BI from the start can help better define the problem and identify behavioural barriers that can potentially undermine the effectiveness of the policy.
- **Consider the behaviour of public and private organisations and not just individuals**. Most BI interventions have focused on individuals as citizens or consumers but less frequently as employees. Organisations are made of individuals and there are transferrable individual-level insights that can be applied to influencing organisational behaviour (OECD, 2018). As policy-relevant decisions are often made by organisations, applying BI to organisations may have widespread policy implications.
- **Investigate long-term effects of BI interventions**. In the early days of BI, the focus was mostly translating promising evidence-based interventions from the behavioural science literature to inform BI interventions that would provide proof of concept or “quick wins” (Sanders, 2018). As relatively little is known about long-term effects of BI interventions, the next phase as suggested by the applications presented in this report would be to prioritise pursuing projects that can be monitored and provide benefits in the longer term.
- **Explore the effectiveness of BI interventions and more broadly policy interventions in different contexts**. BI interventions implemented in different national and subnational contexts may establish which behavioural biases are

common to which societies and what factors moderate or mediate these effects. Applying BI can also serve as a powerful tool to test what works and what does not in different contexts and facilitate bespoke approaches. Additionally, understanding the extent to which BI and comparative methodologies apply to contingent cultural contexts may help advance the reflection on the ethical and distributive impacts of behavioural interventions and nudges. As more countries are integrating BI to policy design and delivery, there may be more opportunities to replicate similar studies and gain a more global understanding of behavioural biases to inform public policy.

- **Invest time and resources in scoping the policy problem.** This is a crucial but often overlooked step of applying BI, which involves understanding a policy problem before planning an intervention. This stage serves to “identify, define, evaluate and select those behavioural problems contained within a wider policy challenge that are particularly suitable for a BI approach” (OECD, 2018).
- **A plurality of robust and cost-effective methods is available for behavioural policymaking.** Appealing to different methodologies, such as randomised controlled trials (RCTs), quasi-experiments, theoretical models and laboratory research, can allow researchers to harness the complementary strengths of these approaches. In addition, the strategic use of multiple approaches to address one question – a process often defined as triangulation – might be a direct way of addressing replicability concerns (Munafò et al., 2018). When applying these tools, it is crucial for researchers to rigorously follow each step in the BI methodology to ensure robust results that can be scaled up to behaviourally informed public policies.
- **From research to policy.** In order to maximise the potential for a BI intervention to be scaled up into a policy strategy, studies should aim for full applicability of experimental results. Dissemination of results should also be “behaviourally informed” and aim for a level of clarity that will be easily accessible and understandable to all relevant stakeholders and policymakers. Importantly, unlike the world of academia, where the incentives are high to disseminate results that are statistically significant, even null or ambiguous statistical results in the world of policymaking can be essential for informing future policy interventions.
- **Always keeping ethics in mind.** Equally important is to pay special attention to ethical considerations at all stages of the research and policy-making process, especially regarding experimentation and testing. Behavioural practitioners and policymakers can rely on available ethical guidelines to ensure they are applying BI responsibly. OECD (forthcoming) serves as one source for these guidelines to ensure BI is always being applied responsibly.

References

- Benartzi, S. et al. (2017), “Should governments invest more in nudging?”, *Psychological Science*, Vol. 28(8), pp. 1041-1055, <https://doi.org/10.1177/0956797617702501>.
- De Grauwe, P. (2012), *Lectures on Behavioral Macroeconomics*, Princeton University Press.
- Goldstein, N., R. Cialdini and V. Griskevicius (2008), “A room with a viewpoint: Using social norms to motivate environmental conservation in hotels”, *Journal of Consumer Research*, Vol. 35, <https://doi.org/10.1086/586910>.

- Kremer M., G. Rao and F. Schilbach (2018), “Behavioral development economics”, in D. Bernheim, S. DellaVigna and D. Laibson (eds.) (forthcoming), *The Handbook of Behavioral Economics*.
- Lehner, M., O. Mont and E. Heiskanen (2016), “Nudging – A promising tool for sustainable consumption behaviour?”, *Journal of Cleaner Production*, Vol. 134(A), pp. 166-177.
- Lunn, P. (2014), *Regulatory Policy and Behavioural Economics*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264207851-en>.
- Milkman, K. et al. (2011), “Using implementation intentions prompts to enhance influenza vaccination rates”, *Proceedings of the National Academy of Sciences*, Vol.108(26), pp. 10415-10420.
- Munafò, M., R. Smith and G. Davey (2018), “Robust research needs many lines of evidence”, *Nature*, Vol. 553(7689), pp. 399–401, <https://doi.org/10.1038/d41586-018-01023-3>.
- OECD (2018), *OECD Regulatory Policy Outlook 2018*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264303072-en>.
- OECD (2017), *Behavioural Insights and Public Policy: Lessons from Around the World*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264270480-en>.
- OECD (forthcoming), *The Behavioural Insights Policymaking Toolkit and Ethical Framework*, OECD Publishing, Paris.
- Parand, A. et al. (2010), “The disparity of frontline clinical staff and managers’ perceptions of a quality and patient safety initiative”, *Journal of Evaluation in Clinical Practice*, Vol. 17(6), pp. 1184-1190, <https://doi.org/10.1111/j.1365-2753.2010.01506.x>.
- Sanders, M., V. Snijders and M. Hallsworth (2018), “Behavioural science and policy: Where are we now and where are we going?”, *Behavioural Public Policy*, Vol. 2(2), pp. 144-167.
- Schoar, A. and P. Tania (2014), *A Behavioral Approach to Financial Coaching, The Financial Health Check*, New American Foundation.
- Schooler, J.W. (2014), “Metascience could rescue the ‘replication crisis’”, *Nature*, Vol. 515(7525), p. 9.
- Tear, M.J. et al. (forthcoming), “Safety culture and power: Interactions between perceptions of safety culture, organisational hierarchy, and national culture”, *Safety Science*, <https://doi.org/10.1016/j.ssci.2018.10.014>.
- Thaler, R. (2005), *Advances in Behavioral Finance, Volume 2*, Russel Sage Foundation, New York.
- World Bank Group (2015), *World Development Report 2015: Mind, Society and Behavior*.

Part I. Individual decision-making

Chapter 2. Real-time information and consumer decisions on energy consumption

This chapter investigates ways that behavioural insights (BI) can be used to improve the effectiveness of smart meters and induce energy savings. It describes the results of an experiment in Ontario, Canada, that tests the provision of real-time feedback on energy consumption through in-home displays.

Introduction

New technologies are quickly developing in the electricity market, encouraging consumers to make well-informed decisions about their electricity consumption.¹ Unlike analogue electricity meters that typically only allow for flat electricity tariffs and information on aggregate usage, the digital (“smart”) meters provide consumers with information on dynamic electricity pricing and consumption in real time. Smart meters are replacing analogue meters in many regions of the world. As discussed in Rivers (2018), in Canada and the United States, approximately half of all residential meters have been replaced by smart meters as of 2016. In Italy, smart meters have been introduced for the totality of residential accounts (about 26 million). In France and the United Kingdom, the rollout of smart meters to households lags behind the rollout in North America. In Ontario, the focus of part of this report, the rollout of smart meters to residential customers was completed by 2010, making it an interesting case study for understanding the potential impacts of smart meters and associated feedback technologies on consumer electricity demand.

Paired with in-home feedback technologies, smart meters have two distinguishing features that can impact the environment through both behavioural and market-based avenues.² First, unlike analogue meters, digital meters record electricity consumption at a fine-grained interval, potentially enabling households to be exposed to prices that vary over time of day. On the other hand, with a standard electricity meter, consumers only find out their consumption when electricity bills arrive at monthly or bi-monthly intervals. Second, smart meters can communicate electricity prices and electricity consumption in real time to households, which provides them with a better informational basis on which to make electricity consumption decisions.

These are *behavioural* avenues through which smart meters and feedback technologies can change consumer behaviour. In addition, smart meters enable time-varying electricity pricing. Economists have long advocated for time-varying wholesale prices to be passed on to consumers, arguing that the flat tariffs normally used in the residential sector suppress potentially cost-effective demand response (Borenstein et al., 2002). Because they are digital devices, smart meters can facilitate the implementation of virtually any type of tariff structure, including those that vary over time. In contrast, with a standard electricity meter, implementing time-varying rates is difficult or impossible. Smart meters enable consumers to conserve electricity when supply is constrained by facilitating dynamic electricity pricing. This is a *market-based* avenue through which smart meters affect the market by influencing electricity demand.

This report, which draws on the findings in Rivers (2018) and where further detail can be found, focuses on a particular type of real-time feedback technology, in-home displays (IHDs), which provide consumers with real-time information about electricity consumption, price and expenditures. Such high-quality information should make electricity consumption more salient to households and therefore increase the ability of the consumer to optimise decisions relating to electricity consumption. However, it is unclear how consumers will respond to the installation of IHDs. In particular, optimising consumers may respond by either increasing or decreasing electricity demand, depending on the nature of their perceptions of electricity consumption and price before the installation of the IHDs. Likewise, IHDs may make consumers more or less responsive to changes in electricity price, depending on how consumers’ pre-IHD beliefs reflected actual electricity prices. Moreover, the installation of an IHD may also increase the attention that consumers devote to their electricity consumption and cause changes in

consumption as a result. Understanding how consumers respond to more information, therefore, rests more on empirical than on theoretical results.

The empirical literature on the impact of real-time feedback via IHDs on electricity demand has produced mixed results. Early pilot programmes developed by electric utilities typically suggest that providing households with real-time feedback on electricity demand causes a substantial reduction in electricity consumption. However, these early studies often do not use methods that would be considered appropriate today or they do not report enough information on methods, leading to doubts about their findings. More recent studies use high-resolution (e.g. hourly) data to compare electricity consumption from households with and without in-home displays, using either quasi-experimental or experimental research designs. These studies suggest that IHDs can induce meaningful reductions in electricity consumption in contexts where the price for electricity is high. However, there are few such high-quality studies and most of those that have been conducted focus on particular contexts such that results may not necessarily generalise to a wider population.

This report also provides a review of a recent study that sheds new light on the effect of real-time IHD feedback on consumer electricity demand. The study evaluates a programme that resulted in approximately 7 000 households in Ontario, Canada, being provided with an in-home electricity display. It uses a quasi-experimental approach to assess the impacts of real-time IHD feedback on household electricity demand, by leveraging the fact that IHDs are rolled out to households over a one-year period. This context enables a longitudinal approach to estimating the impact of IHD feedback, in which household electricity consumption with an IHD is compared to consumption in the same household before receipt of an IHD, controlling for trends experienced by other households whose IHD status does not change.

Based on this approach, several important findings are reported. First, the receipt of an IHD results in a reduction in electricity demand of around 3% overall. This result suggests that either: i) households underestimated their expenditures on electricity prior to receiving an IHD and the additional information caused them to reduce consumption; and/or ii) the receipt of the IHD caused electricity consumption to become more “visible” to households and led them to conserve electricity independently of the response to improvements in the quality of information. The results also suggest that household electricity conservation in response to real-time feedback provided via IHDs is concentrated in the autumn and winter heating seasons. The response by households is roughly uniform throughout the day and does not appear to be caused by the time-of-use pricing schedule.

The study found that household electricity conservation in response to IHD feedback persists for at least five months following the receipt of the display. Although it is not possible to confidently identify the mechanisms by which households respond to the IHD with the data available in this study, this finding suggests that households respond to real-time feedback in part by adjusting thermostat settings downwards or investing in durable energy efficiency improvements that result in lower space heating demand.

The chapter will be structured as follows: first, a background is provided on smart meters, including how they provide feedback and allow for time-varying prices. This is followed by a brief overview of findings from previous studies that measured the impact of real-time feedback on residential electricity demand. Next, the report summarises the theoretical model underlying the experiment, followed by a description of the quasi-experimental design of a case study in which real-time feedback was provided to

residential customers in an electricity distribution area in Ontario, Canada. Results of the case study are then discussed, including the impact of real-time feedback on electricity consumption and how it varies by season, time of day and outdoor temperature. The chapter is then concluded with a summary of the findings.

Context and problem setting

Real-time feedback and time-use electricity pricing

Smart meters differ in two important ways from traditional analogue electricity meters. First, they record electricity consumption using a digital, rather than analogue, technology. Electricity consumption on smart meters is also recorded with a corresponding timestamp, indicating the time of use with hourly or higher frequency. On an analogue meter, in contrast, it is not possible to know when electricity was consumed within a billing period. This difference between the two technologies entails that smart meters enable flexible pricing (i.e. varying within day and across days) while analogue meters do not provide this opportunity. Second, smart meter infrastructure allows communication between the meter and the electricity distribution company. This eliminates the requirement for manual in-place meter reading that is associated with analogue meters. Most smart meters additionally allow communication between the smart meter and the household.

These two differences between smart and conventional meters – regular recording of electricity consumption and communication ability – allow for important changes both in the way that electricity consumption is communicated to households and in the way that electricity consumption is billed. The following sections discuss each of these potential changes. It is worth noting that when smart meters are adopted, households and their electricity distributors can make choices about using these features of smart meters or not. Upon adopting smart meters, certain jurisdictions and households have chosen not to change the way electricity is priced or to make use of feedback on household electricity consumption

Smart meters allow real-time feedback on household electricity consumption

Like an analogue meter, a smart meter is installed outside the house and does not typically display information on electricity consumption in an accessible, intuitive or easy-to-read manner for the average household. On its own, a smart meter provides limited information to a household about electricity consumption. However, most smart meters include features to allow communication between the electricity meter and the household, typically using wireless technology. Using these features, or using near-real-time data relayed by the smart meter to the electricity distribution company, households can obtain feedback on their electricity consumption.³ The provision of this information may encourage households to change electricity consumption behaviour, possibly inducing energy conservation. There are a number of technologies that have been adopted to provide households with real-time information on their electricity consumption, outlined below.⁴

Text message or email

Irregular text messages or email messages can be used to highlight to consumers unusual consumption or changes in prices. For example, Glerup et al. (2010) analyse a feedback scheme in Denmark in which emails or SMS messages are sent to participants when

electricity consumption deviates from average levels by a pre-specified amount and find a 3% reduction in electricity demand as a result.

Internet site or mobile application

It is possible to display information in a useful graphical format by linking a mobile application or Internet website to the distribution company repository of consumption data. For example, Schleich et al. (2013) analyse an Austrian field trial in which consumers were provided with access to a website that displayed useful information relating to electricity consumption (with a one-day lag). They find limited impact of website feedback on consumer electricity demand.

In-home display

In-home displays (IHDs) use a wireless or optical reader to display information from the smart meter in a convenient and accessible manner to the household. Typical in-home displays feature graphics that display electricity consumption and price over the day and month, as well as indicators showing the current price of electricity. For example, Houde et al. (2013) analyse a programme that provided Google employees with an in-home display and find that electricity consumption was reduced by about 5% for several weeks following the receipt of the device.

Smart meters allow dynamic electricity pricing regimes

Smart meters record electricity consumption on an hourly or higher frequency and recording occurs with a time stamp. As a result, smart meters enable the electricity distribution utility to use prices that change over the course of a day or change from one day to the next.⁵ Changes in prices to reflect different costs of electricity provision over time are a market-based mechanism for encouraging energy conservation.

There are a number of pricing schemes that are enabled through the use of smart meters.

Real-time pricing

In a real-time pricing programme, residential consumers are exposed to the wholesale price of electricity. This can provide them with an incentive to conserve electricity during periods when demand is high or when supply is reduced. Real-time pricing is rarely applied to residential customers. Allcott (2011a) examines a case where selected Chicago consumers were exposed to real-time prices.⁶ He finds a reduction in peak-period consumption and a welfare gain for consumers on real-time prices.

Critical period pricing

Under a critical-period pricing tariff, customers pay a flat price for electricity except for during a certain number of “critical” periods during the year, when the consumer electricity rate increases substantially. These critical periods are times of particularly constrained supply, such as hot summer afternoons, when air conditioning demand peaks. The large increases in electricity price during a limited number of hours provide consumers with a substantial incentive to reduce demand during these periods. Jessoe and Rapson (2014) examine a critical peak period electricity scheme and find that consumers indeed respond by reducing demand.

Time-of-use pricing

In a time-of-use pricing scheme, the consumer electricity tariff changes by a predictable amount at predictable periods during the day. For example, during the summer season, a utility might declare the hours of noon to 7 p.m. on weekdays as “peak” periods, in which the price of electricity is double the price in other periods. Time-of-use pricing obtains some of the benefits of real-time pricing without exposing consumers to the fluctuating wholesale price of electricity.

Literature review

This section briefly reviews the findings from articles that focus on the relationship between real-time feedback and consumer electricity demand in the residential sector. A more detailed summary can be found in Rivers (2018). For a recent review of real-time pricing studies, see Faruqui and Sanem (2010).

Studying the effects of varying the price of electricity, Ireland’s Commission for Energy Regulation (2011), now the Commission for the Regulation of Utilities (CRU), conducted a behavioural trial to gauge customer response to various time-of-use tariffs and demand-side management stimuli (enabling technologies). Time meters were installed in 5 028 participating households, which were then assigned to treatment and control groups with the former receiving various combinations of time-of-use tariffs, in-home displays and fridge magnets and stickers that outlined different electricity use time bands and cost per band. The study found that participants equipped with in-home displays reduced their overall energy consumption by an average of 3.2% and their peak demand by 11.3%.

Faruqui and George (2005) looked at California’s Statewide Pricing Pilot involving 2 500 residential and small- to medium-sized commercial and industrial customers. They found that customers who received the Critical Peak Pricing (CPP) rate intervention saw peak electricity use reductions between 8% and 15%. However, when smart thermostats were added to the CPP intervention, peak reductions were even greater, reaching 25% to 30%.

Delmas et al. (2013) conduct a meta-analysis of 59 studies across multiple disciplines in the academic literature, all of which use randomised controlled trials (RCTs) to estimate the impact of information provision on electricity consumption. The study covers a wide variety of behavioural interventions that affect electricity demand, including real-time feedback, social norm comparisons, delayed feedback, audits and other interventions. The results of the meta-analysis suggest that real-time feedback causes a reduction in electricity consumption of about 11% on average. However, the authors caution that estimates of the effects of feedback are inflated in poor quality studies (for example, those that do not control for weather or other confounding factors). Across all types of feedback, they find that the treatment effect in high-quality studies (which represent only a small fraction of all studies) is only about one-quarter as large as the treatment effect for all studies. However, Delmas et al. (2013) neither provide an estimate of the effect of real-time pricing across high-quality studies in their data set nor clarify whether they consider any of the real-time feedback studies in their survey to be high quality.

Faruqui et al. (2010) summarise findings from several pilot experiments using real-time electricity feedback, most of which were published in non-peer-reviewed outlets. The pilots use a number of different interventions, including different types of IHD, different types of payment for electricity and different electricity tariffs, making it somewhat difficult to compare across studies. Faruqui et al. (2010) report that providing real-time

feedback through an IHD to consumers is associated with a reduction in electricity demand of 3% to 13%. However, some of the reviewed pilot projects use very small samples, and the methods used to estimate the treatment effect and design the experiment are not clearly presented in the paper (owing to the large number of interventions surveyed), so it is difficult to ascertain the validity of the results.

Further studies on the effect of real-time electricity feedback include Faruqui and George (2005), who find that critical peak electricity demand in Maryland was reduced among a pilot group of 1 021 households by 18% to 21% with education materials and different rate structures alone, and 23% to 27% with an IHD combined with dynamic pricing programmes. Another study by Faruqui and Akaba (2014), this time in Connecticut, found that the same IHD as tested in Maryland did not reduce energy consumption. However, when all enabling technologies, including A/C switches and in-home displays, were combined with dynamic pricing, customers reduce their energy consumption by 23%.

A report produced by Karkkainen (2004) summarises results from several energy efficiency pilot projects conducted in Europe, including one in Norway that included a sample of 10 894 participants divided into treatment and control groups. The pilot tested the effectiveness of “Ebox” load control relays, which allowed for direct two-way communication of consumption data via the Internet. The pilot saw average peak demand reduced by 11%.

While some studies found no significant effects, others faced methodological issues, related to randomisation, sample size or biases for example, affecting the validity of results. Allen and Janda (2006) studied households in Ohio that received electricity monitors displaying both real-time and historical electricity consumption in kilowatt hours or USD. Baseline data was also collected from the households using utility bill records and semi-structured interviews. Researchers found no statistically significant effect in electricity consumption between the treatment and control groups. Nilsson et al. (2014) presents the results of two field experiments in Sweden, which tested the effects of IHDs on energy consumption. Both studies showed statistically insignificant effects but were limited by small sample sizes of 32 and 42 households. The researchers also note that prior interest in environmental sustainability, energy savings and knowledge of IHDs – as well as the aesthetics of the IHDs themselves – all contribute to the impact these enabling technologies have on consumer behaviour. Westskog et al. (2015) also fail to find statistically significant effects on energy consumption a year after IHDs were installed in a pilot programme in Norway, probably hampered by the small sample size of 33 participants.

Furthermore, Xu et al. (2015) tested IHDs in two recently built apartment buildings in Shanghai, People’s Republic of China. Their sample consisted of 131 respondents, 76 of whom received the IHDs (the other 55 served as control). The treatment group reduced their energy consumption by an average of 9.1% over the control group, and researchers found that introducing IHDs also led to a 12.9% reduction in average standby power usage when compared to the control group. However, there was no randomisation in the sampling and treatment assignment, in addition to little to no discussion of the methodology used. Fenrisk et al (2014) found large effects for 2 groups who opted into an advanced metering infrastructure (AMI) – 27% reduction in electricity demand – but the authors do not discuss disaggregated effects and believe self-selection bias affected the opt-in groups.

In the last few years, several high-quality studies have been published that examine the effect of real-time feedback on consumer electricity demand. Gans et al. (2013) use a quasi-experimental approach based on the roll-out of smart meters with real-time feedback to a subset of Northern Irish households for this purpose. The context they examine, in which customers pre-pay for electricity and experience some of the highest electricity prices in Europe, is likely to produce large conservation impacts. They find that real-time feedback generates a large (11%-17%) reduction in electricity consumption for treated households, which is sustained over several years. It is emphasised that these large impacts are likely context-specific.

Three studies stand out in the domain leveraging IHD. Houde et al. (2013) report on a randomised controlled trial, in which real-time feedback on electricity consumption – with an IHD – was provided to a randomly assigned group of volunteering Google employees. They report a 5% reduction in electricity consumption due to the provision of an IHD but find that the effect does not persist more than a few weeks. Again, the particular context of the study (Google employees) makes it difficult to understand how IHDs might affect consumption in a broader population.

Similarly, Jessoe and Rapson (2014) sampled 437 households to examine the impact of providing an IHD in a context in which households are also exposed to critical peak period pricing (in which prices increase by 2 to 6 times for several hours at a time). They find that households with an IHD are significantly more responsive to critical peak prices than other households. Customers in the group who received both the price and IHD treatments saw their energy consumption decline from 8% to 22%. In contrast, those who only received the price intervention reduced their energy consumption from only 0% to 7% relative to the control group. Researchers attribute the increased energy savings of the IHD group not to price salience but to “consumer learning”.

Some studies also found more modest results in response to real-time feedback. Schliech et al. (2013) ran a trial in Austria with 1 525 residential customers randomly selected into treatment and control groups for a field trial examining the effects of real-time feedback on energy consumption. Feedback group participants reduced their average energy consumption by 4.5% relative to the control group. The researcher’s findings also suggest that electricity consumption is inversely correlated with the frequency of billing and metering. Sulyma et al (2008) conducted a pilot programme testing the efficacy of different price-signalling regimes and technologies in British Columbia, Canada. Two thousand residential customers were randomly assigned to three treatment groups and a control group, with Treatment Groups A and B receiving advanced meters and different communication packages and Group C receiving the same as B as well as an IHD. Group C experienced a 5% reduction in their overall energy consumption and a 9% reduction in peak demand – both attributable to the effect of in-home displays.

Finally, Harding and Lamarche (2016) analyse how the provision of real-time feedback technologies impacts consumer response to time-of-use (TOU) pricing. They find that households with IHDs do not significantly alter their profile of hourly electricity consumption compared to households without them in response to modest price changes.

In sum, the existing literature appears to consist of a fairly large number of studies of questionable quality, which finds varying but often large impacts of real-time feedback on electricity demand. More recently, several high-quality studies have been produced but while the internal validity of these studies appears to be high, it is not clear how well the results from these studies will transfer to other contexts because most have used rather

idiosyncratic populations or treatments. As a result, there remains a relatively significant gap in the understanding of how real-time feedback affects electricity consumption.

Methodology

Theoretical model

Real-time feedback technologies have the potential to reduce electricity demand by providing higher quality and quantity of electricity information to consumers. Based on this assumption, the model for this study reflects the consumption decisions of a “rational” consumer who makes optimal decisions when given information about price and efficiency. A short summary of the theoretical model is outlined below. The paper by Martin and Rivers (2015) provides a more detailed discussion of the model.

The model is based on a representative consumer who has the ability to choose how much electricity to consume. It assumes that the consumer optimises electricity consumption in response to more information on price and efficiency of electricity provided by a single service that is differentiated by time. Perfectly informed, the consumer makes decisions that maximise utility. In contrast, the imperfectly informed consumer misinterprets the price and efficiency, and his/her consumption deviates from optimal levels. In this model, utility maximisation occurs under constraints related to demand for energy services, the consumption of electricity and the consumer budget constraint. The model does not account for a consumer who makes non-optimising decisions even with perfect information.

Experimental design

This section presents a case study on the implementation of time-of-use electricity rates and in-home real-time electricity feedback technologies. The results and analysis presented in this section are based on the paper by Martin and Rivers (2015), which provides a more detailed discussion.

The case study presents an evaluation of a natural experiment in which in-home electricity displays are rolled out quasi-randomly to about 7 000 households served by an electricity distribution company in Ontario, Canada. This section first describes the context in which the programme was offered. It then describes the empirical approach used for understanding the causal effect of real-time feedback on household electricity consumption. Finally, it presents the results of the analysis.

Context

Households within the service area of an Eastern Ontario local electric distribution company (EDC) were offered the opportunity to participate in “peaksaverPLUS”, a demand response programme. Upon agreeing to participate in the programme, the EDC activates a device on the home’s electric hot water heater that allows the utility to remotely reduce the electricity consumption of the water heater during certain high-demand periods of the year (for up to four hours at a time and only between May and October).⁷

It is important to emphasise that the pre-condition for programme participation is ownership of an electric hot water heater. Since there is a very strong correlation between owning an electric hot water heater and using electricity as the primary space heating energy source (i.e. baseboard heaters), it is likely that the vast majority of the households

in the sample primarily use electricity for both space and hot water heating.⁸ The effect of real-time feedback on electricity consumption shown estimated in the report should, therefore, be interpreted as the effect of feedback on households with electric heat and hot water. In addition, it is important to emphasise that households that participate in the programme are not randomly drawn from the population but instead select into the demand response programme. The statistical implications of this selection are addressed below but here it is important to emphasise that the results obtained in this paper reflect the subset of households with electric water heaters that select into a demand response programme. It is not clear how generalisable the results are to the full population since demographic information on households was not available for this study.

In-home display

In return for participating in the demand response programme, participating households received an IHD. The IHD is wirelessly connected to the house's digital electricity meter (all Ontario households have been converted from analogue to digital electricity meters). It displays, in real time, the power consumption by the household in physical units (in kW), the current retail electricity price (in CAD/kWh) and the implied current expenditure on electricity (in CAD/day). It also shows the consumption of electricity over the previous 24 hours as well as over the previous month. Additionally, the IHD is equipped with an LED display, which glows a different colour depending on the current electricity price (e.g. green is off-peak; yellow is mid-peak; red is on-peak).

The IHDs were sent from the utility by mail to each participating household, with instructions for activation. The utility had already pre-paired each IHD with the electric meter at the residence so that upon receiving the IHD, the household could activate the device simply by plugging it into a standard electrical outlet (information on electricity consumption is then transferred wirelessly from the digital electricity meter to the IHD). The data indicates the date that the device was couriered to the customer, and this date is used as the start of the "treatment effect" associated with the IHD. It is important to note that there is no way of knowing if or when the consumer actually installs the IHD and so the effect that estimated throughout the report is an intent-to-treat effect, rather than a treatment-on-the-treated effect. The intent-to-treat effect is a lower bound on the treatment-on-the-treated effect.

Time-of-use electricity prices

In the Electricity Restructuring Act, 2004, the Ontario Energy Board (OEB) was mandated to implement a regulated price plan that included a TOU (Time of Use) pricing structure to more accurately convey the real costs of generation to consumers and to encourage customers to shift demand away from peak periods. Italy and Ontario are the only jurisdictions in the world to implement smart meters for all residential customers as well as an associated TOU pricing plan (Faruqui and Lessem, 2014). The roll-out of the smart meters and implementation of the TOU pricing plan were complete prior to the beginning of the period covered by this study.⁹

Ontario's TOU pricing structure divides each hour into one of three blocks representing off-peak, mid-peak, or on-peak periods. Weekends and holidays are off-peak periods, as are the hours from 7 p.m. to 7 a.m. each weekday. In the summer, hours from 7 p.m. to 7 a.m. are off-peak each weekday. Hours from 7-11 a.m. and 5-7 p.m. are mid-peak, while hours from 11 a.m. to 5 p.m. are on-peak.¹⁰ In the winter, the daytime blocks are

switched, such that peak periods are during the morning (7-11 a.m.) and evening (5-7 p.m.), while the mid-peak period is from 11 a.m. to 5 p.m.

The OEB adjusts TOU prices every six months in response to changes in electricity load as well as the profile of electricity generators in the province. During this period, real electricity prices have been trending upwards in Ontario.¹¹ The ratio of peak/off-peak prices has changed slightly during the study period but has remained between about 1.5 and 2.¹²

The impact of IHDs on electricity consumption is estimated by making use of the staged roll-out of IHDs to electricity consumers. In particular, the impact of the IHD on electricity consumption is determined by comparing a household that has just received an IHD with the same household just before receipt of the IHD and controlling for unobserved confounders using households that are just about to receive an IHD as a control group. Both of these households are programme participants and so are likely similar in important respects (at minimum, both have electric hot water heaters and likely have electric space heaters, for reasons discussed in the prior section).

The research design imposes the assumption that households that are enrolled in the IHD programme early in the year are equivalent to those that are enrolled in the programme later in the year. The identification approach might be compromised if these two types of households are significantly different. There are two reasons to think that the assumption is likely to be valid. First, although the roll-out of IHDs is long enough to exploit it for empirical purposes, from a household's perspective it is still relatively short; there is no reason to think that there is a significant difference between a household that enrolls in a demand response programme a few months before another household. Second, the phased roll-out was in part a response to resource constraints at the utility and this provides a source of exogenous variation in adoption date that is exploited in the analysis.

In addition to these qualitative arguments that suggest the timing of the roll-out is exogenous, it is possible to provide quantitative evidence. To do this, observations of electricity consumption prior to any households receiving an IHD are used (IHD roll-out began in January 2013 and the data on electricity consumption starts in September 2012). A comparison between pre-programme electricity consumption in these households is used to determine if there is any difference between early adopting and late adopting households that could contaminate the estimated treatment effects.

To operationalise this, the data are split into two groups: early adopters and late adopters. Households are split according to the median date of adoption (21 August 2013). Pre-programme electricity consumption in early and later adopters is then compared. Daily electricity consumption is clearly very similar between early-adopting and late-adopting households in the pre-treatment period, following the qualitative arguments above. Additional evidence on this point comes from a regression of pre-programme electricity consumption on the date of IHD receipt. There is no statistical relationship between these two variables. Martin and Rivers (2015) provide more formal statistical evidence that pre-treatment consumption in early-adopting and late-adopting households are identical.

Results and discussion

The main finding of the analysis is that households reduce electricity demand by an average of about 3% once they receive an in-home display (the result is “statistically” different from zero at conventional significance levels).¹³ As described below, this

reduction in electricity demand is maintained for at least several months following receipt of the device. The result is estimated based on a comparison of daily household electricity consumption within the same household before and after receiving an IHD, and controlling for temporal shifts in electricity consumption experienced by all households in the small service area of the utility, for example, due to holidays or changes in weather. The average effect is similar when controlling for household-by-season fixed effects, and also when hourly rather than daily data is employed for estimation. Tables showing this result and other relevant tests are provided in Rivers (2018) and more detail is available in Martin and Rivers (2015).

Temporal variation in household response

The hourly metering data produced by smart meters in Ontario allows for the possibility of breaking down the response by hour of the day.

For most hours of the day, the hourly effect of an IHD is very similar to the average effect over all hours of the day. In fact, the hourly effect is only statistically different from the average effect for 2 hours of the day: the hour up to 7 a.m. and the hour up to 7 p.m.

This is notable for two reasons. First, the stability in the effect across all hours of the day suggests that households are not dynamically responding to real-time information over all hours in the day but rather are permanently adjusting behaviour in a way that generates a relatively uniform response across hours of the day. Second, the result provides preliminary evidence that changes in the time-of-use price within a day are not driving major changes in the response to the IHD (a point explored further below). In particular, the largest reduction in electricity demand is in the hour leading up to 7 a.m., which is on off-peak price. The smallest response is in the hour leading up to 7 p.m., which is on mid-peak or on-peak price, depending on the season.

The study also observes the response according to the season of the year. Unlike the relatively flat response over the course of the day, there is a distinctive seasonal effect of the IHD on consumption. In particular, during the spring and summer months, there is a small and statistically insignificant impact of the IHD on electricity consumption. In contrast, during the winter and fall heating seasons, the IHD causes a roughly 4% reduction in the demand for electricity. This is suggestive evidence that households respond to the IHD in part by reducing the demand for space heating. Further evidence on this point is provided in the following section.

Household response by outdoor temperature

To provide additional evidence on the mechanism through which households are responding to the IHD, an additional regression was performed to examine how the hourly outdoor temperature interacted with the IHD dummy variable. Temperature is divided into equally-sized bins that span the range of temperatures in the data set, in order to enable visualisation of the potentially non-linear relationship between outdoor temperature and the impact of the IHD. This enables the possibility of establishing whether the presence of an IHD produces a differential response at different outdoor temperatures and helps to establish the mechanism by which households respond to the IHD.

The analysis shows that when the outdoor temperature is low, the presence of IHD results in a significant reduction in electricity consumption. In particular, at an outdoor temperature of -8°C or below, household electricity consumption is reduced by 4% to 6%

due to the presence of an IHD (with the larger reduction at lower temperatures). The effect of the IHD on electricity consumption declines near-monotonically as temperature increases until the outdoor temperature is between 2°C and 7°C, at which point the IHD appears to have no effect on electricity demand. At temperatures above 17°C, there is weak evidence that the IHD reduces household electricity demand.

The analysis provides additional evidence that households respond to receiving an IHD by adjusting the thermostat setpoint. When temperatures are extremely cold, suggesting a large heating load, the effect of the IHD is larger. Similarly, when temperatures are extremely hot, there is some evidence that households with an IHD consume less energy than households without. In contrast, when temperatures are less extreme, such that there is little requirement for heating or cooling, the IHD does not appear to have any effect on electricity consumption.

It is possible to make an estimate of the shift in thermostat setpoint that would give rise to the effects observed in this study. To provide an estimate, the HOT2000 building simulation model that is developed by Natural Resources Canada, is used to simulate household heating requirements for different indoor temperature setpoints and outdoor temperatures. Based on model simulations with different indoor setpoints and based on the weather in Ontario, a 1°C reduction in the indoor temperature setpoint is estimated to reduce building energy consumption by about 4% during the heating season. The United States Department of Energy suggests a reduction of 0.6°C (1°F) is sufficient to reduce energy consumption by about 3%.¹⁴ These two studies suggest that a possible interpretation of the findings here is that households responded to an IHD by reducing the thermostat setpoint by about 1°C or slightly less.

Persistence of household response

To establish whether IHDs can be (part of) a cost-effective strategy to encourage households to reduce their electricity consumption, it is critical to know whether the impact of the IHD on consumption is transitory or persistent. Prior studies have shed some light on this (e.g. Gans et al., 2013; Houde et al., 2013) but many have not followed households for sufficiently long periods to observe whether the response is transitory or persistent.

To test the persistence of household responses, a regression is performed in which the IHD treatment dummy is interacted with a variable indicating the number of weeks since the IHD has been received. As above, bins are used to enable the identification of a possibly non-linear response. The results show that the effect of the IHD appears to increase over time, from roughly 2% upon initial receipt to around 4% after households have had the IHD for several months. Importantly, the effect of the IHD on electricity consumption does not appear to be transitory but rather appears to increase fairly steadily over the five-month period over which households are observed following receipt of the IHD. Although it is again not possible to pin down the precise mechanism explaining this response, it is plausibly linked to the increased salience of electricity consumption under IHD adoption, leading consumers to shift habits in a persistent manner, for example by acquiring more energy efficient appliances or by permanently adjusting thermostat setpoints.

Real-time feedback and time of use electricity prices

The programme under study is in Ontario, a province with time-of-use electricity pricing for nearly all residential customers. During the period covered by the data, on-peak prices for electricity were about twice as high as off-peak prices. During the period covered by the data and the rollout of IHDs to customers, the time-of-use tariff for residential households changed twice: once in Spring 2013 when it switched from the winter to summer tariff structure, and once in Fall 2013 when it switched from summer to winter structure. At each switch, prices for electricity were also increased for each block of electricity. It is possible to use these two tariff changes to identify the impact of changes in electricity prices on electricity consumption, both for households with an IHD as well as for households that have yet to receive an IHD. As explained above, it is theoretically not clear whether households with real-time feedback should respond more or less to a price change than households without real-time feedback.

For households without an IHD, the estimated short-run elasticity of electricity demand with respect to price is between -0.17 and -0.37, depending on the time period and the model specification. This is well within the range of other estimates of the short-run elasticity for electricity demand (Lijesen, 2007). For households with an IHD, the estimated elasticity of demand is about -0.2 and does not change appreciably across different time periods. Interestingly, this implies that the elasticity estimated for households with an IHD is sometimes higher and sometimes lower than that estimated for households without an IHD. It is therefore not possible based on this study to conclude that real-time feedback appreciably increases or reduces the sensitivity to time-of-use electricity prices.

Conclusion

This chapter summarises the empirical literature on the effect of real-time feedback on electricity consumption decisions, develops a simple analytical model that describes how an optimising consumer responds to real-time feedback and presents results from an empirical study based on a large-scale roll-out of IHDs to electric utility customers in Canada. Taken together, the results suggest that real-time feedback is likely to cause consumers to reduce electricity consumption. The results also suggest that consumers are unlikely to shift patterns of electricity consumption (i.e. the timing of electricity demand throughout the day) substantially in response to receiving an IHD if differences in prices throughout the day are modest. Finally, the results suggest that households respond to receiving real-time information on electricity price and consumption in part by making one-time decisions of a durable nature – such as adjusting thermostat setpoints, hot water heater settings or upgrading the energy efficiency of household equipment – rather than by responding in real-time to the real-time information.

Notes

¹ This chapter draws on Rivers (2018). Some of the research presented in that article was conducted by Steve Martin and the author of the report, which is also reported in the paper by Martin and Rivers (2015).

² This report focuses only on the direct impacts of smart meters on household electricity consumption. Smart meters also confer other benefits, such as improved ability by the electricity distribution company to detect electricity theft, improved ability to manage electricity flows on the

electricity network, and reduced costs for electricity meter reading. These benefits do not accrue to the household directly, and are not the focus of this report.

³ While smart meters record electricity consumption on an hourly or higher frequency, they typically relay that information to the electric distribution company on a lower frequency, such as daily.

⁴ It is important to note that there have been a number of efforts to supply households with feedback on their electricity consumption that do not rely on real-time consumption information (e.g. Allcott, 2011b; Fischer, 2008). This document focuses on real-time feedback.

⁵ Prices that change over the course of a season are also possible with analogue meters.

⁶ In the case examined by Allcott (2011a), consumers were exposed to the day-ahead forecast of the wholesale price.

⁷ For a household to be eligible for the programme, it must have an electric hot water heater. During the two-year period covered by the data, the utility only implemented load control events for two four-hour periods. Because this report focuses on the response to real-time feedback and not the response to the load control interventions, days on which loads are controlled are removed from the sample. Load control events were declared by the Ontario Power Authority on 24 June and 16 July 2013, from 2 to 6 p.m.

⁸ Using a separate data set – the US Residential Energy Consumption Survey – shows that single family households with electric hot water heaters have roughly an 80% probability of also using electricity for space heat.

⁹ See: [www.ontarioenergyboard.ca/OEB/Industry/Regulatory+Proceedings/Policy+Initiatives+and+Consultations/Smart+Metering+Initiative+\(SMI\)/Smart+Meter+Deployment+Reporting](http://www.ontarioenergyboard.ca/OEB/Industry/Regulatory+Proceedings/Policy+Initiatives+and+Consultations/Smart+Metering+Initiative+(SMI)/Smart+Meter+Deployment+Reporting). Roll out of smart meters to Ontario residential customers was monitored by the Ontario Energy Board in monthly progress reports until June 2012, at which point 99% of eligible customers had smart meters installed. Some Electric Distribution Companies in Ontario implemented time of use pricing as early as 2009 and all EDCs had implemented time of use pricing by 2012. This study uses data from the period September 2012 to 2014.

¹⁰ Summer is defined as the months from May to October.

¹¹ Independent Electricity System Operator, Ontario.

¹² As in most utilities, the cost of the electricity commodity is just one component of the electricity bill received by the customer. Customers also pay a charge for delivery of electricity, as well as a regulatory charge, debt retirement charge and a service charge. Some of these additional charges scale with usage, while others are fixed. In total, the all-in electricity price varies less over price blocks than the electricity commodity charge.

¹³ The identification of this effect leverages the quasi-experimental roll-out of the IHDs to consumers. Martin and Rivers (2015) estimate this effect by performing a regression of the log of electricity consumption on a dummy variable indicating whether the household has received an IHD. The variable is equal to one starting on the date when the household received the IHD and equal to zero on days prior to the receipt of the IHD. In their regression, they control for factors that remain constant for a household and for temporal shocks affecting electricity demand. The paper provides further details on the controls and fixed effects involved in the analysis.

¹⁴ See <http://energy.gov/energysaver/thermostats>.

References

- Allcott, H. (2011a), “Rethinking real-time electricity pricing”, *Resource and Energy Economics*, Vol. 33(4), pp. 820-842.
- Allcott, H. (2011b), “Social norms and energy conservation”, *Journal of Public Economics*, Vol. 95(9), pp. 1082-1095.
- Allen, D. and K. Janda (2006), “The effects of household characteristics and energy use consciousness on the effectiveness of real-time energy use feedback: A pilot study,” in *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 1-7.
- Borenstein, S., M. Jaske and A. Rosenfeld, “Dynamic pricing, advanced metering, and demand response in electricity markets”, Center for the Study of Energy Markets.
- Commission for Energy Regulation (2011), “Electricity smart metering customer behaviour trials (CBT) findings report”, Technical Report, Commission for Energy Regulation.
- Delmas, M.A. et al. (2013), “Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012”, *Energy Policy*, Vol. 61, pp. 729-739.
- Department for Business, Energy and Industrial Strategy (2016), *Smart Meters: Quarterly Report end June, Great Britain*, Statistical Release: Experimental National Statistics, www.gov.uk/government/uploads/system/uploads/attachment_data/file/555971/2016_Q2_Smart_Meters_Report_FINAL.pdf.
- Eurostat (2016), *Private Households by Household Composition, 2006-2015 (Nb of Households in 1 000 and % of Household Types)*, Database, [http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Private_households_by_household_composition,_2006-2015_\(number_of_households_in_1_000_and_%25_of_household_types\)_new.png#file](http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Private_households_by_household_composition,_2006-2015_(number_of_households_in_1_000_and_%25_of_household_types)_new.png#file).
- Faruqui, A. and L. Akaba (2014), “The impact of dynamic pricing on residential and small commercial and industrial usage: New experimental evidence from Connecticut”, *Energy Journal*, Vol. 35(1), pp. 137-160.
- Faruqui, A. and S. George (2005), “Quantifying customer response to dynamic pricing”, *The Electricity Journal*, Vol. 18(4), pp. 53-63.
- Faruqui, A. and N. Lessem (2014), *Year Two Analysis of Ontario’s Roll-out of TOU Rates, Technical Report*, The Brattle Group.
- Faruqui, A. and S. Sanem (2010), “Household response to dynamic pricing of electricity: A survey of 15 experiments”, *Journal of Regulatory Economics*, Vol. 38(2), pp. 193-225.
- Faruqui, A., S. Sanem and A. Sharif (2010), “The impact of informational feedback on energy consumption – A survey of the experimental evidence”, *Energy*, Vol. 35(4), pp. 1598-1608.
- Fischer, C. (2008), “Feedback on household electricity consumption: A tool for saving energy?”, *Energy Efficiency*, Vol. 1(1), pp. 79-104.
- Gans, W., A. Alberini and A. Longo (2013), “Smart meter devices and the effect of feedback on residential electricity consumption: Evidence from a natural experiment in Northern Ireland”, *Energy Economics*, Vol. 36, pp. 729-743.
- Gleerup, M. et al. (2010), “The effect of feedback by text message (SMS) and email on household electricity consumption: Experimental evidence”, *The Energy Journal*, pp. 113-132.

- Harding, M. and C. Lamarche (2016), “Empowering consumers through data and smart technology: Experimental evidence on the consequences of time-of-use electricity pricing policies”, *Journal of Policy Analysis and Management*, Vol. 35(4), pp. 906-931.
- Houde, S. et al. (2013), “Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence”, *The Energy Journal*, Vol. 34(1), p. 87.
- Jessoe, K. and D. Rapson (2014), “Knowledge is (less) power: Experimental evidence from residential energy use”, *American Economic Review*, Vol. 104(4), pp. 1417-38.
- Karkkainen, S. (2004), *Results from the EFFLOCOM Pilots*, Technical Report, EU/SAVE.
- Lijesen, M.G. (2007), “The real-time price elasticity of electricity”, *Energy Economics*, Vol. 29 (2), pp. 249-258.
- Martin, S. and N. Rivers (2015), “Information provision, market incentives and household electricity consumption: Evidence from a large-scale field trial”, *Technical Report, SSRN Working Paper*, No. 2703947.
- Metering.com (2017), “Enedis issues paper on status of AMI rollout project”, <https://www.metering.com/reports/linky-smart-meter-enedis/>.
- Nilsson, A. et al. (2014), “Effects of continuous feedback on households electricity consumption: Potentials and barriers”, *Applied Energy*, Vol. 122, pp. 17-23.
- Office for National Statistics (2016), “Statistical bulletin: Families and households in the UK: 2016”, www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2016 (last accessed in May 2017).
- Office of the Auditor General of Ontario (2014), *Annual Report*, www.auditor.on.ca/en/content/annualreports/arreports/en16/v2_111en16.pdf.
- Rivers, N. (2018), “Leveraging the Smart Grid: The Effect of Real-Time Information on Consumer Decisions”, *OECD Environment Working Papers*, No. 127, OECD Publishing, Paris, <http://dx.doi.org/10.1787/6ad4d5e3-en>.
- Schleich, J. et. al. (2013), “Effects of feedback on residential electricity demand – Findings from a field trial in Austria”, *Energy Policy*, Vol. 61, pp. 1097-1106.
- Smart Energy International (2017), “Enedis issues paper on status of AMI rollout project”, *Smart Energy International*, <https://www.smart-energy.com/resources/reports-and-white-papers/linky-smart-meter-enedis/> (accessed on 27 March 2019).
- Sulyma, I. (2008), “Experimental evidence: A residential time of use pilot,” *Notes*, Vol. 100(6.33), pp. 6-33.
- U.S. Energy Information Administration (2017), *Table 10.10. Advanced Metering Count by Technology Type*, Database, www.eia.gov/electricity/annual/html/epa_10_10.html (last accessed in May 2017).
- Uribe-Pérez, N. et al. (2016), “State of the art and trends review of smart metering in electricity grids”, *Applied Sciences*, Vol. 6(3), p. 68.
- Westskog, H., T. Winther and H. Sæle (2015), “The effects of in-home displays: Revisiting the context”, *Sustainability*, Vol. 7(5), pp. 5431-5451.
- Xu, P. et. al. (2015), “Case study of smart meter and in-home display for residential behavior change in Shanghai, China”, *Energy Procedia*, Vol. 75, pp. 2694-2699.
- Zhou, S. and M.A. Brown (2017), “Smart meter deployment in Europe: A comparative case study on the impacts of national policy schemes”, *Journal of Cleaner Production*, Vol. 144, pp. 22-32.

Chapter 3. Protecting digital consumers

This chapter explains how behavioural insights (BI) can help understand and address the impact of online advertising on consumers. It also proposes practical next steps for policymakers to improve consumer understanding of online disclosures and explores how experimental approaches can help tailor disclosures to consumers regarding personalised pricing.

Introduction

During the last decade, policymakers have substantially increased the use of behavioural insights (BI) in the design and delivery of consumer policy (OECD, 2017a). This has complemented more traditional approaches and improved consumer policymaking by making it a more evidence-based discipline and enhancing the design of consumer policy interventions (OECD, 2017a). Methods such as behavioural experiments and surveys have been helping policymakers better understand various policy questions and providing the evidence with which to help address those questions.

In parallel with the growing interest in BI and consumer policy, the OECD Committee on Consumer Policy (CCP) has incorporated behavioural approaches as part of its broader work on effective consumer policymaking. In 2005, the CCP's first roundtable on economics for consumer policy recognised the potential for behavioural economics to offer a new source of insights for consumer policy, especially in the realm of information disclosure policies (OECD, 2006). In 2010, the CCP released the *Consumer Policy Toolkit*, which provided consumer authorities advice on how to define and respond to consumer problems, a clear six-step framework for policymaking and learnings from the fields of behavioural economics on consumer policymaking (OECD, 2010). Building on the toolkit, the CCP has released several publications focused on how governments and other public policy organisation can and have applied BI.

The CCP has found that the use of BI in consumer policy has mainly been in three areas: i) information disclosure and labelling; ii) regulation; and iii) consumer empowerment.

- i) Information disclosure and labelling have been the most common, typically for price representation and in e-commerce (OECD, 2017a). Some consumer authorities have utilised BI to inform their understanding of deceptive and unfair commercial practices. For example, there have been several enforcement actions that relate to drip pricing in online markets – a practice that can trigger behavioural biases.
- ii) Regulations have also been designed with a view to limiting the ability of businesses to take advantage of certain behavioural biases. For example, the latest EU Consumer Rights Directive (adopted in 2011) bans the use of pre-checked boxes for online sales, e.g. for express delivery options and travel insurance contracts when buying airline tickets (European Commission, 2014). This ban was informed by behavioural literature and its recognition of the power of default options (OECD, 2017a).
- iii) In the area of consumer empowerment initiatives, some consumer authorities have sought to support consumers by providing them with tools to mitigate the effects of behavioural biases. For example, there have been cases where businesses have been required to provide consumers with a simplified version of a consumer contract with the aim of overcoming information overload. In other cases, consumers have been gaining access to their consumption data in machine (i.e. computer) readable formats.¹ This has enabled intermediary services to provide actionable insights based on consumption patterns and transaction histories and support straightforward, effective comparison and decision-making in complex markets (and/or to automate these processes on behalf of consumers). Consumer education initiatives have also been informed by BI and designed with behavioural biases in mind (OECD, 2017a).

Despite the potential benefits, applying BI to certain aspects of consumer policy has not always been straightforward. This has especially been the case when the overarching aim is linked to an abstract or high-level objective. For example, the objective may be to empower consumers to make better decisions through improved online disclosures but it may be unclear whether a consumer should buy, not buy or shop around.

This contrasts with many of the broader policy areas in which BI have been applied, where policymakers (or indeed businesses) often have a clear behaviour that they want to “nudge” consumers or citizens towards (Thaler and Sunstein, 2008). Examples of this type of intervention include auto-enrolment for workplace pensions, along with policy interventions that aim to increase tax compliance (Hernandez et al., 2017), reduce energy use (Allcott, 2011), increase organ donation (Johnson and Goldstein, 2004), reduce litter (Kolodko, Read and Taj, 2016), or promote healthy eating (Shahnazari et al., 2016). In these scenarios, it is likely that policy interventions can be better-tailored and outcomes more easily measured than for, say, the improved online disclosure example given above.

Further, while many of the biases that have been uncovered by BI may be relevant to consumer issues broadly, this does not necessarily imply that wholesale changes to consumer policy are warranted. Instead, BI should be considered in the context of the specific policy question under investigation. In particular, findings from behavioural experiments cannot necessarily be generalised beyond the specific policy question that an experiment was designed to address (OECD, 2018).

Applying BI to consumer policy may also raise new challenges. For instance, nudge-based behavioural interventions may attract criticism if they are perceived to be a form of manipulation. Consumer authorities that are new to BI may also face practical challenges when starting to work in this area, e.g. if they need to dedicate time and additional resources to it, including for capacity building (OECD, 2017a).

This chapter examines the application of BI to consumer policymaking, specifically in online advertising, disclosure agreements and personalised pricing. It brings together and showcases various strands of the CCP’s BI work, including a planned future experiment. The chapter first provides an example of how policymakers can utilise BI to analyse the consumer impact of online advertising. The discussion provides an overview of the CCP’s work in this area, which utilises the existing BI literature to supplement its analysis of the practice.

Second, the chapter examines one of the most common consumer policy interventions: online disclosure requirements. The analysis shows how BI can determine consumer reactions to online disclosures. It also discusses the CCP’s recommendations on designing more effective online disclosures that are based on an understanding of consumers’ behavioural biases.

Third, the chapter explores the behavioural biases that are relevant to personalised pricing and provides initial thoughts on an experimental approach to examine consumer reactions to personalised pricing.

Finally, the chapter provides concluding remarks on the role of BI and consumer policy.

Context and problem setting

Using behavioural insights to better understand the consumer experience: Online advertising and consumer biases

In its analysis of online advertising, the CCP has utilised the BI literature to examine consumer biases and its implications (OECD, 2019). This approach is an example of how policymakers can use existing research on BI as an analytical lens through which a range of market practices and their impact on consumers can be viewed. Consideration of the behavioural lens is valuable because the origins of a problem can sometimes lie in consumers' behavioural biases or in firms' behaviours – which, as this section highlights, can act in ways that exploit consumers' behavioural biases.²

This section reproduces relevant aspects of the CCP's analysis and highlights two key ways in which the supply-side of the online advertising ecosystem can manipulate consumers' behavioural biases.

First, it shows how online advertising might present new ways to mislead consumers regarding the full costs of a product or service, or in respect of unexpected terms and conditions of a sale, in part due to behavioural biases such as the anchoring and endowment effects. Second, it looks at how the default or “status quo” bias might result in consumers disclosing and sharing more personal information than they would choose to, had they actively considered the choice. To provide context, a short overview of online advertising is provided below.

Online advertising: An overview

Advertising is always seeking to persuade, encourage or manipulate consumers into making purchases. It has long employed psychologists and other behavioural scientists in pursuit of these objectives, using them to “probe deep into consumers' minds and build advertising campaigns based on what they found there” (Clay, 2002). Vance Packard's seminal text *The Hidden Persuaders* (1957) attests to how the advertising industry was seeking and applying BI decades before governments and policymakers embraced the potential.

Online advertising is now the dominant form of advertising in many OECD countries and offers businesses the ability to reach consumers in ways that could only have been imagined previously. Online advertising has the potential to benefit consumers through more relevant and timely advertising and by funding a host of “free” online services. However, it also raises some new and complex challenges for consumers and consumer authorities (OECD, forthcoming).

As highlighted in Box 3.1, the OECD 2016 *Recommendation on Consumer Protection in E-Commerce* (OECD, 2016) includes provisions relating to advertising and marketing. In general, these provisions are intended to ensure that consumers understand when they are dealing with online advertising and that such advertising is not false or misleading. There is also a particular focus on consumer protection issues that can be challenging for consumers in the online context, such as pricing, digital content and endorsements (OECD, forthcoming).

Box 3.1. Selected OECD e-commerce recommendation principles on online advertising

- Advertising and marketing should be clearly identifiable as such (para 13).
- Advertising and marketing should identify the business on whose behalf the marketing or advertising is being conducted where failure to do so would be deceptive (para 14).
- Businesses should ensure that any advertising or marketing for goods or services are consistent with their actual characteristics, access and usage conditions (para 15).
- Businesses should ensure that advertised prices do not misrepresent or hide the total cost of a good or service (para 16).
- Endorsements used in advertising and marketing should be truthful, substantiated and reflect the opinions and actual experience of the endorsers. Any material connection between businesses and online endorsers, which might affect the weight or credibility that consumers give to an endorsement, should be clearly and conspicuously disclosed (para 17).
- Businesses should take special care in advertising or marketing that is targeted to children, vulnerable or disadvantaged consumers, and others who may not have the capacity to fully understand the information with which they are presented (para 18).

Source: OECD (2016), *OECD Recommendation of the Council on Consumer Protection in E-Commerce*, pp. 11-12, <http://dx.doi.org/10.1787/9789264255258-en>.

While online advertising shares the same objectives as its analogue forebears, the means of achieving these are radically different. Developments in artificial intelligence (AI) and machine learning, coupled with the routine collection of massive amounts of personal data by online services, allow for the creation of highly detailed profiles about individual consumers, which in turn enables cost-effective, precision-targeted (and retargeted) advertising at an unprecedented scale.

This hyper-personalised advertising at scale has been referred to as Online Behavioural Advertising (OBA), online profiling and behavioural targeting. Boerman et al. (2017) define OBA as “the practice of monitoring people’s online behavior and using the collected information to show people individually targeted advertisements”. The types of information that are being used in OBA include age, gender, location, education level, interests, online shopping behaviour and search history (Boerman, Kruijkemeier and Zuiderveen Borgesius, 2017).

Complementary technologies track user interaction with online ads to determine the effectiveness of advertising campaigns; and to provide the infrastructure for advertising payments to be tied to specific user outcomes, such as “clicks”, webpage visits or purchases. In terms of form, the Internet allows for new ways in which to present text, images, video and audio, and provides for interactive and individually tailored advertising in ways that no prior medium could support.

Given the above, it is not surprising that the Internet has transformed the nature and form of advertising and, as a result, disrupted the advertising and marketing sectors. Recent growth in advertising revenues is being driven by double-digit growth in online advertising (Letang and Stillman, 2016). In the United States, spending on online advertising is expected to exceed spending on television advertising in 2017 (Schuuring et al., 2017). Advertising is Google's primary revenue source, accounting for USD 79 billion in 2016 (Statista, 2018).

Benefits and risks for consumers

As the CCP's analysis of online advertising highlights (OECD, 2019), it can provide both benefits and risk for consumers. Benefits include the potential for more targeted, relevant and timely ads that could see consumers benefit from reduced search costs, greater awareness of relevant products and identification of and access to better deals. Online advertising also funds a range of nominally free online services for consumers, including: search services (e.g. Google); social networking services (e.g. Facebook); and digital news outlets (e.g. HuffPost). These have become a part of the fabric of people's digital experience. If such services were only available on a paid-for basis then some consumers might be worse off.

Risks include longstanding concerns around advertising's potential to mislead and misrepresent, which can now assume a digital form, along with new concerns that are inherent to online advertising. These include (OECD, 2019):

- consumers may not be able to identify some forms of online advertising
- online advertising could reduce consumer trust online
- online advertising may prey on consumer biases and vulnerabilities
- the potential for misleading advertising online
- threats from "malvertising"
- threats associated with increased data collection.

Online advertising can take advantage of consumer biases and cause consumer detriment

The ability of online advertising to target consumers' behavioural biases at scale and to potentially tailor the ads to a consumer's specific vulnerabilities, means that consumer decision-making may be more prone to manipulation through online advertising than for other forms of advertising. Further, online advertising may present new ways in which to mislead consumers regarding the full costs of a product or service, or in respect of unexpected terms and conditions of a sale (OECD, 2019). Consumer authorities will have to remain vigilant to these potential threats. As outlined in Box 3.2, there are several behavioural biases that are broadly applicable to the sphere of consumer policy. The relevant behavioural biases specific to online advertising are examined below.

Box 3.2. Examples of behavioural biases in consumer policy

Anchoring: Consumers “anchor” decisions around information that they think is the most important. Consumers may fail to adjust their perception of the value of the offer sufficiently, even when additional information is provided to them since they cannot stray far from the anchor point.

Availability heuristic: This describes the tendency for consumers to make judgments about the likelihood of an event based on how easily they can recall a relevant example.

Choice/information overload: When faced with either complex products or a bewildering array of choices, consumers can sometimes ignore possible choices, walk away from markets, or choose not to choose. Consumers can also rely on relatively simple “rules of thumb” or “heuristics” to make decisions.

Confirmation bias: This is the tendency of individuals to seek or interpret evidence in ways consistent with their existing beliefs, expectations or a particular hypothesis.

Default and status quo effect: Presenting one choice as default option can induce consumers to choose that option. The power of default is related to the status quo effect, where consumers have a strong tendency to remain at the status quo since the disadvantages of departing from it loom larger than the advantages of doing so.

Endowment effect: Consumers often demand much more to give up an object than they would be willing to pay to acquire it. The value of a good for consumers increases when it becomes a part of a consumers’ endowment.

Fairness: Consumers are generally concerned that market transactions should be fair to other consumers and often concerned about the conditions of supply (e.g. labour condition, use of environmental resources). This means that consumers are concerned not only about their own interest.

Framing: Consumers are influenced not only by the content of the information provided by suppliers but also by how the information is presented. Presenting an option in a certain way may induce consumers to evaluate the choice from a particular reference point.

Hyperbolic discounting/myopia: Consumers’ discount rate tends to rise steeply the shorter the time period being considered. This means that consumers tend to treat the present as if it were more important than other time periods. This explains outcomes such as low retirement savings in the absence of compulsion.

Loss aversion: See endowment effect (above).

Overconfidence: Consumers tend to think that they are more likely to experience an outcome from some action that is better than the average expected outcome. For example, many drivers think that they are safer than the average person, and when consumers are told that 20% of customers will benefit from a particular product, they tend to expect that they will be the part of that 20%.

Priming effect: When consumers are repeatedly exposed to certain objects, for example, through publicity, certain attributes can play an undue role in consumer decisions. Priming can influence preferences by making certain dimensions salient that would otherwise have been considered as less important.

Social norms: Consumers are often guided by the values, actions and expectations of a particular society or group. For example, when people are made aware of what others are doing, it can reinforce individuals' underlying motivations.

Time-inconsistency: While traditional economics assumes that consumers behave in a time-consistent way, i.e. that they are able to make decisions knowing their long-term interest and resist short-term actions that go against that, in reality, choices are not consistent across time periods. Consumers may face a conflict between short-term urges and long-term interests.

Sources: OECD (2017a), *Use of Behavioural Insights in Consumer Policy*, <http://bit.ly/2Ic01fJ> citing Kahneman, D., J.L. Knetsch and R.H. Thaler (1991), "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias", *Journal of Economic Perspectives*, Vol. 5(1), pp 193-206; OECD (2006), "The roundtable on demand-side economics for consumer policy", <https://bit.ly/2Q7U0UK>; OECD (2007), *Roundtable on Economics for Consumer Policy: Summary Report*, www.oecd.org/sti/consumer/39015963.pdf; OECD (2010), *Consumer Policy Toolkit*, <https://doi.org/10.1787/9789264079663-en>; Office of Fair Trading (UK) (2012), *Drip Pricing: UK Experience*, <https://bit.ly/2R1xNM5>; McAuley, I. (2013), "Behavioural economics and public policy: Some insights", *International Journal of Behavioural Accounting and Finance*, Vol.4(1), pp. 18-31; Oxera (2013), *Behavioural Economics and its Impact on Competition Policy: A Practical Assessment with Illustrative Examples from Financial Services*, www.oxera.com/Oxera/media/Oxera/download/reports/Behavioural-economics-and-its-impact-on-competition-policy.pdf?ext=.pdf; Shafir, E. (2008), "A behavioural perspective on consumer protection", *Competition and Consumer Law Journal*, pp. 302-317; Behavioural Insights Team (UK) (2014), *EAST: Four Simple Ways to Apply Behavioural Insights*, www.behaviouralinsights.co.uk/publications/east-four-simple-ways-to-apply-behavioural-insights/.

Endowment, anchoring and framing effects

Practices such as drip pricing and bait pricing may mislead consumers, especially given these practices can take advantage of consumer biases (OECD, 2018). Drip pricing occurs where a company advertises its product at a certain (low) price but later adds on additional non-avoidable fees or surcharges. Drip pricing preys on the anchoring and endowment effects. It can result in consumer loss as consumers feel committed to the purchase decision and stick with it despite the price increasing during the transaction process. According to one online study, it can increase purchase intentions, price and value satisfaction and reduce search intentions (Xia and Monroe, 2004). Bait pricing occurs where a business advertises a product at a low price to attract consumers to their store/site but there is only a low volume of products on sale (that sell out) and consumers end up making a more expensive purchase once there. Like drip pricing, this also makes use of the endowment effect and may be detrimental to online consumers (Ellison and Fischer Ellison, 2009). For these reasons, a number of jurisdictions have taken enforcement action against, or have laws that prohibit, drip pricing and bait pricing (OECD, 2018).

The BI literature also indicates that anchoring and framing effects may also be relevant with regards to the identification of online advertising. They could have an impact in relation to native³ and user-generated⁴ advertising when commercial messages largely look the same as the other content on a site. This could mean that consumers might not understand that such content is advertising and may give the information it conveys greater weight than they otherwise would have given it, due in part to anchoring or framing effects. Anchoring could lead consumers to make mistakes in valuing an offer or in comparing offers (OECD, forthcoming).

To the extent that a business has personal information about a consumer, this may also give them the ability to anchor or frame an advertisement in a way that highlights the characteristics of the product or service that the consumer values, while downplaying other characteristics. This type of conduct could be harmful if it results in consumers being misled or deceived. This could be through misrepresenting the true (financial) cost of a good or service, or by failing to highlight unexpected terms and conditions (for example, where the behaviour amounts to a subscription trap).

Use of social norms and persuasion profiling

Advertisements can use social norms to encourage sales. Social norms can be effective since consumers are often guided by the values, actions and expectations of whatever society or group they consider themselves to be a part of.

Some commentators have raised concerns about online advertisers using “persuasion profiling” to take advantage of the social norms that resonate best with a particular consumer. Such persuasion profiling could be used to target a consumer in real time, given that access to personal information enables a business to know about the consumer’s habits, current location and general vulnerabilities (Calo, 2014). Such targeting could take advantage of time-inconsistency biases, where consumers pursue short-term urges at the expense of their long-term interests. If this form of targeting is used to mislead consumers, then there could be potential for consumer harm (OECD, forthcoming).

Default and status quo bias

Default and status quo biases may lead consumers to disclose and share more personal information than they would otherwise choose to, given the tendency for individuals to go along with whatever the default (or status quo) choice or setting is, even when this may not be in their best interest. Default privacy settings that lead to a high level of disclosure and sharing could hence result in consumers disclosing and sharing more personal information than they would choose to, had they actively considered the choice (Calo, 2014). Conversely, default privacy settings that are more protective of consumers may be an effective way to improve consumer privacy online (OECD, forthcoming).

Using behavioural insights to further understand the impact of online advertising

The OECD’s analysis notes that several of the potential risks associated with online advertising could be better understood by undertaking behavioural experiments to determine the extent of the problem, and possibly to test potential solutions (OECD, forthcoming). In particular, it notes that behavioural experiments could be undertaken to:

- Test consumer identification of online advertising (building on the work already undertaken in Korean, Norway and the United States).
- Test consumer understanding of privacy statements to better understand when such statements might result in consumer harm by misleading consumers.
- Test how consumers react to receiving personalised ads or prices if businesses are required to communicate how those ads or prices have been personalised to the consumer.
- To test the effectiveness of measures that target behavioural biases (for example, “scarcity cues”).

Where issues can potentially be addressed through improved disclosure, the CCP's recent work on improving online disclosures with BI is relevant (see more below).

Improving interventions with behavioural insights: Online disclosures

Online disclosure requirements are one of the policy interventions most commonly implemented by consumer authorities. The analysis presented in this section is based on the CCP report *Improving Online Disclosures with Behavioural Insights*, which provides a more detailed discussion (OECD, 2018).

The objective of the report was to use a BI lens to assess how consumers react to online disclosures. Based on this assessment, the CCP also made high-level recommendations on enhancing the design of online information disclosures in ways that could help ensure consumers were better-empowered online. This approach provided a demonstration of how BI could be used to evaluate a policy intervention.⁵ As such, it is in keeping with Step 6 of the *Consumer Policy Toolkit* (OECD, 2010), which sets out the need for a policy review process to evaluate the effectiveness of a policy.

This section reproduces key parts of the CCP's analysis from its recent report as well as its key recommendations. To provide context, it starts with a short overview of online information disclosures.

The importance of online information disclosure: An overview

Access to good information is essential if consumers are to make decisions in their best interests when shopping online. For this reason, information disclosure requirements have been a key policy tool for empowering online consumers across the OECD. The importance of online disclosure is reflected in the OECD 2016 *Recommendation on Consumer Protection in E-commerce* (OECD, 2016), which dedicates an entire section to this subject.

Online information disclosures provide information about the seller, the goods and services on offer and the transaction itself, including information about payment methods, privacy policies and available dispute resolution and redress options (OECD, 2016). Businesses may make such disclosures through advertising and marketing (ranging from display banners to embedded "native advertising" in online blogs, social media sites and news sites), contractual terms and conditions, and legally-required notices. This information can be conveyed in different ways, including through pop-ups, links, text, images, audio and video. Businesses can provide information at different times during the customer journey, including through pre-transaction advertising and marketing, and during the course of the transaction, including during the payment process.

While online disclosures may be relatively new, economists have long focused on the role of information in correcting market failures that harm consumers. In particular, information economics recognised that, if left to the market, consumers may not always have enough or the right type of information to make informed decisions. This is especially relevant where there are "information asymmetries" (i.e. when sellers know more about the features and quality of their products or services than consumers). Further, the information economics literature recognises that there are costs involved for businesses in providing information and for consumers in searching for and understanding information. It, therefore, posits that disclosures that make product pricing and features more transparent could reduce search costs, potentially improving consumer outcomes. For these reasons, most OECD countries have legislation or guidelines that

ensure that consumers have access to clear, accurate and easily accessible information when shopping online (and off) (OECD, 2018).

The online shopping experience can differ significantly from “bricks and mortar” retail. For example, while consumers may not be able to touch and feel products when shopping online, they usually have access to a wider variety of information concerning those products and may be able to sample digital content. While there may be more information online, however, consumer attention remains a scarce resource both online and off. Further, what works for a recipient of a printed disclosure may not work when transferred to another recipient’s “screen of choice”, be it a computer monitor, tablet or mobile device (Benartzi and Lehrer, 2017). Different delivery channels require businesses to revise both the format and content of information.

Conventional approaches to online disclosure have often assumed that well-informed consumers will reliably make decisions that are in their best interests. However, as highlighted below, consumers can be influenced by behavioural biases that might limit the effectiveness of some forms of online disclosures. Further, BI can highlight why some of the disclosure tactics employed by certain businesses can be effective in eliciting consumer behaviour that is not always in a consumer’s best interest.

As the following highlights, while information disclosure is likely to remain a key tool for empowering consumers, findings from BI indicate that a rethink is required about the usefulness of certain forms of information disclosure.

Consumers’ behavioural biases relevant to online disclosures

The CCP’s analysis identified several behavioural biases that are relevant to consumers’ interactions with online disclosures. The following highlights how consumer detriment can arise when these biases affect whether and to what extent consumers engage with and comprehend disclosures online. Further, in instances where a business fails to adequately disclose either the true price of a product or service or the terms on which the offer is available, these biases can expose consumers to practices that are deceptive, misleading, fraudulent or unfair.

Information overload

Numerous studies have found that consumers are particularly prone to information overload when shopping online (Benartzi and Lehrer, 2017; Office of Fair Trading (UK), 2007). One way in which information overload can manifest itself is in few consumers reading online terms and conditions in full, if at all (see Box 3.3 below).

Further, businesses can potentially take advantage of information overload by making their products, services or prices more complex than required. Bar-Gill (2012) has raised concerns about this in the credit card, mortgage and mobile phone markets.

Anchoring and framing effects

With regards to framing, several studies have shown that the timing, context, layout and form of information disclosures can influence consumers’ ability to comprehend them (Federal Trade Commission (US), 2016).

Anchoring can mean that consumers do not value the entire offer properly, even when additional information is provided. This can lead to sub-optimal choices and consumer detriment. One common anchor point, especially when consumers are also facing information overload, is price.

Reference pricing (which compares a sale price to a pre-sale or competitor's price) can use framing and anchoring effects to inflate the perceived value of an offer. If reference prices are misleading, this has the potential to cause consumer detriment and to distort market outcomes (Office of Fair Trading (UK), 2010). While traditional economic theory suggests this should not have any impact, behavioural studies show that reference prices influence consumers' assessment of value (Ahmetoglu et al., 2010). Drip pricing and bait pricing can also prey on this behavioural bias.

Box 3.3. Do consumers read online terms and conditions?

There is growing evidence and acceptance that most consumers do not read online terms and conditions (T&Cs) in full when making online purchases. Estimates of readership vary depending on the presentation of T&Cs, the product or service they relate to and the way readership is measured (OECD, 2018).

For example, research undertaken for the European Commission found that while between 90% and 95% of consumers accept online terms and conditions, very few read these in full. Readership varied depending on how the terms and conditions were presented. Where consumers had to click through to terms and conditions, only 9.4% opened them, whereas 77.9% of consumers said that they at least scanned terms and conditions that could be scrolled through (Elshout et al., 2016).

The 2017 Ipsos Global Trends survey found that across the 23 countries featured, 64% of respondents agreed with the statement that: "I often don't bother fully reading terms and conditions on a website before accepting them" (Ipsos, 2017). However, self-reporting by consumers may be prone to overstating the actual figure. Server-side surveys indicate that barely 1% of consumers actually read terms and conditions (Ipsos, 2014).

The type of product also influences readership, with higher reported readership rates for mortgages (73%) and car rentals (72%), and lower rates for transactions on peer platforms (17% read them carefully) (Stark and Choplin, 2009; OECD, 2017b).

Readership of End User License Agreements (EULAs) appears to be even lower, with only 0.2% of consumers accessing EULAs (Bakos et al., 2014), a situation that the requirement of obtaining consumer consent does little to improve (Marotta-Wurgler, 2012). It has been calculated that the median time spent on software EULAs is 6 seconds, with at least 70% of users spending less than 12 seconds on the license page (Sauro, 2011).

A documentary film concerning online terms and conditions asserts it would take an average of one month per year for consumers to read the terms they are presented with online (Hyrax Films, 2013).

Sources: OECD (2018), "Improving online disclosures with behavioural insights", <https://doi.org/10.1787/39026ff4-en> (accessed on 10 August 2018); Elshout, M. et al. (2016), *European Commission (2016), Study on Consumers' Attitudes Towards Terms and Conditions (T&Cs) Final Report*, <http://bit.ly/2FrybKl> (accessed 4 November 2018); Ipsos (2014), *Global Trends 2014 - Navigating the New*, <http://bit.ly/2FEMmiA> (accessed 4 November 2018); Stark, D.P. and J.M. Choplin (2009), "A license to deceive: Enforcing contractual myths despite consumer psychological realities", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1340166 (accessed 4 November 2018); Bakos, Y., F. Marotta-Wurgler and D.R. Trossen (2014), "Does anyone read the fine print? Consumer attention to standard-form contracts", <http://dx.doi.org/10.1086/674424> (accessed 4 November 2018); Marotta-Wurgler, F. (2012), "Does contract disclosure matter?", <https://bit.ly/2p0ZVyw> (accessed 4 November 2018); OECD (2017b), *Trust in Peer Platform Markets - Consumer Survey Findings*, <https://doi.org/10.1787/1a893b58-en> (accessed 4 November 2018); Sauro, J. (2011), *Do Users Read License Agreements?*, <https://measuringu.com/eula/> (accessed 4 November 2018); Hyrax Films (2013), *Terms & Conditions May Apply (Film)*, <http://tacma.net/> (accessed 4 November 2018).

The endowment effect and loss aversion

As noted below, drip pricing preys on the endowment effect (Office of Fair Trading (UK), 2010). Once consumers have decided to make a particular online purchase (especially one involving considerable search time and effort), that decision effectively becomes part of their endowment. That is, in their mind, they have already acquired the good or service in question. If the cost of the purchase then increases beyond the price disclosed in the advert due to drip pricing, loss aversion may make it more difficult for consumers to step away from the purchase (OFT, 2012). This could result in consumers making purchases that are not in their best interests; that is, consumers may not shop around enough and may make purchases at higher prices than they would otherwise. Drip pricing can also make it much more difficult for consumers to understand and compare final prices (Greenleaf et al., 2016).

Bait pricing can also take advantage of the endowment effect and be detrimental to consumers (Ellison and Fischer Ellison, 2009). It occurs when a business advertises a product at a low price to attract consumers to its website but fails to disclose that only a limited quantity of the advertised product is available. If that limited quantity has sold out by the time an interested consumer reaches the website, the endowment effect may lead the consumer to purchase a more expensive alternative from the business rather than seeking out the original item elsewhere or comparing prices for the more expensive alternative.

Default biases

Default biases may exacerbate issues related to online disclosures. For example, default settings may opt consumers in for additional services with associated fees or charges through the use of pre-checked boxes or negative option marketing (see below). The effectiveness of pre-checked boxes in influencing consumer behaviour can be demonstrated with an example from Goldstein et al. (2008):

“A large national railroad in Europe made a small change to its website so that seat reservations would be included automatically with ticket purchases (at an additional cost of one to two euros), unless the customer unchecked a box on the online booking form. Whereas 9% of tickets included reservations before the change, 47% did after, earning the railroad an additional [US dollars] \$40 million annually.”

Similarly, default settings are relevant to negative option marketing,⁶ which has raised concerns for several consumer agencies across the OECD. In a Federal Trade Commission (FTC) workshop on the topic, panellists “discouraged the use of pre-checked boxes to obtain consumer consent because online research indicates consumers ignore them” (FTC, 2009). A representative from the National Advertising Division of the Council of Better Business Bureaus noted that pre-checked boxes signal that the information is routine or unimportant and hence, they are not an effective way of communicating with consumers (FTC, 2009)

To the extent that pre-checked boxes or other default settings (including negative option marketing) automatically sign consumers up for additional goods or services, financial commitments, disclosure of personal data or marketing material, it is likely in at least some circumstances that a significant proportion of consumers will fail to uncheck these options despite not actually wanting them or agreeing with them. This has great potential to result in consumer detriment. For example, consumers may be billed for goods or

services they do not want, they may unwittingly share personal information or they may be hassled by unwanted marketing.

Overconfidence and myopia

Overconfidence and myopia may lead consumers to ignore certain types of information including warnings, disclaimers and T&Cs. These biases may also lead consumers to choose the wrong product or service if businesses take advantage of them, for example, by highlighting immediate benefits such as a “free” mobile phone but obfuscating the cost of this over the lifetime of the contract (Bar-Gill, 2012).

Social norms and other factors

Consumer behaviour in relation to disclosures is likely to be influenced by social and cultural norms. For example, if a consumer’s friends and family use an application, consumers may be less likely to read the T&Cs and check the privacy settings. Further, a consumer’s ability to comprehend online disclosures is likely to be influenced by their age, education and familiarity with the Internet, among other things. It is therefore important to test disclosures on the relevant population of interest.

Policy implications

In addition to identifying where behavioural biases can render online disclosures ineffective, or, in the absence of adequate disclosure, result in consumers being treated unfairly, the CCP also examined the policy implications. In particular, it assessed how BI can help policymakers determine when and how best to use online disclosures and how they can provide the basis for improved disclosure design. The resultant high-level recommendations are summarised below.

Information and pricing should be clear and accurate

Protections against false and misleading information remain important, especially given consumers’ susceptibility to behavioural biases. In particular, consumers need to be protected from misleading pricing practices such as drip pricing and bait pricing. Further, to the extent that reference prices are used, these should not be false or misleading.

Material information should not only be in the terms and conditions

As highlighted in Box 3.3, few consumers read online T&Cs in full. For this reason, businesses should not use T&Cs to communicate important information to consumers. Material information should be made clear and salient to consumers, potentially in multiple places on a firm’s website and at various points during the transaction.

Use of images, audio and video should be considered

Information should be made as clear as possible with alternatives to text considered, as appropriate. In some scenarios, images, audio and video can more effectively convey information to consumers than even the clearest and simplest text. As well as considering the use of these alternative media, businesses should consider the overall look of their website, including colour and visual layout, for example.

Timing of disclosures is important

Another key factor to consider is the timing of disclosures. Given the endowment effect, warnings or material information that is only provided towards the end of a purchase may have little impact, since consumers have already made the decision to purchase the good or service. In other scenarios, information that would be quite useful to consumers may be rendered useless if it is received at a time when consumers cannot react. The timing of disclosures is something that should be subject to consumer testing.

Consumer consent should be express

If businesses want consumers to confirm a transaction for goods or services, consumer consent should be expressly sought and obtained. Such consent should not rely on default settings, negative option marketing or pre-checked boxes that consumers are unlikely to notice or change.

Information should be as simple as possible

Given the potential for information overload, online disclosures should be as simple as possible. Simplicity can be achieved by reducing the amount of text, using “signposts” to direct consumers to relevant information, and the use of “layering”, where different levels of information of increasing detail are presented to consumers as needed (for example, through active links).

Personalised disclosures require further consideration

Another way to improve consumer understanding of online disclosures is to better tailor the message to the individual concerned. In particular, if disclosure is only relevant to an identifiable target group, then it should ideally only be shown to that group. However, personalisation raises other potential consumer policy concerns. For this reason, it is an area that requires further research.

Technology-enabled information provision could facilitate comparison shopping and switching in complex markets

In complex markets, consumers might need additional information in order to compare offers in the market. In particular, consumers might need detailed information about their past or likely future usage. When such information is made available in a machine-readable form, it will be easier for consumers to make use of the services offered by public and private intermediaries.⁷

The importance of testing

Finally, the CCP’s online disclosure work also emphasises the importance of undertaking consumer testing – including through the use of behavioural experiments – whenever new disclosure requirements are being considered. The following section highlights how the CCP itself plans to use a behavioural experiment to test the impact of different disclosure requirements in relation to the issue of personalised pricing on line.

Methodology

Building on its work on online advertising and online information disclosures discussed in the preceding sections, the CCP plans to run its first behavioural experiment on personalised pricing online. Although there is little empirical evidence that personalised pricing (as opposed to dynamic pricing or personalised search rankings) is occurring in online markets, interest in this subject has grown. It has become clearer that online vendors are capable of tailoring prices to individuals based on granular personal data. The experiment will be an opportunity to test some of this chapter's recommendations for enhancing the design of online disclosures as listed in the policy implications above. The personalised pricing experiment will also complement the CCP's analysis of online advertising, as both personalised pricing and targeted online advertising are enabled by and are dependent on consumers' personal information.

After providing a brief overview of personalised pricing, this section outlines the relevant consumer biases. It then presents the research orientations that will underpin the experiment, before concluding with initial thoughts around the experimental approach.

Once complete, it is anticipated that the report detailing the experiment's findings will be published and made available at www.oecd.org/sti/consumer.

Personalised pricing: An overview

Definitions of online personalised pricing formulated in the literature typically comprise of three elements: i) that it is a sophisticated form of price discrimination; ii) that it hinges on online vendors' access to and utilisation of consumer data to generate meaningful inferences of what an individual consumer, or group of consumers, is willing to pay for a given good or service; and iii) prices offered to the individual consumer are set on the basis of these insights.

As such, personalised pricing is distinct from (but risks being conflated with) other forms of dynamic pricing that are prevalent online. For services and products where prices are adjusted in response to availability and/or overall demand yield management strategies for selling products and services that are perishable, time-sensitive and/or scarce. This includes event tickets, flights and inter-city rail travel, surge pricing in app-based cab services, and hotel rooms. These strategies can result in different consumers seeing different prices (or the same consumer seeing different prices if they engage with the vendor at different points in time).

Although coverage of personalised pricing frequently highlights concerns and the potential for consumer detriment, the practice is not inherently problematic. The characteristics of a market (i.e. to what extent is it competitive?) and the motives of each vendor adopting the practice, will determine the extent to which it proves beneficial or harmful for consumers in each instance; and/or whether the harms caused to those made worse off by this form of price discrimination outweigh the gains experienced by its beneficiaries (Office of Fair Trading (UK), 2013).

There are, as yet, few irrefutable occurrences of online personalised pricing and no evidence to suggest it is widespread. This may in part reflect the challenges that consumers and those working in their interest face in detecting the practice; and/or vendor hesitancy in deploying a practice that could fuel a consumer backlash if implemented covertly and then exposed, or that consumers could resent even if deployed transparently.

Given the above, the CCP's experiment will be developed on the basis of two untested assumptions: i) that personalised pricing happens at least occasionally and has the potential to become more frequent; and ii) that it is not normally detectable by consumers (although they may suspect it in some scenarios).⁸

The stated difficulty in detecting instances of personalised pricing will mean that – absent effective disclosure – consumers are unlikely to be able to: i) identify a personalised price when they encounter one; or ii) assess whether any personalised price they do encounter is higher or lower than the price they would otherwise have been offered.

Disclosure, therefore, offers a potential means of making the practice transparent to consumers – enabling them to determine whether a personalised price for a given good or service serves their best interests and, as a result, make better-informed purchasing decisions. On that basis, testing the impact of different forms of disclosure of personalised pricing on consumer behaviour will form the crux of the experiment.

Examples of behavioural biases that are potentially relevant to personalised pricing

Framing and loss aversion

Online vendors might exploit this bias by, for example, framing a personalised price offered as being superior to an inflated reference price and/or by creating the impression of scarcity, so that the consumer is compelled to buy through fear of losing out.

Fairness

In addition to considerations of whether a personalised price is fair, consumers might also deem the information asymmetry on which personalised pricing is predicated to be unfair and modify their behaviour accordingly. A recent behavioural experiment examining online targeted advertising and the influence of different approaches to (and degrees of) disclosure on consumers' purchasing intentions, produced evidence showing that when consumers realise that their personal information is flowing in ways they dislike, purchase interest declines. It also found that when third-party sharing of consumer data had occurred in ways that consumers deemed unacceptable, concerns about privacy outweighed people's appreciation for ad personalisation. Offering consumers the means to meaningfully control their privacy settings appeared to buffer any backlash to unacceptable data collection (John, Kim and Barasz, 2018). It would be interesting to see whether similar findings emerge in relation to personalised pricing.

Overconfidence

With regards to personalised pricing, it may be possible to observe whether limited forms of disclosure (e.g. the provider informs the consumer they are being offered a price that is unique to them, based on their purchase history for example) lead the consumer to assume the price they are offered is better than the average.

Research objectives and orientations

Ensuring consumers are able to make well-informed decisions is a key objective for effective consumer policy regimes (OECD, 2010). As noted above, by seeking to ensure that consumers have access to and can comprehend the information required to reach a well-informed decision, disclosure policies play a vital role in this respect. It is

anticipated that the CCP's experiment will seek to address two interrelated questions in relation to disclosure, namely:

- Which approaches to disclosure are most effective⁹ in enabling participants to:
i) identify when a transaction is subject to personalised pricing; and
ii) comprehend the implications of this practice for the transaction (e.g. how it affects the price they are asked to pay)? This could include testing variations in both the content and form of disclosures.
- To what extent does the disclosure of personalised pricing to consumers have a material impact on their decision-making?¹⁰

Proposed experimental approach

A laboratory experiment may be the most feasible option to test these initial questions because it will be in a controlled environment with minimal noise. The results of this laboratory experiment could inform a future field experiment or a natural experiment in the future.

The laboratory experiment will be built around high quality, realistic simulations of e-commerce and m-commerce sites, potentially resembling those of popular online retailers. It is anticipated that smartphones will be used by at least a proportion of participants, given their growing role in online purchasing. In controlled conditions, participants would engage in a range of online purchasing tasks and be presented with various personalised pricing and disclosure scenarios that support the testing of the various research objectives outlined above.

Conclusion

Policy officials can leverage BI as a tool to analyse the behavioural biases that are relevant to consumer policy. The CCP has applied a behavioural approach to analysing three important areas in consumer policy: online advertising, online disclosures and personalised pricing. The CCP has found that online advertising can take advantage of consumer biases, which can mislead consumers to buy full cost products or disclose personal information. Further behavioural experiments are necessary to better understand the associated risks. In the case of online disclosure, the CPP has used a behavioural lens to assess how consumers react to online disclosures. In addition, they have provided practical recommendations for policymakers on how to apply BI to improve online disclosures. Finally, the CCP has applied this behaviourally-driven approach to inform a future experimental design to understand the most effective ways for the consumer to identify personalised pricing and its implications.

Notes

¹ For example, the Midata initiative in the United Kingdom. See: <https://bit.ly/2hkNYBN>.

² This approach is consistent with Step 1 of the 6 steps presented in the *Consumer Policy Toolkit* (OECD, 2010).

³ The retargeting of online advertising occurs when a consumer starts seeing advertising for a product or service they have been researching appear on numerous sites across the web. Consumers can feel like the product or service is “following them” around the Internet.

Retargeting can take various forms and can be based on different information, such as search activities, responses to online advertisements, responses to email advertisements and “clicks”. For examples, see Sloane (2017).

⁴ The Federal Trade Commission has defined native advertising as: “content that bears a similarity to the news, feature articles, product reviews, entertainment and other material that surrounds it online” (FTC, 2015).

⁵ This approach is in line with Step 6 of the *Consumer Policy Toolkit* (OECD, 2010), which sets out the end for a policy review process to evaluate the effectiveness of a policy.

⁶ Negative option marketing refers to a category of commercial transactions in which sellers interpret a customer’s failure to take an affirmative action – either to reject an offer or cancel an agreement – as assent to be charged for goods or services. Negative option marketing can pose serious financial risks to consumers if appropriate disclosures are not made and consumers are billed for goods or services without their consent. See: <https://bit.ly/2CTWimI>.

⁷ See Note 1, above.

⁸ The recent behavioural experiment commissioned by the European Commission (2018) found that less than 20% of participants correctly identified price personalisation when they encountered it.

⁹ In the context of this experiment, disclosure would be deemed effective if it succeeds in enabling participants to identify instances of personalised pricing and comprehend its implications for the transaction at hand.

¹⁰ While it is anticipated that the experiment will observe and capture the impact of disclosure (and various forms of disclosure) on participants’ subsequent decision-making behaviours in the simulated transactions, it will not seek to define optimal outcomes in terms of decision-making and measure whether disclosure “nudges” participants towards these.

References

- Ahmetoglu, G. et al. (2010), *Pricing Practices: Their Effects on Consumer Behaviour and Welfare*, Prepared for the Office of Fair Trading, http://webarchive.nationalarchives.gov.uk/20140402172955/http://oft.gov.uk/shared_oft/business_leaflets/659703/Advertising-of-prices/Pricing-Practices.pdf (accessed on 02 August 2017).
- Allcott, H. (2011), “Social norms and energy conservation”, *Journal of Public Economics*, Vol. 95, pp. 1082-1095, <https://bit.ly/2AMm6je>.
- Bakos, Y., F. Marotta-Wurgler and D.R. Trossen (2014), “Does anyone read the fine print? Consumer attention to standard-form contracts”, *The Journal of Legal Studies*, Volume 43(1), pp. 1-35, <http://dx.doi.org/10.1086/674424> (accessed 4 November 2018).
- Bar-Gill, O. (2012), *Seduction by Contract: Law, Economics, and Psychology in Consumer Markets*, Oxford University Press.
- Behavioural Insights Team (UK) (2014), *EAST: Four Simple Ways to Apply Behavioural Insights*, www.behaviouralinsights.co.uk/publications/east-four-simple-ways-to-apply-behavioural-insights/.
- Benartzi, S. and J. Lehrer (2017), *The Smarter Screen: Surprising Ways to Influence and Improve Online Behavior*, Penguin.

- Boerman, S., S. Kruikemeier and F. Zuiderveen Borgesius (2017), “Online behavioral advertising: A literature review and research agenda”, *Journal of Advertising*, Vol. 46/3, pp. 363-376, <http://dx.doi.org/10.1080/00913367.2017.1339368>.
- Calo, R. (2014), “Digital market manipulation”, *The George Washington Law Review*, Vol. 82(4), pp. 995-1051, http://www.gwlr.org/wp-content/uploads/2014/10/Calo_82_41.pdf (accessed 4 November 2018).
- Clay, R. (2002), “Advertising as science”, *Monitor on Psychology*, Vol. 33/9, p. 38, <http://www.apa.org/monitor/oct02/advertising.aspx>.
- Ellison, G. and S. Fischer Ellison (2009), “Search, obfuscation, and price elasticities on the Internet”, *Econometrica*, Vol. 77/2, pp. 427-452, <http://dx.doi.org/10.3982/ECTA5708> (accessed 30 December 2018).
- Elshout, M. et al. (2016), *European Commission (2016), Study on Consumers’ Attitudes Towards Terms and Conditions (T&Cs) Final Report*, European Commission, Brussels, <http://bit.ly/2FrybKl> (accessed 4 November 2018).
- European Commission (2018), *Consumer Market Study on Online Market Segmentation Through Personalised Pricing/offers in the European Union*, European Commission, https://ec.europa.eu/info/publications/consumer-market-study-online-market-segmentation-through-personalised-pricing-offers-european-union_en (accessed on 10 August 2018).
- European Commission (2014), *DG Justice Guidance Document (The Directive on Consumer Rights)*, European Commission, https://ec.europa.eu/info/sites/info/files/crd_guidance_en_0.pdf.
- Federal Trade Commission (US) (2016), *Putting Disclosures to the Test Workshop: Staff Summary*, <https://www.ftc.gov/system/files/documents/reports/putting-disclosures-test/disclosures-workshop-staff-summary-update.pdf> (accessed on 02 August 2017).
- Federal Trade Commission (US) (2015), *Native Advertising: A Guide for Businesses*, www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses (accessed 30 December 2018).
- Federal Trade Commission (US) (2009), *Negative Options - A Report by the Staff of the FTC’s Division of Enforcement*, <https://bit.ly/2CTWiml> (accessed 4 November 2018).
- Goldstein, D. et al. (2008), “Nudge your customers toward better choices”, *Harvard Business Review*, Vol. 12, <https://hbr.org/2008/12/nudge-your-customers-toward-better-choices> (accessed on 08 August 2017).
- Greenleaf, E. et al. (2016), “The price does not include additional taxes, fees, and surcharges: A review of research on partitioned pricing”, *Journal of Consumer Psychology*, Vol. 24/1, pp. 105-124, <https://bit.ly/2VkNkWo> (accessed 30 December 2018).
- Ham, C.-D. and M.R. Nelson (2016), “The role of persuasion knowledge, assessment of benefit and harm, and third-person perception in coping with online behavioral advertising”, *Computers in Human Behavior*, Volume 62, pp. 689-702, <https://bit.ly/2Op8oq3> (accessed 4 November 2018).
- Hernandez, M. et al. (2017), *Applying Behavioral Insights to Improve Tax Collection: Experimental Evidence from Poland*, World Bank, Washington, DC, <https://bit.ly/2PaOVyJ> (accessed 4 November 2018).
- Hyrax Films (2013), *Terms & Conditions May Apply (Film)*, Directed by C. Hoback, <http://tacma.net/> (accessed 4 November 2018).
- Ipsos (2017), *Ipsos Global Trends: Section 8 - Online Behaviour*, <https://bit.ly/2HJnUK> (accessed 4 November 2018).

- Ipsos (2014), *Global Trends 2014 - Navigating the New*, <http://bit.ly/2FEMmiA> (accessed 4 November 2018).
- John, L., T. Kim and K. Barasz (2018), *Ads That Don't Overstep*, pp. 62-69, <https://hbr.org/2018/01/ads-that-dont-overstep>.
- Johnson, E. and D. Goldstein (2004), "Defaults and donation decisions", *Transplantation*, Vol. 78, pp. 1713-1716, <http://dx.doi.org/10.1097/01.TP.0000149788.10382.B2>.
- Kahneman, D., J.L. Knetsch and R.H. Thaler (1991), "Anomalies: The endowment effect, loss aversion, and status quo bias", *The Journal of Economic Perspectives*, Vol. 5(1), pp. 193-206.
- Kolodko, J., D. Read and U. Taj (2016), *Using Behavioural Insights to Reduce Littering in the UK*, Clean Up Britain, Warwick Business School, <http://www.nudgeathon.com/wp-content/uploads/2016/01/CLUB-REPORT.pdf> (accessed on 25 August 2017).
- Letang, V. and L. Stillman (2016), *Global Advertising Forecast Winter Update*, Magna.
- Marotta-Wurgler, F. (2012), "Does contract disclosure matter?", *Journal of Institutional and Theoretical Economics*, Vol. 168(1), <https://bit.ly/2p0ZVyw> (accessed 4 November 2018).
- McAuley, I. (2013), "Behavioural economics and public policy: Some insights", *International Journal of Behavioural Accounting and Finance*, Vol. 4(1), pp. 18-31.
- OECD (2019), "Online advertising: Trends, benefits and risks for consumers", *OECD Digital Economy Papers*, No. 272, OECD Publishing, Paris, <https://doi.org/10.1787/1f42c85d-en>.
- OECD (2018), "Improving online disclosures with behavioural insights", *OECD Digital Economy Papers*, No. 269, OECD Publishing, Paris, <https://doi.org/10.1787/39026ff4-en> (accessed on 10 August 2018).
- OECD (2017a), *Use of Behavioural Insights in Consumer Policy*, OECD, Paris <http://bit.ly/2Ic01fJ>.
- OECD (2017b), *Trust in Peer Platform Markets - Consumer Survey Findings*, OECD Publishing, Paris, <https://doi.org/10.1787/1a893b58-en> (accessed 4 November 2018).
- OECD (2016), *OECD Recommendation of the Council on Consumer Protection in E-Commerce*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264255258-en>.
- OECD (2010), *Consumer Policy Toolkit*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264079663-en>.
- OECD (2007), *Roundtable on Economics for Consumer Policy – Summary Report*, OECD, Paris, <http://www.oecd.org/sti/consumer/39015963.pdf> (accessed 4 November 2018).
- OECD (2006), "The roundtable on demand-side economics for consumer policy", *OECD Digital Economy Papers*, No. 116, OECD Publishing, Paris, <https://bit.ly/2Q7U0UK>.
- OECD (forthcoming), *Online Advertising: Trends, Benefits and Risks for Consumers*, OECD Publishing, Paris.
- Office of Fair Trading (UK) (2013), *Personalised Pricing - Increasing Transparency to Improve Trust*, OFT, <https://bit.ly/2ydgzmo>.
- Office of Fair Trading (UK) (2012), *Drip Pricing: UK Experience*, <https://bit.ly/2R1xNM5> (accessed 30 December 2018).
- Office of Fair Trading (UK) (2010), *Advertising of Prices*, OFT, http://webarchive.nationalarchives.gov.uk/20140402173016/http://oft.gov.uk/shared_oft/market-studies/AoP/OFT1291.pdf (accessed on 02 August 2017).

- Office of Fair Trading (UK) (2007), *Internet Shopping: An OFT Market Study*, OFT, <http://webarchive.nationalarchives.gov.uk/20140402163042/http://oft.gov.uk/OFTwork/markets-work/internet> (accessed on 01 August 2017).
- Oxera (2013), *Behavioural Economics and its Impact on Competition Policy: A Practical Assessment with Illustrative Examples from Financial Services*, www.oxera.com/Oxera/media/Oxera/downloads/reports/Behavioural-economics-and-its-impact-on-competition-policy.pdf?ext=.pdf.
- Packard, V. (1957), *The Hidden Persuaders*, D. McKay Co.
- Sauro, J. (2011), *Do Users Read License Agreements?*, <https://measuringu.com/eula/> (accessed 4 November 2018).
- Schuuring, M. et al. (2017), *The Digital Marketing Revolution Has Only Just Begun*, BCG, <https://www.bcg.com/publications/2017/sales-consumer-insights-digital-marketing-revolution-has-only-just-begun.aspx> (accessed on 10 January 2018).
- Shafir, E. (2008), “A behavioural perspective on consumer protection”, *Competition and Consumer Law Journal*, pp. 302-317.
- Shahnazari, S. et al. (2016), “Promoting healthy eating habits”, MaRS Solutions Lab, Canada, https://www.marsdd.com/wp-content/uploads/2016/03/MSL_ABI-Working-Paper.pdf (accessed on 25 August 2017).
- Sloane, G. (2017), *Snapchat Gives Advertisers New Tracking Tool to Target Ads*, AdAge, <http://adage.com/article/digital/snapchat-advertisers-pixel-tracking-tool-target-ads/311129/> (accessed 10 January 2018).
- Stark, D.P. and J.M. Choplin (2009), “A license to deceive: Enforcing contractual myths despite consumer psychological realities”, *NYU Journal of Law & Business*, Vol. Spring, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1340166 (accessed 4 November 2018).
- Statista (2018), *Google: Ad Revenue 2001-2016*, <https://www.statista.com/statistics/266249/advertising-revenue-of-google/> (accessed on 23 January 2018).
- Thaler, R. and C. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth and Happiness*, Yale University Press, New Haven.
- Xia, L. and K. Monroe (2004), “Price partitioning on the Internet”, *Journal of Interactive Marketing*, Vol. 18/4, <http://dx.doi.org/10.1002/dir.20017> (accessed 30 December 2018).

Part II. Organisational decision-making

Chapter 4. Cartel deterrence and the labour market for managers

This chapter examines the impact of different antitrust regimes on deterring cartel behaviour. It presents the results of a laboratory experiment that sheds light on how competing regulatory frameworks may affect managers' incentives to collude and, in turn, managers' labour markets.

Introduction

Although they share a similar underlying rationale, the United States (US) and the European Union (EU) antitrust regimes differ in several particular areas. One such area is the punishment of cartels, possibly the most worrisome competition law infringement (OECD, 2005).

The U.S. Department of Justice (DOJ hereafter) pursues individual convictions of both civil/administrative and criminal nature along with corporate fines since the inception of the Sherman Act (1890): “[v]iolations are punishable by fines and imprisonment”.

Within the EU, Article 23 of Council Regulation (EC) No. 1 of 16 December 2002 introduces the legal basis for implementing fines in case of violations: “2. The Commission may by decision impose fines on undertakings and associations of undertakings where, either intentionally or negligently: (a) they infringe Article 81 or Article 82 of the Treaty; [...] 5. Decisions [...] shall not be of a criminal law nature”. Additionally, until relatively recently, EU jurisdictions typically only applied corporate fines. The United Kingdom (UK) and an increasing number of EU countries (Austria, Belgium, Denmark, Estonia, France, Germany, Greece, Hungary, Ireland, Italy, Poland, Portugal, Romania, the Slovak Republic, Slovenia and Spain) now also include in their jurisdictions the possibility of individual penal sanctions for cartels.

The goal of this chapter is to provide experimental evidence on the relative effectiveness of individual fines vs. corporate fines to deter cartels. This is somewhat related to two different antitrust frameworks: the US antitrust framework, which applies individual fines; and the EU antitrust framework, which is mainly based on corporate fines. The issue at stake has clear implications for competition policy and can be of interest to academic economists as well as competition authorities and antitrust practitioners.

In addition, we analyse the extent to which such antitrust regimes may influence the labour market for managers, not only through the type of contract they are offered – fixed salary vs. fixed salary plus a variable component – but also through salary levels. To the best of our knowledge, this type of analysis – the relationship between antitrust regimes and manager labour markets – is novel in the literature.

More specifically, our research goals are:

- Do antitrust law regimes involving individual fines have larger cartel deterrence power than antitrust regimes encompassing only corporate fines?
- Do antitrust regimes cause distortions in managers’ labour market, namely by making particular types of contracts more appealing to shareholders?

Our research paper is theoretical and experimental. The theoretical model looks into the incentives for Bertrand oligopolists to form cartels. Each firm is operated by a manager, whose contract (offered by the principal-shareholders) is a combination of a fixed and a variable salary component that depends on the firm’s revenue. As usual, within a principal-agent framework, managers choose whether or not to accept the contract and, in case they do accept, in addition to setting prices, managers can choose to form an explicit cartel, thus sustaining monopoly prices. As outlined above, we look at manager incentives for collusion under two antitrust regimes – one where, if the cartel is detected, fines are revenue-based and are levied on the firm (corporate fines); and another where fines are salary-based and levied on the manager (individual fines).

Our underlying rationale for following an experimental approach as a complement to our theoretical approach is the following. It is well known and well acknowledged that the empirical literature in the broad field of economics of cartels suffers from sample selection bias. The main reason is that observed empirical data consists of detected cartels, which may not be representative of the whole population of cartels (and which includes those that have not been detected).

In addition, empirical work usually suffers from the inability of measuring all the possible factors that may affect an outcome or variable of interest. As well explained in Chamberlain (1948), “[t]he data of real life are necessarily the product of many influences other than those which it is desired to isolate [...] [u]nwanted variables cannot be held constant or eliminated...”. By contrast, the toolkit of experimental economics allows us to prevent the limitations inherent to empirical work. Indeed, experiments allow us “to study in isolation and under known conditions the effects of particular forces” (Ruffle and Normann, 2011). In contrast to empirical work, laboratory experiments generate data under controlled conditions, which, by changing the experimental conditions and environment, allow us to make meaningful comparisons. In other words, the researcher has the possibility of changing just one variable of interest in each treatment of the experiment, maintaining all else equal. This methodology more adequately reveals any nexus of causality between variables, whereas with empirical data, more often than not, we cannot unambiguously uncover causal relationships.

Naturally, experiments also have disadvantages, especially those in the industrial organisation field: subjects are typically students, who play the role of a fictitious economic agent (usually firms), in a highly stylised economic environment with fairly low stakes (and usually with no downside, as subjects do not typically lose money in an experiment). Nevertheless, many of the criticisms usually made to the application of experiments in industrial organisations are not fully convincing or, at least in some cases, can be easily dismantled (Ruffle and Normann, 2011; Hinlopen and Normann, 2011).

The experimental part of the research paper involves quadropoly (four firms) markets, where identical firms selling a homogenous product play a repeated Bertrand game with inelastic demand. We choose experimental markets with four firms since there is experimental evidence of greater gains and higher incentives from cartel behaviour in such markets (Fonseca and Normann, 2012; 2014).

Our main experiment consists of four treatments:

- A baseline treatment where subjects cannot communicate (“NoChat”).
- A treatment where subjects can communicate and choose whether or not to form a cartel, without the presence of a competition authority (“ChatNoLaw”).
- A treatment where subjects can communicate and the competition authority levies individual fines on firms’ managers if a cartel is detected (“ChatUS”).
- A treatment where subjects can communicate but only corporate fines are levied in case of cartel detection (“ChatEU”).

Each treatment is made up of 12 sessions of 12 subjects each, which are divided into 2 sets of 6 sessions (which we refer to as Phase 1 and Phase 2). Subjects in the first set of 6 sessions (Phase 1) play the role of firm owners and choose contracts for their managers in the second set of 6 sessions (Phase 2). In the second set of sessions, subjects play the role of managers in four different markets, where in each market they are randomly matched with three other subjects from the same session. In a market, managers are

expected to make a pricing decision per period, until the market eventually ends (more details below).

In each market, managers are offered a particular salary by their firm owner. The salary has a fixed component and a revenue-based component. A contract can be “low-powered” or “high-powered”. Both types of contract have the same fixed component, but high-powered contracts have a larger revenue-based component than low-powered contracts. Each subject makes decisions in two markets under a high-powered contract and in the other two markets under a low-powered contract.

In the baseline treatment (NoChat), subjects make their pricing decisions in each market in the absence of any communication. By contrast, in the other three treatments, subjects have the choice of whether to start communicating via a chat messenger. As a result, we endogenise cartel formation as in Fonseca and Normann (2014). Endogenising cartel formation allows us to study cartel deterrence, which is one of the goals of the paper.

In our first treatment, ChatNoLaw, cartel formation has no negative consequences to participants: no competition law exists and cartels, if formed, attract no sanctions whatsoever. By contrast, competition law is assumed to exist in the ChatUS and ChatEU treatments and an antitrust authority monitoring each market may punish cartels if they are detected. We follow Fonseca and Normann (2014) and various experiments on leniency programmes (e.g. Apesteguia et al., 2007; Hinloopen and Soetevent, 2008; Bigoni et al., 2012) and assume that, with a fixed probability, the competition authority may audit a market. If a market is audited and managers in that market have decided to communicate, the cartel is automatically detected and the authority will levy a punishment that differs depending on the applicable antitrust regime.

In the ChatUS treatment, this punishment is individual in nature and is assumed to take the form of an individual fine that falls on the manager’s salary. Under an antitrust regime with individual convictions, offenders could face jail time, but we are, for obvious reasons, unable to implement such a punishment in an experiment. Therefore, we make the assumption that individual punishments in such antitrust regimes can be made equivalent to a monetary reduction in the manager’s salary. By contrast, in the ChatEU treatment, the punishment takes the form of a corporate fine, i.e. there is a reduction in firms’ revenues (and only indirectly, through the variable salary component, managers’ salaries).

In a nutshell, our results are the following. From a theoretical viewpoint, we find that particular type of fine for collusion does play an important role in managers’ decisions. Shareholders, anticipating managers’ choices, find it optimal to offer fixed wage salaries in a corporate fine antitrust regime, but when fines are of an individual nature, they choose instead a combination of a fixed and a variable component or, in some cases, a purely variable contract.

From an experimental viewpoint, our main conclusion is that, while both type of fines have some success in deterring collusion, the observed behaviour diverges significantly from the theoretical expectation. One possible explanation for this divergence is (unobserved) manager heterogeneity: not only may managers be different from one another (e.g. in their degree of risk aversion), but they may also diverge from one another in terms of strategy choices, e.g. strategies played under competition or strategies played in case of deviation from the cartel agreement.

Not only do the questions we address in this research project have academic relevance, but they also have implications for competition policy. In fact, a recent ongoing debate has evolved on the criminalisation of EU competition policy (Shaffer et al., 2015). In addition, Shaffer et al. (2015) point out that there is a worldwide trend in adding a criminal law component to competition policy. For example, among BRICS (Brazil, Russia, India, China and South Africa) countries, Brazil and South Africa have started to introduce individual sanctions (fines and imprisonment) for cartel offences; also, a growing number of Latin American and Asian countries have been doing so. We hope to be able to contribute to this ongoing discussion with the outputs of this research paper.

Context and problem-setting

In this subsection, we give an overview of the European Union – France, Germany and the United Kingdom (UK), the top three European countries in gross domestic product (GDP) terms – Canada and the United States legal frameworks for dealing with cartels. This overview reveals the contrast between the EU and the US in punishing detected cartels. France, Germany and the UK stand closer to the US regime; in fact, these countries have a cartel enforcement regime within which managers and executives may be punished through monetary fines and/or imprisonment. Nevertheless, Canada and the United States tend to have, in their jurisdictions, higher maximum individual punishments than these three European countries.

EU legislation

The legal framework within which the European Commission (EC), through the Directorate-General for Competition (DG Competition), fights cartels is provided by Article 101 of the Treaty on the Functioning of the European Union (TFEU, Box 4.1).

Box 4.1. Treaty on the Functioning of the European Union, Section 101

“1. The following shall be prohibited as incompatible with the internal market: all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the internal market, and in particular those which:

- *directly or indirectly fix purchase or selling prices or any other trading conditions*
- *limit or control production, markets, technical development, or investment*
- *share markets or sources of supply*
- *apply dissimilar conditions to equivalent transactions with other trading parties, thereby placing them at a competitive disadvantage*
- *make the conclusion of contracts subject to acceptance by the other parties of supplementary obligations which, by their nature or according to commercial usage, have no connection with the subject of such contracts.*

2. Any agreements or decisions prohibited pursuant to this Article shall be automatically void.

3. *The provisions of paragraph 1 may, however, be declared inapplicable in the case of:*

- *any agreement or category of agreements between undertakings, – any decision or category of decisions by associations of undertakings*
- *any concerted practice or category of concerted practices, which contributes to improving the production or distribution of goods or to promoting technical or economic progress, while allowing consumers a fair share of the resulting benefit, and which does not:*
 - *impose on the undertakings concerned restrictions, which are not indispensable to the attainment of these objectives*
 - *afford such undertakings the possibility of eliminating competition in respect of a substantial part of the products in question.”*

Source: European Union (2012), *Consolidated Version of the Treaty on the Functioning of the European Union*, <https://eur-lex.europa.eu/legal-content/en/txt/?uri=celex%3a12012e%2ftxt> (accessed on 21 March 2019).

The first paragraph of Article 101 of the TFEU prohibits all agreements between competitors that prevent, restrict or distort competition within the EU and that may adversely affect trade between member states. In addition, Article 101 also lists a series of agreements between firms that, if put into place, will be automatically declared legally invalid. Finally, Paragraph 3 of Article 101 provides exceptions, whereby agreements that would normally be declared null are instead considered legally valid if they bring about efficiency gains due to innovation (improved production, distribution of products or promotion of technical or economic progress) and do not impose unnecessary restrictions or significantly restrict competition.

Discussing the method for computing fines set out by EU competition law is beyond the scope of this paper (see European Commission, 2013). However, the essence of our theoretical and experimental approaches involves the distinction between corporate fines (with no criminalisation), as under the EU framework,¹ and punishment mechanisms targeting both firms and individuals (criminalisation, with the possibility of prison sentences), as under the US antitrust regime.

The European Commission (EC) is the institution responsible for cartel enforcement process in the EU regarding cartels affecting both competition in the EU and cross-border trade between member states. A detected cartel member may appeal the EC's decision to the General Court (GC) that has both the power to annul, reduce or increase the fine imposed by the EC, as well as to review the entire investigation. In the second stage of appeals, the EC, a cartel member or both can appeal to the European Court of Justice (ECJ). The ECJ has the power to annul, reduce or increase the fines imposed by the GC; however, the ECJ does not have the power to reconsider the investigation by analysing facts and evidence that the GC used to corroborate its decisions Hellwig and Hüscherlath (2017) detail the cartel enforcement process in the European Union.

Anti-cartel legislation in major European countries

Germany

In Germany, the anti-cartel law is the Act against Restrictions on Competition (ARC). Section 1 of the ARC corresponds exactly to Article 101(1) of the TFEU. The Federal Cartel Office (FCO) enforces the ARC and there are two separate proceedings according to the severity of the infringement. The FCO deals with minor infringements with just administrative proceedings included in the ARC. For more serious infringements, for which the FCO imposes fines, proceedings follow both the Code on Administrative offences and the Code on Criminal Procedure.

For the purposes of our paper, it is relevant to note that the German anti-cartel legislation prescribes sanctions for both firms and individuals. Prison sentences, of up to five years, are possible only for bid-rigging in tender proceedings. The maximum fine for a firm is equal to 10% of its worldwide turnover of the last completed business year. The maximum individual fine is EUR 1 million for serious violations (cartel activity, territory allocation, quotas, bid rigging) and EUR 100 000 for minor violations.

France

In March 2009, the President of the French Republic and Parliament made a reform (Lasserre, 2009) of the competition enforcement system, by creating an independent Antitrust Authority (*Autorité de la concurrence*). The current antitrust law underpinning the cartel enforcement process contemplates both corporate fines up to 10% of global turnover of the group to which a firm belongs to and individual sanctions (fines) of up to EUR 75 000 and/or imprisonment of up to 4 years (Autorité de la concurrence, 2016). These punishments are for all types of cartel activity (price fixing, market sharing agreements and bid-rigging). However, criminal convictions have been extremely rare.

United Kingdom

The 1998 Competition Act provides the legal framework within which the Competition and Markets Authority fights cartels, agreements between companies, decisions by associations of firms and concerted practices that restrict, prevent and distort competition within the UK. More specifically the Competition Act prohibits the following agreements and practices (Practical Law Competition, 2019):

- Directly or indirectly fix purchase or selling prices.
- Limit or control production, markets, technical development or investment.
- Share markets or sources of supply.
- Apply dissimilar conditions to equivalent transactions with other parties, placing them at a competitive disadvantage.
- Make the conclusion of contracts conditional on acceptance of unrelated obligations.

The Competition and Markets Authority can impose fines up to 10% of the firm's global worldwide turnover in the last financial year. In addition, under the Enterprise Act 2002, individuals that arrange with other persons to implement actions such that two or more

firms undertake cartel activity (price fixing, sharing markets or customers, limiting production or supply, or bid-rigging) may be imprisoned for up to five years. For cartel activity undertaken before 1 April 2014, for cartel conduct to be considered a criminal offence, attracting the prosecution of individuals, such individuals had to act dishonestly. The Theft Act 1968 defines the meaning of “dishonestly” (Practical Law Competition, 2019). The Court may disqualify for up to 15 years a corporate manager responsible for cartel activity. To sum up, individuals guilty of cartel offences may be subject to imprisonment and disqualification in their managerial roles.

North American anti-cartel law

United States

The United States introduced antitrust law with the Sherman Act in 1890. In 1914, the US Congress created the Federal Trade Commission (FTC) with the Federal Trade Act. Any agreements or decisions between firms or individuals associated with fixing pricing, market division or bid-rigging always represents a violation (“per se” breach of the law) of the Sherman Act (Federal Trade Commission, n.d.).

The enforcement regime of antitrust laws is carried out by the FTC and the U.S. Department of Justice (DOJ) Antitrust Division and it is both civil and criminal in nature. Indeed, under the Sherman Act, both firms and individuals may be criminally prosecuted. Sanctions include corporate fines of up to USD 100 million along with individual fines of up to USD 1 million and a maximum of 10 years in prison. Criminal convictions are not uncommon. Moreover, the federal law allows an increase of the maximum fine up to twice the gains the conspirators have obtained or twice the monetary loss suffered by the victims if either of those two amounts is greater than USD 100 million (Federal Trade Commission, n.d.).

Canada

The Competition Act (14 months older than the US Sherman Act) is the relevant antitrust law in Canada. Section 45 of the Competition Act deems illegal any agreement and decision between individuals or companies leading to fix, maintain, increase or control prices, allocate markets, territories or customers, or fix, maintain, control, reduce production or supply of a product. The Commissioner of Competition has the power to pursue antitrust investigations through the Competition Bureau. In case of solid evidence of a breach of the Competition Act, the case passes on the Public Prosecution Service for criminal prosecutions before either the Federal Court or the Courts of Province. The Federal Court or a Court of Province may impose a fine up to CAD 25 million and/or up to 14 years in prison. To convict an individual, the prosecutor must prove that such an individual has undertaken a cartel offence and had the intention of doing so (Competition Bureau, 2018).

Cartel enforcement in practice

In this section, we review some legal practice by presenting cartel cases and enforcement processes for the European Union (France, Germany and the UK), Canada and the United States.

EU cartel case

The trucks cartel case (European Commission, 2017) gave rise to the highest fine levied by the European Commission since 1969: EUR 3.8 billion. The firms involved in the cartel were MAN, DAF, Daimler, Iveco and Volvo/Renault, with which the EC reached a settlement, and Scania, which refused to settle. Scania along with the other 5 companies established a cartel for a 14-year period, from 1997 to 2011.

The Commissioner for Competition, Margrethe Vestager, said: “This cartel affected very substantial numbers of road hauliers in Europe since Scania and the other truck manufacturers in the cartel produce more than nine out of every ten medium and heavy trucks sold in Europe. These trucks account for around three-quarters of inland transport of goods in Europe and play a vital role in the European economy. Instead of colluding on pricing, the truck manufacturers should have been competing against each other – also on environmental improvements” (European Commission, 2017).

The Commission’s investigation ascertained that between 1997 and 2004 senior managers of the involved companies met a number of times and held phone conversations as well. From 2004 to 2011, the truck firms’ German subsidiaries organised the cartel through electronic exchange of information. Overall, the cartel had the following goals: i) to co-ordinate prices at “gross list” for medium and heavy trucks in the European Economic Area (EEA); ii) to co-ordinate the timing for the introduction of emission technologies in compliance with the European emission standards; and iii) to pass on to consumers the costs associated with the adoption of the emission technologies.

The investigation was initiated after MAN applied for leniency. As a result, the EC carried out raids in January 2011. Subsequently, in November 2014, the EC issued a statement of objections. In July 2016, MAN, DAF, Daimler, Iveco and Volvo/Renault accepted the settlement decision. The EC levied a fine of EUR 1 billion to Daimler, EUR 881 million to Scania, EUR 753 million to DAF, EUR 670 million to Volvo/Renault and EUR 495 million to Iveco.

German price-fixing cases

We review two recent cases (Bundeskartellamt, 2017a; Bundeskartellamt, 2017b) entailing price fixing between manufacturer and retailer. The first case refers to a decision of the Bundeskartellamt (German Competition Authority) to fine two companies operating in the clothing industry EUR 10.9 million. Although the German anti-cartel law allows for individual fines, no individual was fined in this case.

The two companies involved in the proceedings were the clothing manufacturer Wellensteyn International GmbH & Co. KG (Wellensteyn) and the retailer Peek & Cloppenburg KG, Düsseldorf (P&C Düsseldorf). P&C Düsseldorf implemented the collusive agreement started by Wellensteyn and, as a reward, it received merchandise return options.² Wellensteyn entered into agreements with retailers in Germany to maintain its minimum prices for products sold.

These agreements consisted of the prohibition of online sales and in the requirement for retailers to not reduce prices even at the end of the seasons. The manufacturer (Wellensteyn) monitored the prices charged by retailers to check whether they were adhering to the agreement. Moreover, retailers actively monitored competitors to ensure they were complying with the agreement. Complaints filed by some retailers prompted the antitrust investigation, which began in March 2013. The cartel was active between April 2008 and February 2013.

The second case involves vertical price-fixing in the sale of furniture. The German Competition Authority imposed a total fine of EUR 4.4 million on 5 manufacturers and 4 managers. As a result, this case is a good example where both corporate and individual fines were applied. The companies fined were Hülsta-Werke Hülst GmbH & Co. KG, Stadtlohn, Rolf Benz AG & Co. KG, Nagold, Heinz Kettler GmbH, Ense-Parsit, Aeris GmbH, Haar, and Zebra Nord GmbH, Hatten-Sandkrug.

The five manufacturers established agreements with their retailers to maintain minimum prices as well as discount ranges to final consumers. The agreements dictated retailers to charge prices equal or greater than the recommended manufacturer price and included a monitoring system of retailers' pricing strategies and a punishment system whereby the manufacturer would refuse to supply non-compliant retailers. For discretionary reasons, the Bundeskartellamt did not prosecute the retailers.

French cartel case

On 20 September 2010, the Antitrust Authority fined banks and financial institutions for violating Article 101 of the TFEU (Autorité de la concurrence, 2010). More precisely, they convicted these firms for colluding on interbank fees for cheque image exchanges and collecting those fees between 1 January 2002 and 1 July 2007. In particular, the authority imposed the following fines:

- Banque de France (French Central Bank), EUR 346 500 on the first breach and EUR 3 500 on the second.
- BPCE (resulting from a merger between the Banques Populaires (BP) and the Caisses d'Épargne (CE)) assuming the rights and obligations of BP Participations, EUR 37.7 million on the first breach and EUR 380 000 on the second.
- BPCE assuming the rights and obligations of CE Participations, EUR 52.3 million on the first breach and EUR 530 000 on the second.
- La Banque Postale, EUR 32.5 million on the first breach and EUR 330 000 on the second.
- BNP-Paribas, EUR 62.7 million on the first breach and EUR 630 000 on the second.
- Crédit Agricole, EUR 82.1 million on the first breach and EUR 830 000 on the second.
- Crédit Mutuel, EUR 2.9 million on the first breach and EUR 30 000 on the second.
- Crédit du Nord, EUR 6.9 million on the first breach and EUR 70 000 on the second.
- Crédit Industriel et Commercial, EUR 20.9 million on the first breach and EUR 210 000 on the second.
- LCL (former Crédit Lyonnais), a subsidiary of Crédit Agricole, EUR 20.7 million on the first breach and EUR 210 000 on the second, for which Crédit Agricole is held responsible for EUR 15 million on the first breach and EUR 152 000 on the second.
- HSBC, EUR 8.9 million on the first breach and EUR 90 000 on the second.
- Société Générale, EUR 52.9 million on the first breach and EUR 530 000 on the second.

UK cartel case

At the end of March 2016, the Competition and Markets Authority (CMA) opened an investigation into two segments of the market for furniture: supply of drawer wraps and drawer fronts at the manufacturers' level (CMA, 2016). Regarding the first product (drawer wraps), three companies, Thomas Armstrong (Timber) Ltd, Hoffman Thornwood Ltd and BHK (UK) Ltd have been found guilty of cartel arrangements. As BHK has applied for leniency, the CMA did not impose a fine on this firm, conditional on its continued co-operation with the investigation.

In what concerns the second product (drawer fronts), Thomas Armstrong (Timber) Ltd and Hoffman Thornwood Ltd have admitted their role in two other illicit cartels: one between 2006 and 2008, and the other between 2006 and 2008 and in 2011, to share the market and co-ordinate pricing strategies through bid-rigging, as well as share confidential and competitively sensitive information. The CMA imposed a total fine on the two companies of GBP 2.8 million.

Canada and the United States

The Antitrust Division of the DOJ has levied average annual corporate fines of USD 28 million in the early 1990s. Since 1994, average annual fines have increased dramatically to over USD 300 million, to reach USD 560 million in the period 2005-07. The average fine per firm has also increased significantly. In addition, individual sanctions – both fines and imprisonment – form an important part of the Antitrust Division's punishment for cartel activity. Over the period from 1990 to 2007, the Antitrust Division of DOJ imposed jail sentences on 284 individuals (Shaffer et al., 2015; Connor, 2008b), with a mean of 24 per year over the period 2005-07 (Connor, 2008b); whereas, the mean annual number of fined individuals is 26.6 during the period of 1990 to 2006 (Connor, 2008b). In addition, individuals were fined an average of USD 147 100 (Connor, 2008b).

In 1889, Canada was the first country in the world to introduce the criminalisation of cartels. Nevertheless, defendants usually receive a suspended sentence. Therefore, the enforcement of individual sanctions for cartel offences has been “softer” than in the US. For instance, from 1998 to 2008, 11 individuals were convicted, nine of which paid fines.

Literature review

The literature on the effect of antitrust jurisdictions on the likelihood of collusion is very recent.

Theoretical literature

Dargaud et al. (2013) develop an interesting theoretical analysis comparing two types of fines: a profit-based fine (like that used in most EU jurisdictions) and a delegation-based fine that targets the manager in a more direct way (as the one used in the USA). They find that a delegation-based fine is output-distortive but is more effective in deterring cartels. Their paper differs from our theoretical approach (described below) in three dimensions:

1. They assume Cournot competition whilst we assume Bertrand competition.
2. They focus on a duopoly whilst we focus on an oligopoly with N firms.
3. The model is framed within the managerial delegation literature of Vickers (1995) and Fershtman and Judd (1987).

Within this literature, the manager is assumed to maximise a combination of firm's profits and firm's output, where the latter crucially depends on the powers delegated on the manager by shareholders (indirectly, by maximising this combination of profits and output, the manager is also maximising his own remuneration).

By contrast, in our approach, the (risk averse) manager maximises his (expected) utility, which depends on his remuneration consisting of a fixed and a variable (revenue-dependent) component. In other words, in Dargaud et al. (2013), the profit-based fine directly reduces firm's profits and appears as a fixed cost in the manager's objective function; by contrast, the delegation-based fine is a function of output and, not surprisingly, has output-distortive effects under collusion. In our case, as the manager is assumed to maximise his expected utility, the equivalents (in our setting) to a profit- or delegation-based fine directly affect the manager's remuneration.³

In addition, Aubert (2009) addresses the interplay between managerial incentives to collude or compete and incentives to exert effort. The incentives to collude introduce distortions in managers' effort, leading to internal firm inefficiency. In a model where managers privately choose market conduct (competition or collusion) and level of effort, high individual sanctions (monetary and imprisonment) for cartel behaviour deter collusion and incentivise deviations. Jail sentences are more efficient than individual monetary fines.

A recent body of the literature draws attention to the way in which fines can be calculated and its distortionary economic effects as well as their deterrence power. Dargaud et al. (2016) contribute to this literature by providing a theoretical analysis of the distortionary effects of two types of fines: profit-based and damage-based. Katsoulacos et al. (2015) also contribute to this discussion by providing a theoretical comparison of a fine based on the cartel overcharge with three other fine regimes: fixed fines, revenue-based fines and profit-based fines. Their analysis clearly lends support to an overcharge-based fine. Furthermore, Bageri et al. (2013) focus on distortions introduced by current methods of computing fines, whereas Katsoulacos and Ulph (2013) theoretically derive the optimal fine and study the impact on deterrence.

Our theoretical and experimental approaches involve shareholders who must design and offer wage contracts (with a fixed and a variable component, that is, with an underlying revenue-sharing mechanism) to managers. As such, our paper also contributes to the principal-agent literature. Related theoretical contributions in the economics literature involve revenue-sharing contracts in the video rental industry (Dana and Spier, 2001) and gate revenue sharing in team sports (Szymanski and Késenne, 2004). In addition, Wu (2017) theoretically explores the relationship and interplay between contracts that incentivise managerial effort the allocation of talent.⁴ The organisation theory literature studies revenue-sharing contracts in various professional services such as law, accounting, management consulting, advertising and architecture firms (Greenwood and Empson, 2003).

Empirical literature

The empirical literature devotes attention to revenue-sharing contracts in agricultural sharecropping (Allen and Lueck, 1992), flexible compensation which can serve as a device for risk-sharing as well as affect the financial structure of firms by encouraging stockholders to issue a greater amount of debt (Ichino, 1994) and box-office revenue sharing contracts as a device for flexible movie pricing in Spain's distribution market (Gil and Lafontaine, 2012).

Experimental literature

Regarding the design of contracts, Anderhub et al. (2002) analyse a menu of contracts including fixed and flexible wage for studying incentive compatibility along with “fair sharing” and reciprocity. Fehr et al. (2007) study behaviour within a principal-agent framework where principals can choose among an incentive contract with enforceable monitoring, a bonus contract without fining and a trust contract. They find that about 90% of principals prefer to choose the bonus contract, in contrast to the standard theories of efficiency; also, effort exerted by the agents and the average payoff for both principals and agents is higher than in the incentive contract with monitoring.

The authors interpret these results as subjects caring about fairness and opposing situations of distrust and hostility created by incentive contracts with fining mechanisms. Moreover, Karakostas et al. (2017) study efficiency and fairness building on the work of Anderhub et al. (2002) and Fehr et al. (2007). They consider a revenue-sharing contract, a bonus contract and a trust contract among which principals can choose from. The vast majority of experimental subjects choose the revenue sharing contract and evidence suggests that such contract results in greater effort and a fairer distribution of profits, on average, than the bonus contract.

The experimental approach of our paper also extends the recent and relatively small experimental literature on cartels.⁵ Cooper and Kühn (2014) study three forms of communication and analyse which of these aids collusion in duopoly experiments. One of their key findings is that the type of communication involving contingencies facilitates collusion to a greater extent. In addition, Fonseca and Normann (2012) compare experimental Bertrand oligopoly markets with a varying number of firms and with and without communication, investigating how the number of firms affects collusion. As theory suggests, they find that the higher the price, the lower is the number of firms, both with and without communication. More interestingly, they find evidence that gains from communication are non-monotonic in the number of firms. Fonseca and Normann (2014), in a Bertrand-oligopoly experiment with endogenous cartel formation, find evidence that quadropolies form cartels more often than duopolies because of a smaller hysteresis gain if the cartel breaks down.

Among experimental contributions in the area of leniency programmes and cartel deterrence, Apesteguia et al. (2007), Hinloopen and Soetevent (2008), Hamaguchi et al. (2009) and Bigoni et al. (2012) are noteworthy. These experimental contributions assess the overall effectiveness and impact of leniency programmes on cartel activity. In particular, Bigoni et al. (2012) design an experiment for evaluating several antitrust policies: fines, leniency and rewards compared to a benchmark treatment with no antitrust law that the authors call “laissez-faire”. They find that, on the one hand, the presence of fines decreases the emergence of cartels and, therefore, displays substantial deterrence power; on the other hand, the existence of fines results in higher prices in comparison to cartels formed under “laissez-faire”, because cartel members appear to use the fine as a device for costly punishments. The treatment with leniency results in lower average cartel’s prices than antitrust without leniency mainly due to a smaller number of formed cartels; however, the authors do not find evidence of lower prices than those under the laissez-faire treatment, primarily because cartels formed under leniency appear to be more stable than cartels formed under laissez-faire. Under the rewards policy, the authors find evidence of prices declining to the competitive level.

Policy literature

Wils (2002; 2005) discusses individual sanctions and the criminalisation of cartels with reference to the EU context, concluding that imprisonment can lead to increase cartel deterrence. In addition, Bartalevich (2014) offers a comparative analysis of EU and US antitrust policy. Buccirossi and Spagnolo (2007) argue that although the introduction of individual criminal penalties can improve deterrence, well-designed and well-implemented leniency programmes can attain the same goal.

Methodology

Theoretical model

This section is based on a theoretical model developed by the authors and which we describe herein as simple a manner as possible, avoiding, to the extent possible, mathematical notation and highlighting the underlying economic intuition. This theoretical model was used to “calibrate” the experimental parameters, that is, the specific design of each of the experimental treatments (described below). In particular, the examples presented in this section borrow heavily from the instructions presented to subjects in the experiment and are particularly helpful to describe how the relatively complex setup that forms the basis of our theoretical model was conveyed to experimental subjects who, in most cases, have no economics background.

Market description

Consider an oligopolistic market with $n > 1$ firms selling homogeneous goods and K identical consumers. Each consumer wants to buy exactly one unit of the good. As goods are homogeneous, consumers will buy the good from the firm(s) charging the lowest price (as long as this price does not exceed the consumers’ reservation price – the maximum price that they would be willing to pay). Given the prices, there are two possible market configurations (see Box 4.2):

1. If one firm charges a price $\check{p}$ below the prices charged by all the remaining firms, this firm supplies the entire market (as there are no capacity constraints) and rival firms sell nothing. The revenue of this firm is $\check{p} \times K$ and the revenue of the remaining firms is zero.
2. If there are $m \leq n$ firms charging the same price $\check{p}$ and this price is lower than the price set by the remaining $n - m$ firms, the m firms will equally split the K consumers. The revenue of each of the m firms is $\check{p} \times (K/m)$ and the revenue of the remaining firms is zero.

Suppose that firms interact for an infinite number of periods and discount future profits at a constant (and equal across firms) rate. The choice of the price charged by each firm in each period is the responsibility of its manager. Managers are assumed to be symmetric, risk-averse and their utility only depends on their wages. Wages have two components:

1. A fixed component, which the manager receives independently of the firm’s revenues in each period.
2. A variable component, which corresponds to a percentage of the firms’ revenues in each period.⁶

Box 4.2. Example 1

Consider a market with $n = 4$ firms (which we denote by A, B, C and D) and $K = 24$ consumers that pay, at most, EUR 10 for each unit of the good.

1. Suppose that Firm A sets a price of EUR 8.5, Firm B chooses a price of EUR 7.5, Firm C sets a price of EUR 8.8 and Firm D sets a price of EUR 7.6. As Firm B sets the lowest price, it sells all 24 units (at a price of EUR 7.5) and makes revenues of EUR 180 ($= 24 \times 7.5$). Firms A, C and D do not supply any customer and, therefore, make EUR 0 in revenues.
2. Suppose now that Firm A and Firm B both set a price of EUR 7, Firm C sets a price of EUR 8.8 and Firm D sets a price of EUR 7.4. As Firms A and B set the same (lowest) price, they share the customers equally. Hence, both firms sell 12 units at a price of EUR 7 each unit, each generating revenues of EUR 84 ($= 12 \times 7$). As Firms C and D do not supply any customer, they make EUR 0 in revenues.

There are no costs of producing the good. Thus, in each period, the only cost for shareholders is the wage paid to managers. Consequently, the shareholders' profit is the difference between the firm's revenue and the manager's wage.

For illustrative purposes, let us go back to Example 1 (Box 4.2) and see how managers' payoffs would be calculated in that case.

Box 4.3. Example 2

Suppose that the fixed wage component is equal to EUR 30 and the variable component is 10% of the firm's revenue.

1. The managers of Firms A, C and D only receive the fixed component of their wage (as their firms make no revenues). Therefore, the wage of managers of Firms A, C and D is EUR 30 and the shareholders of these firms make losses – their payoff is -EUR 30. In contrast, the manager of Firm B receives the fixed component (EUR 30) plus 10% of his firm's revenue (EUR 180). Therefore, his wage is equal to EUR 48 ($= \text{EUR } 30 + \text{EUR } 18$) and the shareholders of Firm B receive EUR 132 ($= \text{EUR } 180 - \text{EUR } 48$).
2. The managers of Firms C and D only receive the fixed component (as their firms make no revenues), i.e. their wage is EUR 30. The shareholders of these firms make losses: their payoff is -EUR 30. In contrast, the managers of Firms A and B receive the fixed component (EUR 30) plus 10% of the revenues of their firms (EUR 84). Thus, their wage is EUR 38.4 ($= \text{EUR } 30 + \text{EUR } 8.4$) and the shareholders of (each of) these firms receive EUR 45.6 ($= \text{EUR } 84 - \text{EUR } 38.4$).

Assume that the two types of players in the market – shareholders and managers – make their decisions according to the following timing:

1. Each firm's shareholders offer his manager a contract with a wage that contains a fixed as well as a variable component.⁷
2. Each manager chooses whether or not to accept the contract offered by the firm's shareholders. If the manager accepts the contract, he is the sole person responsible for setting the price of his firm in all periods. Otherwise, if he rejects the contract, he will receive a (fixed) income in all periods, which we denote by w^u and refer to as "outside option income".⁸

As the market duration is assumed to be infinite, managers will choose the price in each period in a way that maximises their discounted utility, given by the discounted sum of their (expected) wage in each period.

An important assumption we make – the reasoning for which becomes clear upon carefully analysing Examples 1 and 2 – is that shareholders themselves do not have an outside option. As we mentioned above, managers can choose not to accept the contract and receive the outside option income. However, we explicitly assume that shareholders cannot (or would not) design contracts that lead to the manager not accepting it, in which case the firm could not operate as it would have no manager.⁹

Competitive benchmark

Let us characterise the market outcome when managers set prices non-cooperatively. This is the equilibrium counterfactual, which would be observed if managers did not co-ordinate prices.

As is well known in the literature, if products are homogeneous and firms compete in prices, the non-co-operative equilibrium is a repetition of the Bertrand-Nash equilibrium of the stage game in all periods. Thus, in each period, the price will be equal to the marginal production cost (which is assumed equal to zero). If the contract has a non-zero fixed component, shareholders will incur losses in each period, because they will receive no revenues (as the price is equal to the marginal cost) but must pay the fixed wage component.

In this context, the key question we wish to address is: what contract will shareholders (optimally) offer managers? That is, anticipating how a firm's manager will, in equilibrium, set the firm's price, what contract would shareholders choose to offer to managers so as to maximise their payoff? In our competitive benchmark, the answer to this question is:

Result 1. Suppose that managers set prices non-co-operatively. Any contract with a fixed component equal to the outside option income is optimal (regardless of the variable component).

Let us understand the economic intuition behind this result. As shareholders aim at maximising profits (or, in this case, minimising losses), they want to set the lowest possible value for the fixed component. However, in order for managers to accept the contract, they must receive a discounted utility greater than or equal to the one they would receive if they rejected the contract (and received the outside option income in each period). It follows, therefore, that shareholders will offer a fixed wage component equal to the outside option income (w^u). Finally, as the per-period equilibrium revenues

are zero, the discounted utility for managers and shareholders' profits do not depend on the value of the variable wage component.

Collusion

Looking at the competitive benchmark, it becomes clear that shareholders would strictly prefer that managers set their prices co-operatively if the variable wage component is relatively low.¹⁰ Let us then determine which contracts shareholders may offer to provide incentives for their managers to set prices co-operatively.

To ensure the tractability of the model, we focus on collusive agreements that involve the managers of all firms (full collusion). We will also focus on agreements aiming at co-ordinating all prices at the monopoly price level, i.e. the price that would prevail if there were a single firm active in the market (perfect collusion). It follows straightforwardly that the maximum price that managers may set is the consumers' reservation price, which we denote by $\bar{p}$. If all managers agree to set this price, each manager has the temptation to unilaterally deviate and set a price slightly below $\bar{p}$. By doing so, the manager can supply the entire market (instead of sharing it equally with the other managers). Thus, for the agreement to be sustainable there must exist a credible punishment for managers that disrupt it. We assume the simplest punishment scheme: if a manager deviates from the agreement, it becomes permanently broken (i.e. there is no possibility for renegotiation after a deviation). In other words, after a deviation, managers will set prices equal to the marginal cost (which means zero revenues for the firms) forever and essentially revert to the competitive benchmark.

In order to obtain a price co-ordination outcome, shareholders must offer wage contracts such that:

- Managers are willing to accept, i.e. their discounted utility must be greater than (or equal to) the outside option income in all periods.
- Once the collusive agreement has been established, managers do not have unilateral incentives to disrupt the agreement (by slightly undercutting the price).
- Shareholders must prefer the price co-ordination outcome to the competitive benchmark, that is, shareholders' payoffs must be higher when managers abide by the collusive agreement than when they set their prices non-co-operatively.

Shareholders will offer a contract that maximises their (discounted) profit subject to the above constraints. It is straightforward that the shareholders' utility is decreasing in the fixed and variable component of the manager's wage. Thus, shareholders will offer the lowest values for these components that the manager is willing to accept, i.e. in the limit, the fixed and variable components that leave the manager indifferent between accepting the contract and rejecting it. Given a contract, i.e. a combination of a fixed and a variable component, managers will accept it as long as their expected discounted utility is not lower than their discounted utility when they receive the outside option income in all periods.

It follows, therefore, that any contract that provides the same expected utility as the outside option income allows shareholders to extract the highest possible surplus from their managers (and still ensure contract acceptance). However, the terms of the contract, i.e. the way the wage is distributed between the fixed and variable components, affects the incentives for managers to co-ordinate prices.

As we have discussed above, our interest is in understanding how different antitrust regimes affect economic agents' choices. Specifically, we will analyse how two different cartel fine policies affect the optimal contracts offered to managers. These two fine policies are inspired by the corporate fines regime that was observed until recently in most EU jurisdictions and on the individual fines regime that is observed in the US (in addition to its corporate fines). We do not examine the effect of the associated risk of imprisonment.

To make the two scenarios comparable, we will make similar assumptions. In both scenarios, we will assume that if the cartel is detected by the antitrust authority, managers may restart the agreement in the following period. In addition, there will be no aggravated fine if the cartel is caught again in a future period (for recidivism) and the probability of the cartel being caught is also not affected.¹¹ We also assume that if the market is audited, enough evidence is collected for the antitrust authority to be able to convict the firms or managers in case of price co-ordination. This is modelled by assuming that with a given probability, the antitrust authority audits a market, detects the cartel if it is active and imposes a fine (on the firm or on the manager, depending on the antitrust regime). Deviators are assumed to also pay the fine if the market is audited and the cartel is detected.¹²

Corporate fines

For simplicity, under a corporate fine regime, we assume that cartel behaviour (if detected) is punished with a fine that targets firms, in the form of a percentage of its revenues. Also, for simplicity, this fine is assumed not to depend on the cartel duration. Suppose that, in a given period, the market is audited: if the managers are co-ordinating prices, a fine equal to a percentage of the firm's revenues is imposed in all firms;¹³ if managers are not co-ordinating prices, then no fine has to be paid. Notice that if the variable wage component is positive, managers will be indirectly affected by the fine, insofar as they will receive a share of a lower "pie" (because firms' net revenues will decrease in the amount of the corporate fine).

Box 4.4. Example 3

Suppose that, in every period, there is a 20% chance that the market will be audited. If the market is audited and managers are co-ordinating prices, the antitrust authority imposes a fine of 35% of the revenues of all firms. The managers' contract is composed of a fixed component of EUR 30 plus 10% of the firms' revenues (variable component). If all managers co-ordinate on the maximum price (EUR 10):

1. If the market is not audited, the revenues of each firm are EUR 60 ($= 24 \times \text{EUR } 10/4$), the managers' wage is EUR 36 ($= \text{EUR } 30 + 10\% \times \text{EUR } 60$) and each shareholder gets EUR 24 ($= \text{EUR } 60 - \text{EUR } 36$).
2. If the market is audited, the revenues of each firm are EUR 39 ($= (100\% - 35\%) \times \text{EUR } 60$), the managers' wage is EUR 33.9 ($= \text{EUR } 30 + 10\% \times \text{EUR } 39$) and each shareholder gets EUR 5.1 ($= \text{EUR } 39 - \text{EUR } 33.9$).

As the market is audited with a probability of 20%, the expected wage of each manager is EUR 35.58 ($= 20\% \times \text{EUR } 33.9 + 80\% \times \text{EUR } 36$) and the expected profit of each shareholder is EUR 20.22 ($= 20\% \times \text{EUR } 5.1 + 80\% \times \text{EUR } 24$).

As explained above, managers will only collude if their expected discounted utility when co-ordinating prices is greater than (or equal to) their expected discounted utility if they unilaterally deviate (i.e. if they slightly undercut the price to supply the whole market). An increase in the severity of the fine or in the probability of cartel detection has two countervailing effects on the incentives for managers to collude:

- On the one hand, the managers' expected gains from collusion decrease because, with some probability, the firms' expected revenues will be lower and, therefore, the managers' variable wage component will be lower. This decreases the incentives for managers to collude.
- On the other hand, as deviations are still punished, the gain from a defection also decreases. This increases the incentives for managers to stick to the agreement.

In our theoretical model, once these two effects are taken into account, we obtain the following result:

Result 2. Under a corporate fine regime, shareholders will optimally offer fixed-wage contracts equal to the outside option income.

As shareholders are risk-neutral and managers are risk-averse, it is optimal to have all the risk on the shareholders' side. Shareholders achieve such an outcome by offering a purely fixed contract (i.e. without a variable component). As shareholders have all the bargaining power when designing the contract, they will be able to extract all the managers' surplus, i.e. they can offer wage contracts that leave managers with no surplus over their outside option income. With such a contract, managers will be indifferent between accepting and rejecting the contract (since they will receive the same utility) and, if accepting the contract, they will be indifferent between colluding or not (since they will receive the same utility regardless of the revenues of the firm).

Individual fines

Under an individual fine regime, such as in the US, cartel behaviour is punished with individual sanctions on the manager, as it is considered a criminal offence.¹⁴ Following Dargaud et al. (2013), we assume that this sanction is equivalent to a monetary reduction of the manager's remuneration, e.g. the reputational damage caused by the fine.

We assume that the fine is a percentage of the manager's wage and is independent of the cartel duration. Thus, with a positive probability, the market is not audited, and managers receive their (full) wage; but with a positive probability, the market is audited and if managers are co-ordinating prices, the cartel is detected and managers have to pay a fine (i.e. they receive a percentage lower than 100% of their wages).

As in the case of corporate fines, we assume that, if the cartel is detected, managers may restart the agreement in the following period. Furthermore, if managers disrupt the collusive agreement, they will have a short-run deviation gain but, after the deviation, they will revert to the competitive benchmark and permanently set a price equal to the marginal cost (which means zero revenues for firms).

Result 3. Under an individual fine regime, no fixed-wage contract provides incentives for managers to co-ordinate prices.

Box 4.5. Example 4

Suppose that, in every period, the market is audited with 20% probability. If the market is audited and managers are co-ordinating prices, all managers must pay a fine that corresponds to 10% of their wages. As in Example 3, the managers' contract is composed of a fixed component of EUR 30 plus 10% of the firms' revenues. As seen in Example 3, if all managers co-ordinate on a price of EUR 10, the revenues of each firm will be EUR 60. Thus:

1. If the market is not audited, the manager's wage is EUR 36 ($= EUR\ 30 + 10\% \times EUR\ 60$) and each shareholder receives a profit of EUR 24 ($= EUR\ 60 - EUR\ 36$).
2. If the market is audited, the manager's wage is EUR 32.4 ($= 90\% \times EUR\ 36$), each manager pays a fine equal to EUR 3.6 ($= 10\% \times EUR\ 36$) and each shareholder receives a profit of EUR 24 ($= EUR\ 60 - EUR\ 36$).

As the cartel is detected market with probability 20%, the expected wage of each manager is EUR 35.28 ($= 80\% \times EUR\ 36 + 20\% \times EUR\ 32.4$). Notice that shareholders bear no risk, as their individual profit is EUR 24, regardless of the market being audited or not.

Let us understand the intuition behind this result. Suppose that shareholders offer a fixed wage $\tilde{w}$ to the managers (i.e. managers receive the same wage regardless of their firm's revenues) and managers agree to co-ordinate prices. In periods where the market is not audited, managers will receive $\tilde{w}$; but in periods where the market is audited, managers receive only a fraction of $\tilde{w}$ (as the fine falls on the manager). Suppose now that a manager unilaterally deviates from the agreement. In the deviating period, the manager receives $\tilde{w}$ if the market is not audited, and a fraction of $\tilde{w}$ if the market is audited. However, in all periods following the deviation, the manager will receive $\tilde{w}$ for sure (since managers will permanently revert to competition and, therefore, no further fines will need to be paid). Therefore, the expected discounted utility of a manager is surely higher if he deviates than if he sticks to the agreement. In sum, if managers receive a fixed-wage they will surely not abide by the collusive agreement.

Notice that Result 3 is in sharp contrast to the result in the case of corporate fines (Result 2). This occurs because, under corporate fines, the "costs" of the cartel are on the shareholders' side. Thus, if the manager receives a fixed salary, his salary is the same regardless of whether there is price co-ordination (and the levying of corporate fines).

We also obtain the following result:

Result 4. Under an individual fine regime:

1. If the discount factor is sufficiently high, any purely-variable contract (i.e. without fixed component) provides incentives for managers to collude. As a result, the optimal contract is the purely-variable contract that leaves managers with the same expected discounted utility as if they rejected the contract (and received the outside option income in each period).

2. For lower values of the discount factor, the optimal contract must have strictly positive fixed and variable components. Among all possible contracts with positive fixed and variable components, shareholders would like to offer the contract that makes collusion among managers more likely.

Let us grasp the intuition behind Result 4. Under an individual fines regime, a wage contract without variable component does not provide incentives for managers to collude (Result 3). Thus, the question is whether a purely variable contract is optimal or not. If the discount factor is low, a contract without fixed component is not optimal because managers do not place much value in future payoffs and will not resist the temptation of disrupting the agreement to (significantly) increase the revenue of their firms in the present (and, therefore, their wage). As a result, when the discount factor is low, the contract must have both variable and fixed components strictly positive. If the discount factor is high, this temptation for unilateral deviation vanishes.

Experimental design

To answer our research questions, we develop and implement a computerised experiment involving experimental four-firm markets, where subjects play a repeated Bertrand game with inelastic demand and identical firms selling a homogenous product. We choose experimental markets with four firms because for such markets we have experimental evidence of having greater gains and higher incentives from explicit collusion (Fonseca and Normann, 2012; 2014). The marginal cost is assumed to be 0 for simplicity, and we assume a mass of simulated consumers $M = 24$ with reservation price equal to 10. Consequently, subjects (firms) can charge a price between zero and ten. The firm charging the lowest price supplies the entire market; in case of ties, firms split the revenue evenly.

Furthermore, as subjects have the choice whether to start communicating via a chat messenger, we are able to endogenise cartel formation as in Fonseca and Normann (2014). Endogenising cartel formation allows us to study cartel prevalence, cartel effectiveness and cartel prices, which is one of the goals of this paper. We also assume the presence of an antitrust authority monitoring each market. Following Fonseca and Normann (2014) and various experiments on leniency programmes (e.g. Apesteguia et al., 2007; Hinloopen and Soetevent, 2008; Bigoni et al., 2012), the antitrust authority monitors each market with some probability, assumed to be $\tau = 20\%$, in order to detect cartel behaviour. If a market is audited, we assume that a cartel, if active, is detected and the antitrust authority will levy a punishment which differs across jurisdictional frameworks. The cartel is formed only when all the four firms in a given market decide to communicate.

Our experiment consists of four treatments across two phases of a multi-day experiment. The four treatments are:

1. a baseline treatment without communication, NoChat
2. a treatment where firms may decide to form cartels without the threat of punishment, ChatNoLaw
3. a treatment where individual fines are levied in case a cartel is detected, inspired by the US jurisdiction, ChatUS
4. a treatment where corporate fines are levied in case of cartel detection, ChatEU.

In Phase 1, subjects play the role of managers in an indefinitely repeated Bertrand market with four firms. Following Dal Bó and Fréchette (2011), we simulate an indefinitely repeated market by imposing a 10% probability it will terminate at the end of each experimental period; the expected duration is therefore 10 experimental periods. This is a sufficiently long expected duration to allow for collusion to take place while keeping the expected total number of periods in the session as a whole (i.e. 40) low enough not to induce subject fatigue, which may compromise decision quality. Once a market ends, the software randomly rematches participants randomly into new sets of four and a new market starts. We do this four times to obtain four markets per session.¹⁵

At the start of the first period of each market, subjects are told the parameters of their contract, which determine their payoff. We implement two wage contracts, which differ in the revenue component. The two contracts are $X = (30; 0.2)$ and $Y = (30; 0.02)$, which we refer to respectively, as “high-powered” and “low-powered”. The rationale for the choice of these contract parameters is provided below.

To control for the possibility that the order in which we implement contracts may affect subjects’ decision-making, we implement two orderings at the session level: $XYXY$ and $YXYX$. Subjects were compensated for the total earnings in two of the four markets in which they took part.

After the fourth market ended (i.e. at the end of Phase 1), subjects were told they would then take the role of firm owners. Their task would be to choose which of the two contracts they had just experienced as managers they would like to offer to a future manager of their firm. Their firm would exist in a future session of the experiment (Phase 2). Their earnings as firm owners would be equal to the revenue of the firm minus the manager salary minus any potential fine.

Ultimately, the main purpose of Phase 1 was to generate contract choices. Subjects played the role of managers in Phase 1 so that they had sufficient experience of the markets to make a better-informed choice of contracts when in the role of firm owner. We ran 6 sessions with 12 subjects for each of the 4 treatments, totalising 288 subjects and therefore generating 288 contract choices.

In Phase 2, subjects only played the role of managers, who were told that their contract was chosen by the firm owner, who was a participant in an earlier session. Every manager in a market in Phase 2 had the same contract. Managers were also told that the profits resulting from their pricing decisions would go to the firm owner. As in Phase 1, subjects took part in four indefinitely repeated markets with a continuation probability of 10%. In each market, they were assigned to a different firm owner, such that they were exposed to both types of contracts. We also implemented two orderings at the session level: $XYXY$ and $YXYX$. In this second phase, we ran three sessions of each ordering type, with 12 subjects per session for each of the 4 treatments, for a total of 288 subjects.

The fine in both Phase 1 and Phase 2 markets was set at 35% of the firm’s revenue in the ChatEU treatment and 35% of a manager’s salary in the ChatUS treatment.

The sessions took place in the FEELE lab at the University of Exeter Business School between February and March 2018. The software used to run the experiment was z-Tree (Fischbacher, 2007). We recruited participants from a pool of volunteers using the FEELE lab’s ORSEE system (Greiner, 2015). Sessions lasted on average one hour. The average earnings were GBP 13.10 (min: GBP 9.00, max: GBP 19.00).

Results and discussion

Introductory note and test hypotheses

As mentioned above, each of the four treatments – NoChat, ChatNoLaw, ChatUS and ChatEU – was carefully designed, building on the results from our theoretical approach.

In the context of our research Goal 1 (“Do antitrust law regimes involving individual fines have larger cartel deterrence power than antitrust regimes encompassing only corporate fines?”), the main theoretical conclusion we arrive at is that *for a given contract that a manager chooses to accept*, it is indeed true that a corporate fine regime generates stronger incentives for cartel formation than an individual fine regime. That does not necessarily imply that cartels are expected to emerge for that given contract: the answer to this crucially depends on wage levels and on their impact in the manager’s incentive compatibility constraint – that is, on the difference between the expected (discounted) utility under a collusive outcome and the expected (discounted) utility in case of deviation.

This allows us to manipulate wage levels in such a way that a specific contract may generate incentives for collusion under one antitrust regime, but not under the other. Indeed, this is the underlying rationale for the wage level choices in the low- and high-powered contracts. Our expectation is that the low-powered contract generates incentives for collusion under a corporate fine regime (ChatEU) but not under an individual fine regime (ChatUS), because this contract is expected to satisfy the incentive compatibility constraint in the former, but not in the latter. By contrast, the high-powered contract was designed so that there are incentives for collusion in both regimes.¹⁶

Under the NoChat treatment, explicit cartel formation is not allowed, as there is no communication. By contrast, in the ChatNoLaw treatment, there is no punishment associated with cartels. Therefore, the incentive compatibility constraint is satisfied with both types of contract. Table 4.1 summarises these hypotheses.

Table 4.1. Summary of hypotheses

	NoChat	ChatNoLaw	ChatUS	ChatEU
Low powered contract	No Cartels	Cartels	No cartels	Cartels
High-powered contract	No Cartels	Cartels	Cartels	Cartels

Following on from this line of reasoning, our hypothesis on cartel prevalence may be further refined, in particular because the crucial element underlying cartel formation is the incentive compatibility constraint. But managers may be heterogeneous (e.g. with different risk attitudes or with different discount factors): this implies that cartel formation may not be observed in the data even when it is theoretically expected to emerge. In order to understand this, recall that for a cartel not to be formed, all that is required is for a single manager to express his unwillingness to engage in communication with other managers. Theoretically, this would occur if that manager’s incentive compatibility constraint was not satisfied, which in our setup may only happen for manager-specific reasons, e.g. that manager’s risk attitude and/or discount factor may be very different from that of the remaining managers. Assuming that such manager heterogeneity may exist, but assuming it to be “equal” across treatments, that is, assuming that the same population of heterogeneous managers needs to make cartel formation

decisions across treatments, we may conclude that cartels may not be formed when they are expected to be formed in Table 4.1. For instance, if one manager has a particularly low discount factor, he may “block” the emergence of (at least) one cartel in the ChatEU treatment. Naturally, that same manager with a particularly low discount factor would also block (at least) that cartel in the ChatUS treatment and, possibly, in the ChatNoLaw treatment. Therefore, building on the incentive compatibility constraint and on the possibility that managers are heterogeneous, we can refine Table 4.1 and posit the following hypotheses:

- *Hypothesis 1.* For a given contract (low-powered or high-powered), the number of cartels formed in the ChatNoLaw treatment should be higher than in the ChatEU treatment, which in turn should be higher than in the ChatUS treatment, and the latter should be higher than zero (the number of cartels formed, by definition, in the NoChat treatment).
- *Hypothesis 2.* In the ChatNoLaw, ChatUS and ChatEU treatments, the number of cartels formed should be higher with the high-powered contract than with the low-powered contract.

Regarding pricing behaviour, our theoretical framework is built upon the premise that managers, when they collude, choose the maximum price possible (“full collusion”). Indeed, taking into account that there is a punishment associated with cartels in the ChatEU and ChatUS treatments, it is perfectly logical that managers maximise their “upside”, that is, their collusive payoffs, because they face an expected cost. Similarly, in the ChatNoLaw treatment, expected utility maximisation dictates that managers should collude by choosing the maximum prices that consumers are willing to pay. Nevertheless, it may be that managers are able to co-ordinate on a different price level. The relevant aspect here is that the decision to form a cartel is heavily related to the incentive compatibility constraint, that is, the difference between the expected (discounted) payoffs of the collusive outcome and of a unilateral deviation. In short, if a cartel is formed, the incentive compatibility constraint must be satisfied. But in that case, it must also be true that colluding leads to higher payoffs than not colluding (and competing): the unilateral deviation payoffs are, in our theoretical approach, equal to those of Bertrand competition except for the deviation period. This means that if the incentive compatibility constraint is satisfied, then forming a cartel is preferable to Bertrand competition in the first place. Therefore:

- *Hypothesis 3.* Conditional on a cartel having been formed, managers should be able to co-ordinate on the same price level in the ChatNoLaw, ChatEU and ChatUS treatments, regardless of the contract type (low- or high-powered).

As mentioned above, at the end of Phase 1, managers were asked to play the role of firm owners and choose a particular contract that would be offered to managers in Phase 2. As firm owners, they would receive further payoffs depending on managers’ decisions in Phase 2. Looking at Table 4.1, it becomes relatively clear to outline our theoretical expectation for this choice. Clearly, from a theoretical viewpoint, we expect the manager incentive compatibility constraint to be satisfied in the ChatEU treatment regardless of the type of contract. Therefore, in that treatment, managers should have the incentives to collude, which implies that firm owners with a profit maximisation perspective should choose the low-powered contract – as it would be expected to achieve the same revenues but at a lower cost (the manager’s wage) to the firm owner. The same is true in the ChatNoLaw treatment. However, in the ChatUS treatment, the low-powered contract does not satisfy the incentive compatibility constraint. As such, in order to foster collusion,

firm owners would need to choose the high-powered contract. We summarise this in the following way:

- *Hypothesis 4.* Given a choice between the low- and high-powered contracts, firm owners should choose the low-powered contract in the ChatNoLaw and ChatEU treatments, but the high-powered contract in the ChatUS treatment. They should be indifferent between the two in the NoChat treatment.

In order to test these hypotheses, we use only Phase 2 data, wherein subjects (managers) are making price and cartel formation decisions based on contracts that were chosen by real firm owners. The pricing/communication stage in Phase 1 was designed only to provide experience to those subjects in order for them to make an informed decision regarding their contract choice as firm owners, so we will only use those outcomes as predictors of contract choices.

Cartel prevalence

We begin by looking at the frequency with which subjects agreed to form a cartel. We computed the total number of initiated cartels for each of the 216 four-firm markets in the ChatNoLaw, ChatEU and ChatUS treatments. Figure 4.1 displays the distribution of the number of unique cartels initiated during the lifetime of a market.

The NoChat and ChatNoLaw provide benchmark cases. In the former, cartels are never possible (therefore, zero cartels were formed during the lifetime of all markets), while in the latter there is no punishment associated with cartel formation (and, therefore, once a cartel is formed, it lasts until the market ends). As expected, all observations record either zero or exactly one cartel formed, in accordance with Table 4.1 and Hypothesis 1.

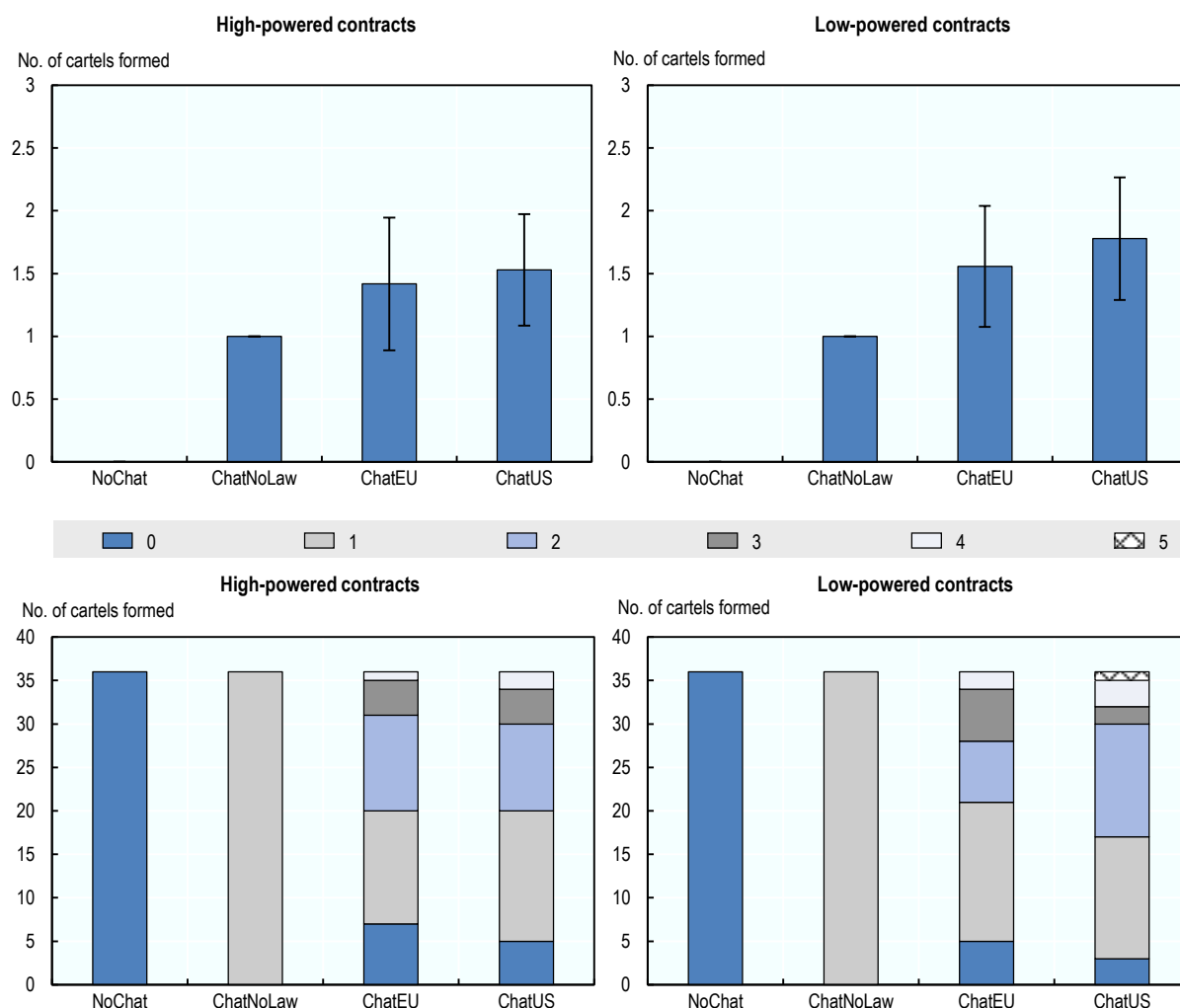
By contrast, in the ChatEU and ChatUS treatments, the antitrust authority may detect and break cartels (if formed), which means that more than one cartel may be observed over the lifetime of a market. Interestingly, in a significant percentage of markets, no cartels are formed in both treatments. Therefore, there would appear to be a significant deterrent effect from both individual and corporate fines. Overall, the distributions of cartel formation are very similar in the ChatEU and ChatUS treatments for both high-powered and low-powered contracts. To test for this, we regressed the number of cartels formed in a given market on a dummy for high-powered contracts, a dummy for ChatUS, as well as a set of market dummies to account for learning effects. We omit the NoChat and ChatNoLaw data since there is no variance in the outcome measure in these treatments. The results of the regression are summarised in Table 4.2.

There is no significant difference in the frequency of cartel formation between ChatEU and ChatUS either for low-powered ($(1,11) = 0.74, p = 0.407$) or high-powered contracts ($F(1,11) = 0.25, p = 0.630$). This means that Hypothesis 1 is not supported by the data in what concerns the ChatEU and ChatUS treatments. Also, there is no significant difference in the frequency of cartel formation between high- and low-powered contracts, either in ChatEU ($(1,11) = 0.33, p = 0.578$) or ChatUS ($F(1,11) = 0.17, p = 0.688$). This means that, in what concerns the ChatEU and ChatUS treatments, Hypothesis 2 is also not supported by the data.

Another way in which we can measure cartel prevalence is to quantify the extent to which cartels were active during the lifetime of a market. This analysis complements that of cartel formation, as it allows us to ascertain how long managers were prepared to wait before re-forming a cartel after it was detected. In the ChatNoLaw treatment, cartels were

active in 96% of periods in which it was possible to form a cartel. In the ChatEU and ChatUS treatments, cartels were active in 59% and 63% of possible cases.

Figure 4.1. Cartel formation frequency conditional on treatment and contract type



Notes: Top panel: average number of cartels formed (error bars are 95% robust confidence intervals). Bottom panel: distribution of cartels formed.

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

We regressed a variable equal to the proportion of periods where a cartel was active in a market on a dummy for high powered contracts (HighPowered), a set of treatment dummies (ChatEU, ChatUS) and their interaction with HighPowered, and a set of market dummies (Market2, Market 3, Market 4) that account for any learning across markets. Results from the regression are available in Table 4.2.

Table 4.2. Determinants of cartel formation coefficient and statistical significance (standard errors)

Dependent variable:	Cartels started	Active cartel
High-Powered Contract	-0.138 (0.242)	-0.017 (0.074)
ChatUS	0.222 (0.258)	0.011 (0.063)
High-Powered Contract x ChatUS	-0.111 (0.269)	0.043 (0.085)
Market #2	1.583*** (0.212)	0.349*** (0.064)
Market #3	0.917*** (0.158)	0.404*** (0.055)
Market #4	1.000*** (0.160)	0.343*** (0.054)
Constant	0.681*** (0.159)	0.233** (0.081)
R2	0.278	0.257
N	144	144

***: significant at the 1% level.

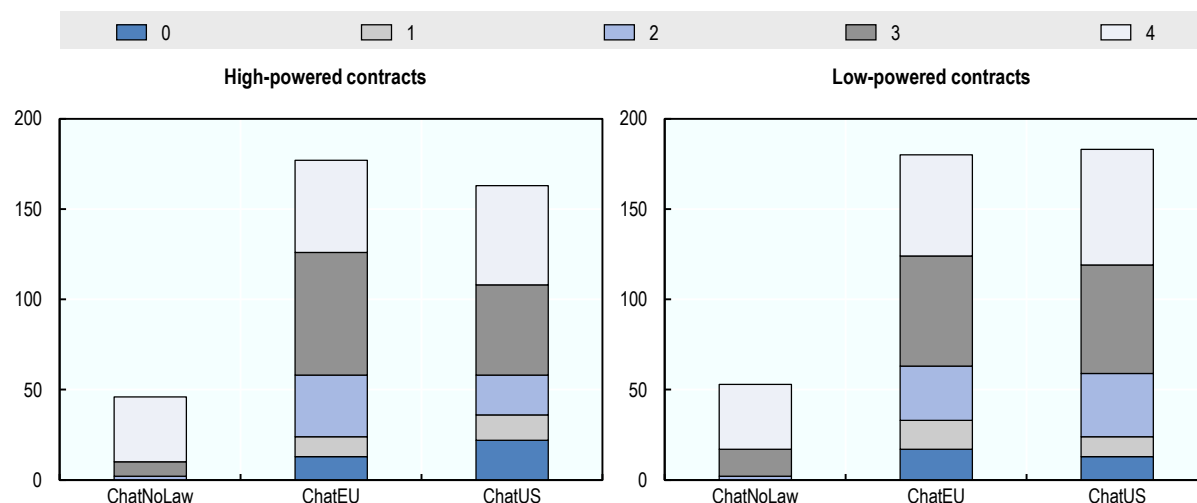
Note: Robust standard errors clustered at the session level in parentheses.

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

Cartels are not more prevalent in the ChatEU treatment than in ChatUS treatment, both for low-powered contracts ($(1,17) = 0.03, p = 0.862$) and high-powered contracts ($F(1,17) = 0.59, p = 0.451$). Again, the evidence does not appear to support Hypothesis 1 (relative to these 2 treatments). We also do not detect a difference in the prevalence of cartels as a function of manager contracts in the ChatNoLaw ($F(1,17) = 1.63, p = 0.219$), ChatEU ($F(1,17) = 0.21, p = 0.650$) or ChatUS ($F(1,17) = 0.04, p = 0.851$) treatments. This also implies Hypothesis 2 does not appear to be supported by the data.

We next document the voting decisions taken by managers whenever the opportunity to form a cartel arose. Remember that cartels were only formed if there was a unanimous decision by managers. Figure 4.2 displays the distribution of voting outcomes in each treatment.

In the ChatNoLaw treatment, the overwhelming proportion of outcomes was a unanimous agreement to form a cartel. Since there was no cartel detection, the number of votes is substantially lower than in the ChatEU and ChatUS treatments, as there were fewer opportunities to vote in favour of cartel formation (if formed, a cartel would only break down when the market ended). The distribution of voting outcomes in the ChatEU treatment is similar to its counterpart in the ChatUS treatment. We estimated random-effects models of the individual decision to vote in favour of forming a cartel whenever that vote was available to managers in a market – that is, excluding the very first period in a market, as well as any period in which the cartel was already active. The regressors are dummy variables for whether the cartel had been caught in the 4 previous periods ($t-1$, $t-2$, $t-3$ and $t-4$), a dummy for ChatEU, a dummy for high-powered contract and its interaction with the ChatEU dummy, as well as 3 market dummies. Results are presented in Table 4.3.

Figure 4.2. Distribution of votes in favour of forming a cartel

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

Table 4.3. Determinants of individual and aggregate votes for cartel formation coefficient and statistical significance (standard errors)

Dependent variable:	Individual vote	Total votes
Caught (t-1)	0.361*** (0.047)	1.771*** (0.182)
Caught (t-2)	0.261*** (0.044)	1.377*** (0.207)
Caught (t-3)	0.086** (0.038)	0.596*** (0.184)
Caught (t-4)	0.051 (0.038)	0.197 (0.159)
ChatEU	-0.011 (0.031)	-0.028 (0.212)
High Powered	-0.062*** (0.024)	-0.322** (0.127)
EU x High Powered	0.090** (0.036)	0.468 (0.285)
Period	-0.041*** (0.004)	0.014 (0.033)
Market 2	0.227*** (0.043)	-1.004** (0.391)
Market 3	0.610*** (0.050)	-1.625** (0.773)
Market 4	1.015*** (0.078)	-1.508 (1.082)
Constant	0.671*** (0.079)	1.648*** (0.364)
R2	0.14	0.230
N (# markets)	4 464 (144)	792 (144)

*: significant at the 10% level.

**: significant at the 5% level.

***: significant at the 1% level.

Note: Robust standard errors clustered at the session level. Data excludes the initial period of each market or any period in which cartel was already active.

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

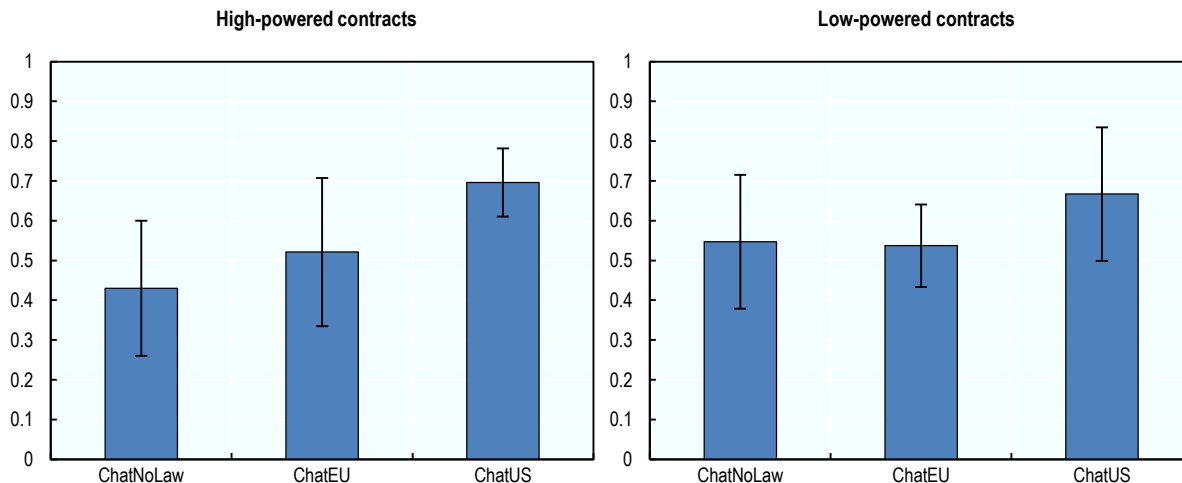
The results support the finding that cartels are most likely to re-form immediately after an audit takes place. The likelihood of a cartel being formed drops dramatically if it takes longer than two periods to reform. Interestingly, conditional on an audit having taken place, cartels are less likely to be formed under high-powered contracts in the ChatUS treatment but the reverse is true for the ChatEU treatment. This suggests that Hypothesis 2 for the ChatEU treatment appears to be supported by the data, but not for the ChatUS treatment. By contrast, as above, Hypothesis 1 does not appear to be supported by the data.

Pricing behaviour

We now turn to pricing behaviour conditional on the cartel being active or not. We start by quantifying the proportion of periods in which firms were able to co-ordinate and charge the same price in a given period.

When firms could not or did not form a cartel, they were able to co-ordinate on a price in 6% of cases (2 instances in NoChat, 39 cases in ChatEU and 37 cases in ChatUS). In all cases, the price they co-ordinated on was 10. In contrast, firms were able to charge the same price in 57% of periods in which a cartel was active. Of those instances, the chosen price was almost always 10. The cases in which firms co-ordinated on a different price were $p = 9.99, 7.5, 4.5, 3$ (one observation each), as well as 5 and 4 (two observations each). Figure 4.3 displays the relative frequency of instances in which firms co-ordinated on the same price based on market-level averages.

Figure 4.3. Proportion of cases in which cartels were active and firms co-ordinated on prices



Note: Error bars denote cluster-robust 95% confidence intervals.

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

To undertake the correct statistical inference, we must control for the market being played within a session, as well as the fact that markets are not statistically independent of each other. Results of the estimation are available in Table 4.4.

Table 4.4. Estimates on the propensity of managers to co-ordinate on the same price when a cartel is active

Dependent variable: SamePrice	Coefficient and statistical significance (standard errors)
ChatEU	-0.020 (0.090)
ChatUS	0.118 (0.102)
High-Powered	-0.116** (0.043)
High-Powered x ChatEU	0.067 (0.064)
High-Powered x ChatUS	0.114* (0.065)
Market 2	0.180*** (0.042)
Market 3	0.288*** (0.074)
Market 4	0.366*** (0.067)
Constant	0.338*** (0.086)
R2	0.16
N	196

*: significant at the 10% level.

**: significant at the 5% level.

***: significant at the 1% level.

Note: Each observation is the relative frequency of instances in a given market where firms all charged the same price and a cartel was active in that period. Standard errors are clustered at the session level.

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

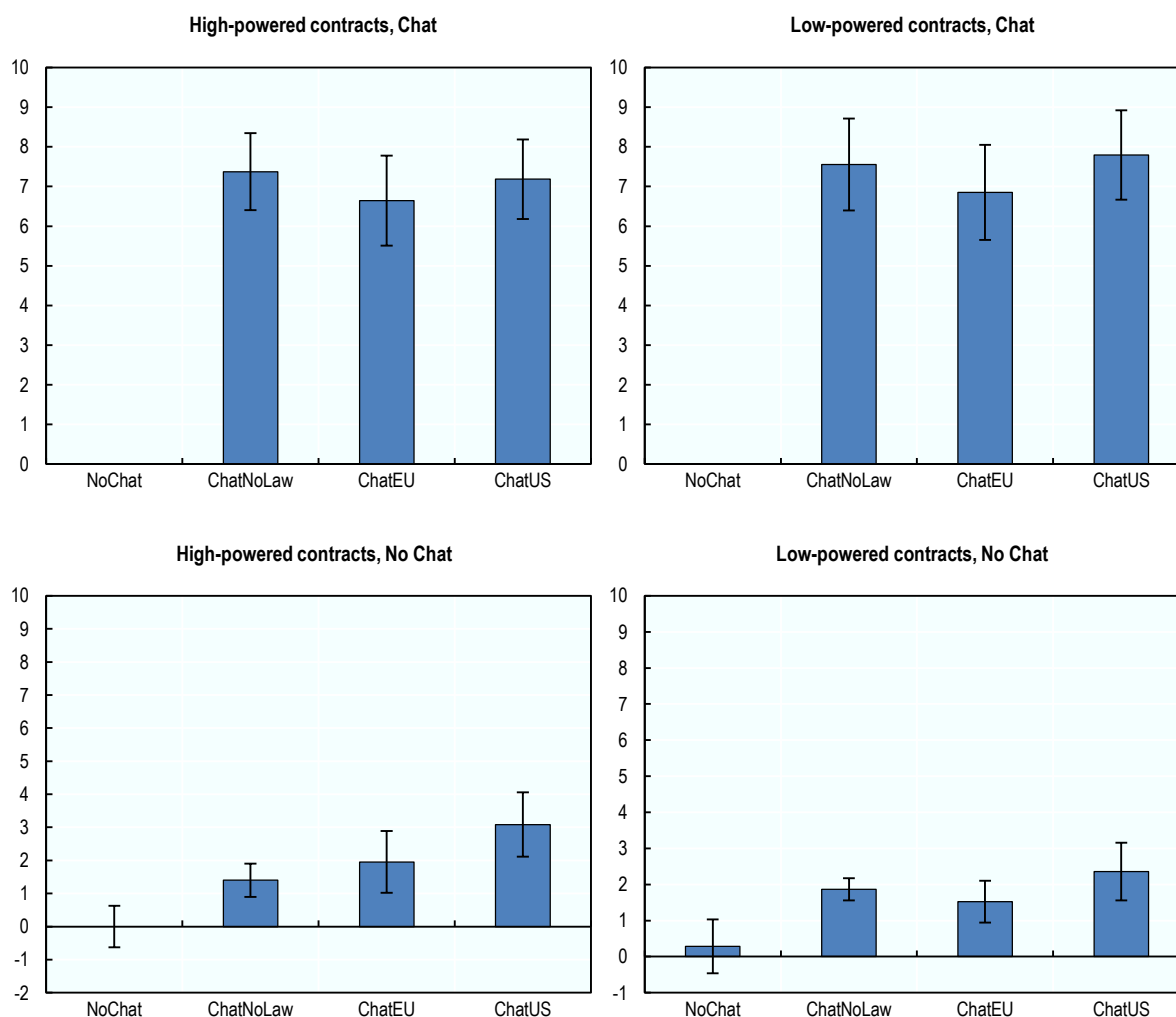
Conditional on a cartel being active, there is no statistically significant difference in the ability to co-ordinate prices across treatments between ChatUS and ChatEU when contracts are low-powered: $(1,17) = 2.38, p = 0.142$; however, when contracts are high-powered, there is a marginally significant difference: $F(1,17) = 3.52, p = 0.078$). This implies that Hypothesis 3 is supported by the data in what concerns the comparison between these 2 treatments for a given contract. We also find that managers have greater ease in co-ordinating on prices in later markets than in earlier markets. In addition, we do find that managers in the ChatNoLaw treatment find it harder to co-ordinate on prices in high-powered contracts than in low-powered contracts ($(1,17) = 7.43, p = 0.014$). A similar result was obtained in the ChatUS treatment ($(1,17) = 3.10, p = 0.096$) but not in the ChatEU treatment ($F(1,17) = 1.08, p = 0.313$). This implies that Hypothesis 3 is only partially supported by the data when we compare managers’ ability to co-ordinate prices within a treatment but across contracts: only in the ChatEU treatment do we find that the type of contract plays no role in managers’ ability to co-ordinate prices.

We now move to the analysis of the average selling price. Figure 4.4 displays the average selling price conditional on treatment, whether a cartel was active and the type of manager contract. We complement the figure with econometric analysis which accounts for time trends within a market as well as learning across markets. The econometric estimates are Table 4.5.

Average selling prices are higher when cartels are active than when they are not. Focusing only on the former case, we find no significant difference across different antitrust regimes conditional on the type of contract. However, we do find an effect of contract type conditional on the antitrust regime: average prices in the ChatUS treatment

are lower when manager contracts are high powered ($\chi^2(1) = 8.59, p = 0.003$). Conversely, when cartels are not active, average prices in the ChatUS treatment are significantly higher when contracts are high-powered ($\chi^2(1) = 9.69, p = 0.002$). This suggests that managers may have attempted more forcefully to collude tacitly after the cartel was detected in ChatUS than in ChatEU, because of the harsher punishment regime in the former treatment. Ultimately, the four-firm Bertrand environment is too competitive and frustrated those attempts. Understanding whether tacit collusion would have been successful in duopolies or triopolies is an interesting future line of inquiry.

Figure 4.4. Average selling prices conditional on contract type and a cartel being active



Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

Table 4.5. Estimates of the determinants of selling price

Dependent variable: Selling price	Coefficient and statistical significance (standard errors)
Active Cartel	5.686*** (0.527)
NoChat	-1.578*** (0.358)
ChatEU	-0.342 (0.250)
ChatUS	0.489 (0.373)
Active Cartel x ChatEU	-0.361 (0.697)
Active Cartel x ChatUS	-0.246 (0.731)
High Powered	-0.466** (0.219)
Active Cartel x High Powered	0.286 (0.413)
High Powered x NoChat	0.053 (0.402)
High Powered x ChatEU	0.896* (0.488)
High Powered x ChatUS	1.196*** (0.384)
Active Cartel x High Powered x ChatEU	-0.921 (0.698)
Active Cartel x High Powered x ChatUS	-1.629*** (0.556)
Initial Period	0.302 (0.224)
Period	0.047* (0.028)
Market 2	-0.333 (0.248)
Market 3	0.078 (0.434)
Market 4	-0.064 (0.694)
Constant	1.866*** (0.152)
R2	0.59
N (Markets)	2 736 (288)

*: significant at the 10% level.

**: significant at the 5% level.

***: significant at the 1% level.

Note: The unit of observation is the selling price of market j in period t . Standard errors are clustered at the session level.

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

Voting behaviour

We conclude the analysis by looking at contract choices made by firm owners. Figure 4.5 displays the frequency with which firm owners offered low powered contracts to managers.

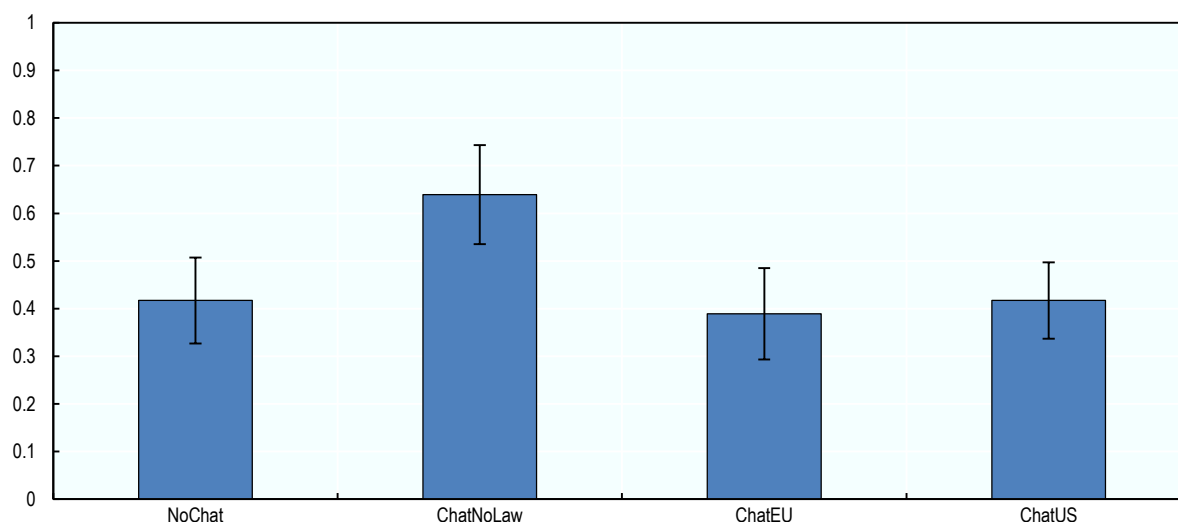
In the NoChat, ChatEU and ChatUS treatments, the proportion of low powered contracts offered is 40%. In contrast, in the ChatNoLaw treatment, the proportion of low-powered contracts goes up to over 60%.

There is no significant difference in the frequency of low-powered contract choices between the NoChat and ChatEU ($F(1,23) = 0.18, p = 0.677$); NoChat and ChatUS ($F(1,23) = 0.00, p = 1.00$) and ChatEU and ChatUS treatments ($F(1,23) = 0.20, p = 0.662$). The latter clearly provides evidence contrary to Hypothesis 4. In contrast, the proportion of low-powered choices in the ChatNoLaw treatment is significantly higher than in the NoChat ($F(1,23) = 10.40, p = 0.004$), ChatEU ($F(1,23) = 12.46, p =$

.002), and ChatUS treatments ($F(1,23) = 11.38, p = 0.003$). This provides some evidence in support of Hypothesis 4 for the ChatNoLaw treatment.

To provide some additional information, we conducted a short unincentivised survey after the experiment had concluded. One of the open-ended questions pertained to the reasoning behind the contract choice. Based on the responses we came up with a series of categories, which we feel capture the responses.

Figure 4.5. Proportion of low-powered contracts offered



Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

The first category is Self Interest, which includes mentions of cost minimisation, or maximising profits, or simply stating selfishness as a motive. Examples of responses which were coded under this heading include: “More profit for me”; “Because the manager only getting 2% of the revenue means more revenue for me”.

The second category is Own Experience, which included instances where subjects referred to their own behaviour or outcomes in the markets in which they took part as managers as a justification for their choice. Examples include: “More money was made when variable pay was higher in my groups”; “I chose it because I earned more ECU in these sessions and so would hope that other people did too”.

The third category was Incentive in Contract to Collude (ICC), which refers to instances in which subjects explicitly referred to giving high-powered contracts in order to incentivise collusion explicit or tacit. Examples included: “Realised that if managers were acting rationally then my profit would be higher if the contract has 20% of revenue in, as it should encourage collusion”; “20% people more likely to co-operate, thus profits should be higher”.

Interestingly, a non-trivial proportion of subjects mentioned incentives, though not in the way the ICC is predicated to operate. Some subjects argued that low-powered incentives were less likely to induce deviations, which in turn should lead to more price stability. We categorised these instances as Incentives (Non-ICC). Examples include: “With a

lower variable component, people tended to “go rogue” less as it paid off much less and so wasn’t worth it, so everyone stuck to putting 10. This means there would be a greater chance of picking a firm that had some form of profit”.

We also categorised under this heading entries that referred to incentives that related to pricing behaviour but did not explicitly mention the formation of cartels. Examples include inducing more competitive behaviour: “So that the manager would more carefully choose his price, since he would be receiving a considerable portion of the revenue if his price was the lowest”; “I wanted to motivate the manager not to co-operate”; or “The 20% variable was a higher motivation to make better sales to the consumer using an established unit price. I was more encouraged to think about my decision if more was to potentially be earned”.

Another category we generated was Altruism, which refers to instances where the subject expressed concern for the welfare of the manager, or general welfare. Examples include: “20% is the nice choice for managers”; “Give the other manager an opportunity to earn as much as possible”; or “Feels more ethical even in a simulation. Plus, it means that the people who play in the future have more of a chance of receiving ECU in revenue”. All other instances are summarised under Other.

Table 4.6 summarises the relative frequency of each coding category for each of the four treatments. In all treatments, the most frequently cited motive is narrow self-interest, expressed either via minimising the amount paid to managers or as maximising own payoff in a narrow, non-strategic sense (in that the potential gains from collusion are not explicitly mentioned). In NoChat and ChatNoLaw, incentivising the manager in a generic sense is the second most frequently used motive, closely followed by motivating the manager to engage in collusion, which becomes the second most frequent motive in ChatUS and ChatEU. Importantly, other-regarding concerns account for a sizeable fraction of responses, as does the firm owner’s experiences in the role of manager as to what was the most profitable environment.

This may help understand why we do not observe a larger difference between ChatUS and ChatEU in the fraction of high-powered contracts: not enough firm owners were either motivated or sufficiently strategically sophisticated to understand the potential for motivating managers through their contract choice.

Table 4.6. Relative frequency (%) of self-reported motives for contract choice NoChat

	NoChat	ChatNoLaw	ChatEU	ChatUS	Overall
Altruism	8	8	10	7	8
ICC	11	7	14	18	13
Incentives (Non-ICC)	13	11	8	10	10
Own Experience	4	4	7	14	7
Self-Interest	36	56	32	32	39
Other	28	14	29	19	23

Source: Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.

There are two important issues to take from the contract choice data. First, while low-powered contracts do (theoretically) provide the incentive to collude in the ChatEU treatment, it appears that 60% of firm owners do not choose that contract (to offer to managers). This contradicts Hypothesis 4 regarding the ChatEU treatment.

An alternative explanation for the high percentage of high-powered contracts is reciprocity manifested through gift exchange: firm owners could offer a generous contract to managers in the expectation that the latter will reciprocate with higher (attempts at) collusion. However, the higher frequency of low-powered contracts in the ChatNoLaw treatment undermines that explanation. The gift exchange hypothesis should predict the same behaviour across treatments: if firm owners want their manager to collude, they should be generous. The (standard) game-theoretic approach says the opposite: firm owners should not offer high-powered incentives unless there is an incentive-compatibility constraint to satisfy and managers need to be incentivised to collude. This is borne out by the data: in the ChatNoLaw treatment, there is no inherent downside risk in communicating and communicating greatly facilitates collusion. Therefore, the managers' incentive-compatibility constraint is always satisfied and we observe more low-powered contracts being offered than in the ChatEU and ChatUS treatments, where managers face the risk of being fined if they communicate.

Whilst the NoChat treatment is (apparently) puzzling, recall Hypothesis 4 in that regard: the contract choice should make no difference to firm owners. Therefore, one should expect an approximately equal number of low- and high-powered contracts being chosen. They anticipate that managers will play the Bertrand-Nash equilibrium of the stage game (i.e. $p \leq 0.01$) so the expected loss from the variable component of the wage is negligible. This justifies their choosing such a high proportion of high-powered contracts.

Discussion

We find that both individual and corporate fines have some success in deterring managers from forming cartels, however, by and large, the experimental results provide very little support to the theoretical predictions we made. Therefore, it is important to understand why this is so. Whilst we do not provide a definite answer here, we point to potential explanations that may help in understanding in what ways (and why) experimental subjects may diverge from our theoretical expectation.

Let us start by considering the decision to form a cartel. Our theoretical approach suggests that managers should base their decision on the sign (positive or negative) of the following expression, where U represents each manager's expected discounted utility (cartel formation constraint):

$$U_{\text{collusion}} - U_{\text{competition}}$$

Similarly, once a cartel is formed, a manager's decision to abide by the cartel agreement instead of deviating and undercutting the agreed price should depend on the sign (positive or negative) of this expression (incentive compatibility constraint):

$$U_{\text{collusion}} - U_{\text{deviation}}$$

These two expressions depend on several variables as well as several assumptions. Let us start with the variables, some of which are "controlled" through experimental parameters, whilst others are manager-specific. The experimental parameters that affect these expressions are: the contract (fixed and variable component), the type of fine (individual or corporate), the magnitude of the fine, the probability of cartel detection, the discount

factor, the number of consumers, the maximum price they are willing to pay and the number of firms in each market. The above description of our experimental design explains in detail the specific parameters chosen for each of these variables.

But there is one variable that is manager-specific, that is, which cannot be “controlled” in the experiment and which depends on each individual: the manager’s risk aversion. A more risk-averse manager will have (all else equal) a lower expected utility of a “lottery”, that is, of a specific random event that may affect his final outcome. For example, take a coin toss: with 50% probability it lands “heads” and the individual wins EUR 10 and, with 50% probability, it lands “tails” and the individual wins nothing. The more risk-averse an individual is, the lower will his expected utility associated with this coin toss be. Applying this to our setup, this implies that more risk-averse individuals will have lower expected utilities associated with the collusive, the deviation and the competition outcomes. Whilst, on its own, an increased risk aversion will affect all outcomes in a similar manner, thus preserving the sign of the above expressions, that may not be the case in conjunction with other “deviations” from our theoretical approach.

In our experimental design, we have relied on our theoretical approach and assumed that all managers had a similar degree of risk aversion as well as a similar discount factor. In combination with the remaining “experimental” parameters, this allowed us to determine the sign of the above expressions and, thus, to determine, in each treatment and for each contract, whether or not we expected cartels to emerge (and subsist over time). Therefore, one possible explanation for the discrepancy between our theoretical predictions and our experimental results may be associated with manager heterogeneity: heterogeneity in the risk aversion parameters and heterogeneity in the strategies chosen under each circumstance.

Recall that, in the experiment, a cartel would be formed if all members agreed to form the cartel. Under our theoretical approach’s assumption of symmetry among managers, if the above expressions were positive for a manager, they would be positive for all managers. But clearly, all that is necessary for a cartel not to emerge in the experiment is for one of the two above expressions not to be positive for a given manager. And this could happen because his risk aversion parameter in combination with other “deviations” from our theoretical approach is significantly different from those of the remaining managers. If that were the case, we would expect a cartel not to be formed, although the number of “votes” to form a cartel was very high. When we look at our results for cartel prevalence, it becomes clear that this may be one possible explanation for our results: when we consider only the number of cartels formed, neither Hypothesis 1 nor Hypothesis 2 are supported by the data. However, when we look instead at the number of votes to form a cartel, Hypothesis 2 is supported by the data in the ChatEU treatment. This means that manager heterogeneity may help in explaining the divergence from our theoretical expectation.

It is therefore important to understand what other “deviations” from our theoretical approach may explain the observed results. These “deviations” are linked with the theoretical assumptions made when evaluating the sign of the above expressions. Let us start with *Ucollusion*, that is, the manager’s expected discounted utility along the collusive path. Our primary assumption is that once a cartel is formed, i) managers are able to co-ordinate on a single price and that ii) that price is equal to 10, the maximum price consumers would be willing to pay for the good (full collusion). In other words, our underlying assumption for the cartel is that, if formed, it would successfully behave as a monopolist. Also, this implies that the cartel would generate the highest possible payoff

that any co-ordination strategy could achieve. From an economic viewpoint, as we have discussed above, this is logical. However, the data suggests that this was not the only co-ordination strategy used, although it was, by far, the most prevalent.

Regarding *Udeviation*, the manager's expected discounted utility if he disrupts the agreement, the main assumption made is related to the periods subsequent to the deviation period. In particular, if a manager deviates from the cartel agreement (whichever it is, that is, whatever is the agreed price level), his firm captures the whole market in the deviation period, but is then likely to be punished by the remaining cartel members in subsequent periods. The assumption we have made was that the remaining cartel members reverted to the Bertrand-Nash equilibrium strategies in all subsequent periods. That is, all chose a price of zero in all periods after an observed deviation. Whilst this is a standard assumption in the literature, it is also true that it is the harshest possible punishment that the remaining cartel members can impose on someone who deviates. In addition, whilst this is undoubtedly a punishment on a deviator, it also affects the remaining cartel members, whose firms are thus assumed to receive no revenues until the market ends and its managers are assumed to receive only the fixed component of their wages. Although we do not report this above, there are several cases of cartels that subsisted even after a deviation. In other words, in those cases, a deviation did not trigger the Bertrand-Nash equilibrium. Casual observation of the experiment as it was evolving suggests that managers continued to talk with one another after a deviation occurred. In some cases, the remaining cartel members would inquire about the deviator in order to understand why that happened. And in several cases, all managers were able to co-ordinate prices again. Clearly, in those cases, *Udeviation* would be higher than we predicted and this would affect the incentive compatibility constraint. In other words, this may affect cartel survival and help explain why it was that firms were only able to co-ordinate on prices in 57% of all periods in which a cartel was active.

In addition, whether or not "punishment strategies" differed across contracts may also help explain our results. For example, suppose that the reversion to Bertrand-Nash equilibrium strategies is more likely in the low-powered contract. This could be rationalised because with the low-powered contract, the variable component was low and, therefore, the difference (in utility) between a punishment strategy based on reversion to Bertrand-Nash equilibrium strategies or an alternative strategy (including, for instance, the possibility that the agreement is re-enacted) is also low. In that case, one would expect price co-ordination to be more difficult under high-powered contracts. With a higher variable component: a manager knows, when deviating, that everyone will be hurt if there is a reversion to Bertrand-Nash equilibrium strategies and may thus find it easier to convince the remaining cartel members to re-enact price co-ordination, which will increase firms' revenues and, through the variable component, managers' payoffs. Therefore, all else equal, after a deviation occurs, one could expect more successful price co-ordination with high-powered contracts.

Finally, our main assumption associated with *Ucompetition*, the manager's expected discounted utility if there is no price coordination, is that managers choose the Bertrand-Nash equilibrium strategies, i.e. they set a price of zero, which generates no revenues and managers only receive the fixed component of their wages. This is a crucial assumption to determine whether or not a cartel is formed, as it will affect the sign of the first expression above (cartel formation constraint). There are two dimensions associated with this assumption: a first when we think of a cartel being formed for the first time; and a second when, within a market, we think of a cartel formation decision after it has been detected.

Consider the decision to form a cartel for the first time. Arguably, this is when the assumption of Bertrand-Nash equilibrium strategies makes more sense: managers have not yet communicated with one another and will have observed “market behaviour” in those periods. It may be that Bertrand-Nash equilibrium strategies were not played, but it is very likely that prices will have been relatively low.

By contrast, after a cartel has been formed and detected by the competition authority, managers will know their co-operation history. In particular, they will know whether or not they were successful in co-ordinating prices whilst the cartel operated. Therefore, it is possible that after the cartel was detected, cartel members no longer feel the need to explicitly communicate again (and still be able to set high prices). Explicitly forming a subsequent cartel by accepting to communicate exposes members to the risk of being detected (and fined). If they can successfully coordinate prices without explicit communication, they may well prefer to do so. Naturally, this means that tacit collusion is enacted. In other words, it may be that *Ucompetition*, either because managers do not choose Bertrand-Nash equilibrium strategies or because they tacitly collude, is higher than we theoretically predicted and this may explain why it was that in the ChatEU and ChatUS treatments there were several instances where cartels were not re-enacted after detection. In addition, it may well be that such alternatives to Bertrand-Nash competition were more likely with a particular type of contract. As above, with high-powered contracts, managers may be more likely to engage in alternative strategies other than Bertrand-Nash competition in order to generate revenues for their firms which, indirectly, benefit them through the variable component of their wages.

We do not think that, on their own, either of these explanations is sufficient to explain our results. All may have played a role in explaining such a significant divergence from our theoretical predictions. Naturally, it is our intention to explore this issue further in future research.

Conclusion

This paper sets out to understand how the antitrust regime influences cartel behaviour. Specifically, our focus is on the punishment regime when cartels are detected: are there significant differences in cartel behaviour when the fines imposed are of an individual nature (levied on the firms’ managers) as opposed to being corporate in nature (levied on the firms’ revenues)? We combine this question with the particular characteristics of managers’ labour contract, namely the type of contract they are offered – fixed salary vs. fixed and a variable component – but also salary levels.

Our approach is both theoretical and experimental. In the theoretical approach, we develop a framework to analyse managers’ incentives to form and sustain cartels, as well as to determine firm shareholders’ optimal contract choices. We find that the different antitrust regimes induce different optimal contract choices by firms’ shareholders. In particular, shareholders choose fixed wage contracts under and antitrust regime with corporate fines but when fines are individual in nature, the optimal contract may contain a variable component or, in some cases, be purely variable.

In our experimental approach, we use our theoretical framework to carefully design an experiment where firms’ managers interact with one another in various different settings. We consider a setting where they cannot communicate with one another, as well as three settings where they can – one without the threat of punishment for collusion and

two others where the punishment for collusion is either through individual or corporate fines.

Interestingly, while we find that both individual and corporate fines have some success in deterring managers from forming cartels, the experimental results diverge significantly from our theoretical predictions. These differences encompass: differences in the cartel prevalence; differences in the pricing behaviour; and, importantly, differences in the contractual choices made when subjects take on the role of shareholders.

In our brief discussion of possible factors underlying those differences, we point out that not only may managers not follow the strategies predicted in our theoretical framework (namely when reacting to a deviation from the collusive agreement) but they may also diverge from these in different ways. Manager heterogeneity – both in manager characteristics and in manager behaviour – may, therefore, be one of the main explanatory factors for the divergence between our theoretical predictions and experimental results.

Inevitably, our results provide more questions than answers and it is our intention to continue our research on this theme. Nevertheless, it is also our belief that by highlighting the significant differences between the theoretical models of collusion and observed (experimental) behaviour, we are contributing not only to the fine-tuning of those models but also to practitioners who are faced with real situations of collusion and need to make decisions.

Notes

¹ Hellwig and Hüschelrath (2017), among others, confirm that the EC imposes only monetary fines.

² Wellensteyn granted refunds to P&C Düsseldorf for clothes returned by costumers after purchase.

³ As Dargaud et al. (2013), we assume that the criminal sanctions directly applicable to managers in the US jurisdiction can be modelled as a utility reduction of the manager's utility.

⁴ In equilibrium, the least talented are production workers, business owners are those gifted with low talent, managers receiving a bonus that does not vary with ability are medium talented, and managers paid with a flexible compensation that varies according to ability and firm size are highly talented.

⁵ Potters and Suetens (2013) provide a good review of Industrial Organization experiments since 2000.

⁶ The main element in the calculation of this variable component is the percentage of revenues that accrues to the manager, which, naturally, lies in between 0% and 100%.

⁷ Notice that one of these salary components can be equal to zero.

⁸ Naturally, managers who do not accept the contract may have several outside options to choose from, e.g. unemployment, employments offers in other markets, etc. For the purpose of our model, we assume that the best “outside option” has an associated fixed income per period.

⁹ To be more precise, all that we require is that the shareholders' outside option yields a payoff that is lower than that which would be obtained when the firm receives no revenues and the shareholders' payoff is negative in the amount of the manager's fixed component. For example, a significant reputational cost could be assumed to exist and shareholders would incur it in case their

firm is left without a manager. This would be sufficient to ensure that shareholders always design contracts that induce manager's acceptance.

¹⁰ If the variable wage component is lower than 50%, this implies that for every additional EUR 100 in revenues, managers receive less than EUR 50 through their variable wage component and shareholders receive more than EUR 50 in additional profits. Therefore, an increase in firm's revenues increases shareholders' profits and thus contributes towards reducing the possible loss associated with the manager's fixed wage component.

¹¹ This is not entirely realistic but recidivism would introduce significant complexity into the analysis.

¹² This assumption differs from Dargaud et al. (2013), who assumed that no fine is paid in the deviation period.

¹³ Recall that we only focus on collusive agreements that involve all firms.

¹⁴ To be more precise, the US antitrust regime is both a corporate fine and individual fine regime; that is, both coexist. In our model, we focus only on individual fines and assume corporate fines not to exist.

¹⁵ It is widely acknowledged in the experimental economics literature that subjects require repetition of a game in order to learn what the optimal course of action is. Dal Bó and Frechette (2011) argue that in the context of repeated games, it is necessary for subjects to also learn how to play the repeated game.

¹⁶ In relation to our previous remark – that a corporate fine regime generates stronger incentives for cartel formation than an individual fine regime – this implies that the incentive compatibility constraint is satisfied under both regimes with a high-powered contract, that is, the difference between the expected utility of the collusive outcome and a unilateral deviation is positive but it is higher (in magnitude) under a corporate fine regime.

References

- Allen, D. and D. Lueck (1992), "Contract choice in modern agriculture: Cash rent versus cropshare", *Journal of Law and Economics*, Vol. 35, pp. 397-426.
- Anderhub, V., S. Gächter and M. Königstein (2002), "Efficient contracting and fair play in a simple principal-agent experiment", *Experimental Economics*, Vol. 27, pp. 5-27.
- Apesteguia, J., M. Dufwenberg and R. Selten (2007), "Blowing the whistle", *Economic Theory*, Vol. 31, pp. 143-166.
- Aubert, C. (2009), "Managerial effort incentives and market collusion", *TSE Working Paper Series*.
- Autorité de la concurrence (2016), "Control of anticompetitive practices", Autorité de la concurrence, France, http://www.autoritedelaconcurrence.fr/user/standard.php?lang=en&id_rub=510&id_article=2856 (accessed on 21 March 2019).
- Autorité de la concurrence (2010), "Decision No. 10-D-28 of 20 September 2010 on prices and associated conditions applied by banks and financial institutions for processing cheques submitted for encashment purposes", Autorité de la concurrence, France, http://www.autoritedelaconcurrence.fr/doc/10d28_en.pdf (accessed on 21 March 2019).
- Bageri, V., Y. Katsoulacos and G. Spagnolo (2013), "The distortive effects of antitrust fines based on revenue", *Economic Journal*, Vol. 123, pp. 545-557.

- Bartalevich, D. (2014), “EU competition policy and US antitrust: A comparative analysis”, *European Journal of Law and Economics*, Vol. 44(1), pp. 91-112.
- Bigoni, M. et al. (2012), “Fines, leniency and rewards in antitrust”, *RAND Journal of Economics*, Vol. 43, pp. 368-390.
- Buccirossi, P. and G. Spagnolo (2007), “Optimal fines in the era of whistle-blowers. Should price fixers still go to prison?”, in V. Ghosal and J. Stennek (eds.), *The Political Economy of Antitrust*.
- Bundeskartellamt (2017a), “Bundeskartellamt imposes fines on account of vertical price fixing in the sale of furniture”, Bundeskartellamt, Germany, http://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Kartellverbot/2017/B1-164-13_B1-167-13_B1-87-14_B1-47-15.pdf?__blob=publicationFile&v=3 (accessed on 21 March 2019).
- Bundeskartellamt (2017b), “Fines imposed on account of vertical price fixing in the clothing industry”, Bundeskartellamt, Germany, http://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Kartellverbot/2017/B2-62-16.pdf?__blob=publicationFile&v=2%20 (accessed on 21 March 2019).
- Chamberlain, E. (1948), “An experimental imperfect market”, *Journal of Political Economy*, Vol. 56, pp. 95-108.
- CMA (2016), “Supply of products to the furniture industry: Anti-competitive arrangements,” Competition and Markets Authority, UK, <https://www.gov.uk/cma-cases/supply-of-products-to-the-furniture-industry-suspected-anti-competitive-arrangements> (accessed on 21 March 2019).
- Competition Bureau (2018), “Investigating cartels”, Competition Bureau, Canada, https://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/eng/h_02760.html (accessed on 21 March 2019).
- Connor, J.M. (2008a), “Anti-cartel enforcement by the DOJ: An appraisal”, *The Competition Law Review*, Vol. 5(1), pp. 89-121.
- Connor, J.M. (2008b), “Cartel enforcement at the Antitrust Division, U.S. Department of Justice, 1990-2007”, *GCP, the Online Magazine for Global Competition Policy*, pp. 1-12.
- Cooper, D.J. and K.-U. Kühn (2014), “Communication, renegotiation, and the scope for collusion”, *American Economic Journal: Microeconomics*, Vol. 6(2), pp. 247-78.
- Dal Bó, P. and G. Fréchette (2011), “The evolution of cooperation in infinitely repeated games: Experimental evidence”, *American Economic Review*, Vol. 101, pp. 411-429.
- Dana, J.D. Jr and K.E. Spier (2001), “Revenue sharing and vertical control in the video rental industry”, *Journal of Industrial Economics*, Vol. 49, pp. 223-245.
- Dargaud, E., A. Mantovani and C. Reggiani (2016), “Cartel punishment and the distortive effects of fines”, *Journal of Competition Law and Economics*, Vol. 12(2), pp. 375-399.
- Dargaud, E., A. Mantovani and C. Reggiani (2013), “The fight against cartels: A transatlantic perspective”, Working Paper, Institute d’Economia de Barcelona.
- European Commission (2017), “Antitrust: Commission fines Scania EUR 880 million for participating in trucks cartel”, European Commission, http://europa.eu/rapid/press-release_IP-17-3502_en.htm (accessed on 21 March 2019).
- European Commission (2013), *EU Competition Law: Cartel legislation*, Publications Office of the European Union, Luxembourg, http://ec.europa.eu/competition/cartels/legislation/cartel_compilation_en.pdf.

- European Union (2012), *Consolidated Version of the Treaty on the Functioning of the European Union*, EUR-Lex, European Union, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012E%2FTXT> (accessed on 21 March 2019).
- Federal Trade Commission (n.d.), “The antitrust laws”, Federal Trade Commission, United States, <https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/antitrust-laws> (accessed on 21 March 2019).
- Fehr, E., A. Klein and K.M. Schmidt (2007), “Fairness and contract design”, *Econometrica*, Vol. 75, pp. 121-154.
- Fischbacher, U. (2007), “z-Tree: Zurich toolbox for ready-made economic experiments”, *Experimental Economics*, Vol. 10(2), pp. 171-178.
- Fonseca, M. and H.-T. Normann (2014), “Endogenous cartel formation: Experimental evidence”, *Economics Letters*, Vol. 125, pp. 223-225.
- Fonseca, M. and H.-T. Normann (2012), “Explicit vs. tacit collusion – The impact of communication in oligopoly experiments”, *European Economic Review*, Vol. 56, pp. 1759-1772.
- Fershtman, C. and K.L. Judd (1987), “Equilibrium incentives in oligopoly”, *American Economic Review*, Vol. 77, pp. 927-940.
- Gil, R. and F. Lafontaine (2012), “Using revenue sharing to implement prices: Evidence from the movie exhibition contracts”, *Journal of Industrial Economics*, Vol. 60, pp. 187-219.
- Gonçalves, R. et al. (2018), “Cartel deterrence and manager labour market in USA and EU antitrust jurisdictions: Theoretical and experimental evidence”, Unpublished, Research Paper for the OECD, Católica Porto Business School, Porto.
- Greenwood, R. and L. Empson (2003), “The professional partnership: Relic or exemplary form of governance?”, *Organization Studies*, Vol. 24, pp. 909-933.
- Greiner, B. (2015), “Subject pool recruitment procedures: Organising experiments with ORSEE,” *Journal of Economic Science Association*, Vol. 1(1), pp. 114-125.
- Hamaguchi, Y., T. Kawagoe and A. Shibata (2009), “Group size effects on cartel formation and the enforcement power of leniency programs”, *International Journal of Industrial Organization*, Vol. 27, pp. 145-165.
- Hellwig, M. and K. Hüschelrath (2017), “Cartel cases and the cartel enforcement process in the European Union 2001-2015: A quantitative assessment”, *The Antitrust Bulletin*, Vol. 62(2), pp. 400-438.
- Hinloopen, J. and H.-T. Normann (2011), *Experiments and Competition Policy*, Cambridge University Press.
- Hinloopen, J. and A. Soetevent (2008), “Laboratory evidence on the effectiveness of corporate leniency programs”, *RAND Journal of Economics*, Vol. 39, pp. 607-616.
- Ichino, A. (1994), “Flexible labor compensation, risk sharing and company leverage”, *European Economic Review*, Vol. 38, pp. 1411-1421.
- Karakostas, A., A. Sonntag and D.J. Zizzo (2017), “Contract choice: Efficiency and fairness in revenue-sharing contracts”, *Scandinavian Journal of Economics*, Vol. 119(4), pp. 962-986.
- Katsoulacos, Y. and D. Ulph (2013), “Antitrust penalties and the implications of empirical evidence on cartel overcharges”, *Economic Journal*, Vol. 123, pp. 558-581.
- Katsoulacos, Y., E. Motchenkova and D. Ulph (2015), “Penalizing cartels: The case for basing penalties on price overcharge”, *International Journal of Industrial Organization*, Vol. 42, pp. 70-80.

- Lasserre, B. (2009), “The new French competition law enforcement regime”, *Competition Law International*, pp. 15-20.
- OECD (2005), “Hard core cartels: Third report on the implementation of 1998 Council Recommendation 7”, OECD, Paris.
- Potters, J. and S. Suetens (2013), “Oligopoly experiments in the current millennium”, *Journal of Economic Surveys*, Vol. 27(3), pp. 439-460.
- Practical Law Competition (2019), “UK cartels and restrictive agreements: A quick guide”, *Thomson Reuters Practical Law*, [https://uk.practicallaw.thomsonreuters.com/0-381-9658?transitionType=Default&contextData=\(sc.Default\)&firstPage=true&comp=pluk&bhcp=1](https://uk.practicallaw.thomsonreuters.com/0-381-9658?transitionType=Default&contextData=(sc.Default)&firstPage=true&comp=pluk&bhcp=1) (accessed on 21 March 2019).
- Ruffle, B. and H.-T. Normann (2011), “Introduction to the special issue on experiments in industrial organization”, *International Journal of Industrial Organization*, Vol. 29, pp. 1-3.
- Shaffer, G.C., N.H. Nesbitt and S. Weber Waller (2015), “Criminalizing cartels: A global trend?”, in A. Duke, J. Duns and B. Sweeney (eds.), *Research Handbook on Comparative Law*, Edgar Elgar.
- Szymanski, S. and S. Késenne (2004), “Competitive balance and gate revenue sharing in team sports”, *Journal of Industrial Economics*, Vol. 52, pp. 165-177.
- Vickers, J. (1985), “Delegation and the theory of the firm”, *Economic Journal*, Vol. 95, pp. 138-147.
- Wils, W.P. (2005), “Is criminalisation of EU competition law the answer?”, *World Competition*, Vol. 28, pp. 117-159.
- Wils, W.P. (2002), “Does the effectiveness enforcement of Articles 81 and 82 EC require not only fines on undertakings but also individual penalties in particular imprisonment?”, in C. Ehlermann and I. Atanasiu (eds.), *European Competition Law Annual*, Hart Publishing, Oxford.
- Wu, Y. (2017), “Incentive contracts and the allocation of talent”, *Economic Journal*, Vol. 127, pp. 2744-2783.

Chapter 5. Fostering a safety culture in the energy sector

This chapter examines how regulatory policymakers can foster elements of a strong culture of safety in the energy sector. It presents the results of an online experiment with regulators and regulated entities in Canada, Ireland, Mexico and Oman that tests the effects of messenger, feedback and social norms on changing behaviours around safety.

Introduction

Individual-level errors, such as inattention, forgetfulness and procedural violations, have long been regarded as the principal factor behind safety incidents and disasters (Reason, 1990). However, the origins of safety risks can often be organisational rather than individual in nature. The conditions under which individuals work and how individuals perceive these conditions are fundamental drivers of safety performance.

Safety culture, defined as the set of “shared values, beliefs, attitudes, norms and practices related to safety within an organisation” (TRB, 2016; Cooper, 2000), is a key aspect of the larger organisational culture and a crucial element for the prevention of organisational accidents. There is evidence from an analysis of global incidents that a poor safety culture contributes to many high-consequence accidents, such as the nuclear safety system failure at the Fukushima Daiichi plant in Japan in 2011 and significant organisational and cultural lapses that contributed to the BP Deepwater Horizon oil spill in 2010.

Prevention of such incidents strongly supports further research on safety culture and action for regulators to better serve the public interest. Regulators have a role to play in working with regulated entities and sharing responsibility to advance safety culture across the industries that they oversee. A key aspect of this duty requires them to lead the way by understanding their own organisational cultures and behaviours, their cultural strengths and vulnerabilities, and how these factors can influence the broader safety and regulatory system. Equally important is to understand the cultural and organisational changes and behaviours of regulated entities and industry to ensure that a safety culture is effectively implemented and impact the sector as a whole.

The use of behavioural insights can provide a powerful tool to understand barriers and opportunities to develop a safety culture within regulators and in regulated entities in high-risk sociotechnical systems. There is growing interest amongst governments and policy institutions to utilise behavioural science to enhance organisational behaviour – from government institutions themselves to external organisations that interact with and are regulated by governments. The increased application of behavioural insights (BI) can help countries across the world to regulate better based on actual and not assumed behaviour.

The OECD has developed extensive knowledge in the application of BI, including in the work of regulatory agencies across sectors and countries (OECD, 2017). As part of the OECD cross-sector work on applying BI to public policy, a project was developed on BI and safety culture, under the auspices of the OECD Network of Economic Regulators and with the support of the Government of Canada (Natural Resources Canada and National Energy Board of Canada); Ireland’s Commission for Regulation of Utilities; Mexico’s Agency for Safety, Energy and Environment; and Oman’s Authority for Electricity Regulation.

The aim of the project was to conduct computer-based experiments with representatives from both regulators and regulated entities in high-risk industries, in order to test the application of BI to strengthen different dimensions of safety culture. The experiment was designed to capture participants’ perception of:

1. The perceptions of workers from regulators and regulated entities regarding safety culture in their respective areas (safety culture).
2. The extent to which different actors would respond to the potential application of behavioural insights to common safety problems (scenarios/vignettes).

This second element constitutes one of the main contributions of the study, as it is one of the first instances of the application of behavioural scenarios and vignettes (a common tool in behavioural and psychological research) to study safety improvement. This allowed us to address a number of novel research questions, including: Do employees of high-risk industries respond differently depending on whether a new safety guideline is introduced by a manager vs. a peer (messenger effect)?; When it comes to safe behaviour, do regulators and regulated entities workers pay more attention to what people do or the norm for what they ought to do (social norms)?; Do participants react more strongly when feedback about their organisations' safety performance is provided in comparison with the performance of other organisations (social benchmarking/feedback)?

In addition, the study presents a uniquely international sample, which included regulators and regulated entities from diverse contexts: Canada, Ireland, Mexico and Oman. This allowed us to systematically examine whether there was heterogeneity in how different cultures respond to the same behavioural scenarios. The results indicated that there exist important international differences in perceptions of safety culture and not all behaviourally informed initiatives aimed at improving safety performance can work as effectively in all country contexts. Potential explanations are discussed in the results section of this chapter.

Equally, there is an asymmetry in the perception of safety culture by regulators and regulated entities; and this perception also differs at different levels of the organisational structure – with frontline workers generally perceiving the organisational culture to be less safe than managers consider it to be. Finally, the study suggests that not all behavioural principles are equally powerful in strengthening elements of safety culture; and notably, feedback on safety performance was perceived as the most impactful tool for behaviour change among the ones tested.

Overall, the study draws important implications on how context can enhance safety in high-risk industries and points at novel methodologies to study safety culture in different countries. We hope that the results of the experiments can inform guidance on fostering strong safety culture by:

- Describing practical approaches to assessing and addressing behavioural barriers and enablers to strengthening elements of safety culture.
- Understanding key decision points within regulators and regulated entities where action can be taken to strengthen elements of safety culture.

In the sections that follow, we summarise the context, methodology and findings of the study and provide a discussion of the challenges and opportunities for applying BI to safety culture.

Context and problem setting

What is safety culture?

There is no internationally agreed upon definition of “safety culture”, however, at its core safety culture is an aspect of the larger organisational culture, including the organisation's values, beliefs, attitudes, norms, practices, competencies and behaviours related to safety (TRB, 2016; Cooper, 2000). In the literature, there is a clear understanding that safety culture impacts safety performance (Smith et al., 2009). For instance, one study analysed 15 major petrochemical accidents between 1980 and 2010, and noted that poor safety culture contributed to 12 of the 15 accidents (Fleming, 2012).

The measurement of safety culture is often conducted through employee surveys (Choudhry et al., 2007; Flin et al., 2000; Guldenmund, 2000) and is assumed to reflect employee perceptions of the organisation's activities and policies regarding safety. Safety culture perceptions are then believed to predict safety-related behaviours, for example, reporting safety concerns and adhering to safety rules. While there is some debate in the literature regarding the degree to which safety culture perceptions map onto organisational accidents (Antonsen, 2009; Kvalheim, Antonsen and Haugen, 2016), it is generally assumed that where safety beliefs and behaviours are shared and positive, then the safety culture is considered strong (Christian et al., 2009; Clarke, 2000; Guldenmund, 2000; Health and Safety Commission, 1993; Singer et al., 2009).

It is widely acknowledged that regulators have an important role in promoting safety culture, however, a number of reviews and commissions of accidents have highlighted the importance of the responsibility of the industries in combination with regulators to promote a safety culture, acknowledging limits of regulation and that regulators cannot create a safety culture on their own (TRB, 2016).

Applying behavioural insights to foster strong safety culture

Safety culture is related to behavioural insights in many ways. Guldenmund (2000) describes a safety culture as a framework that has unconscious assumptions for safety at its core. Surrounding the core assumptions are beliefs and values which are, in turn, surrounded by observable safety culture elements (e.g. safety behaviour). Therefore, behaviours for safety are an essential and, importantly, a visible component of safety culture.

The safety culture paradigm itself is a recent development in a series of safety innovations going back 70 years. The first such innovation comprises a number of technological innovations, safety standards and compliance frameworks going back to the 1950s. Further innovation occurred with a focus on risk assessment and mitigation, and safety management systems in the 1980s. Industrial safety practice became focused on safety culture in the 2000s (Cox and Flin, 1998; Langford et al., 2000). Many tools have been developed in the safety culture space, including awareness drives, regular discussions on safety (e.g. "safety moments"), and of course rules that provide clear boundaries on behaviour. These tools and strategies may take many forms, but at their core, they attempt to apply health and safety rules to moderate behaviour group attitudes and resultant behaviours.

While many hazards have been eliminated, systems put in place to predict risk and efforts made to improve culture regarding safety, further attempts to maintain safety ("Safety 2") require consideration of safety-critical behaviour (Krause et al., 2001). Furthermore, there remains the issue of the "person-centred problem", which is not addressed by other methods (Reason, 1990; Talabi et al., 2015a; 2015b; Dekker, 2001; Holden, 2009). Where conscious behaviour according to unwritten cultural rules and written instructions are not enough for safety, additional manipulation of unconsciously made choices might be useful (Lindhout and Reniers, 2017). This is an opportunity to turn to a BI approach. BI would add to concepts of safety culture and behaviour-based safety by creating safety systems that enable safe behaviour and decision-making (Krause et al., 2001). The future of safety culture is likely to absorb BI, including concepts from psychology, behavioural economics, decision-making and influence.

Given the position that behaviours occupy within a wider safety culture conceptualisation – as visible artefacts of several various psychological and cultural processes – it is important to not overstate BI as a solution for all safety issues. BI interventions ought to form one component of a larger safety management system. Yet, some argue that change in culture and change in behaviour are complementary processes for improving workplace safety (DeJoy, 2005). Indeed, developing safety procedures in the absence of safety awareness/knowledge might be dangerous. Notwithstanding, health and safety practitioners must remember that unsafe behaviour and decision-making is not the only cause of accidents, though it may be the last link in a causal chain.

Noting this limitation of BI, the approach has typically been applied to existing policy levers (including regulation, incentives and information) to enhance their effectiveness for shaping behaviour. Past safety culture strategies have focused on the system, BI focuses on the individual. BI interventions would amplify efforts such as making safety behaviour easy, providing timely information to supplement safety behaviour, highlighting the social nature of safe behaviour and making safe behaviour decisions attractive.

Key BI principles in safety culture literature

Messenger

There exists a wealth of research demonstrating that we process the same information differently depending on who we received it from (Clark et al. 2013; Eckel and Gintis, 2010). In behavioural sciences, this is commonly referred to as “the messenger effect”. For example, individuals are more likely to believe a message when it comes from an expert or authority figure. They are also more likely to conform to the behavioural aspects of the message, decreasing violations overall. People also appreciate information more from people they have a positive feeling for or who are a bit like themselves, like in peer-to-peer sharing of knowledge or ingroup-outgroup dynamics.

Social influences

Humans are social creatures and look to the behaviour of others for information on how they themselves should behave (Bicchieri, 2006; Goldstein et al., 2008). There are several psychological mechanisms by which this occurs and we investigate two.

Social benchmarking

People pay attention to feedback in almost everything they do and often cannot adjust their behaviour without it. Providing a benchmark can reduce mistakes and make the consequences of decisions more salient. For example, providing pre- and post-shift hearing test results to workers can increase the use of hearing protection in subsequent shifts, overcoming what is known as the “present bias” (Zohar et al., 1980). However, we do not always get personal feedback on what we do and often look to the behaviour of others and the feedback they receive. This kind of social benchmarking can be useful in positioning our own behaviour.

Social norms (speaking up)

Evidence also suggests that people tend to survey their social and physical environment for attitudinal and behavioural cues and they care deeply about what their neighbours do. This is especially true when their neighbours belong to their same social in-group. Social

norms act as a standard, informing individuals of what others think and do. We can be strongly influenced by our group memberships and there are many examples where individuals will automatically follow the behaviour of their peers to comply with social norms (Dolan et al., 2012). There are a number of implementations that exist to take advantage of this human trait, which have been applied to domain as diverse as public health, environmental behaviour, international development (Selinger and Whyte, 2011; Nolan et al., 2008; Thaler and Sunstein, 2008; de Groot-Mesken and Vlakveld, 2014; Goldstein et al., 2008a; 2008b; Haines, 1996; Hansen and Jespersen, 2013; Branson et al., 2012; Sunstein, 2006; Ariely et al., 2003; Oullier et al., 2010; Cialdini, 2005*; Avineri, 2014; Linkenbach and Perkins, 2003; 2005; Perkins et al., 2010). In the context of our research, we are particularly interested in applying social norms to create a norm of speaking up about unsafe practices at the workplace.

Reciprocity

The power of “reciprocity” for inducing co-operation is also a well-replicated effect in the behavioural literature (Fehr et al. 2002; Rand et al., 2014). As social beings, people like to keep promises and reciprocate. Therefore, when people observe that others are taking the time to do things for them, they are more likely to continue that engagement. For example, behaviour change can be achieved by writing down a promise or commitment to do something (e.g. meeting a deadline). The examples of implementations that take advantage of commitments and reciprocity are numerous (Thaler and Sunstein, 2008; Breman, 2006; Karlan and Zinman, 2007; Hansen and Jespersen, 2013; Oullier et al., 2010).

Safety culture context of countries in the analysis

Each of the countries selected for the study possesses peculiar characteristics with regards to how they regulate the energy sector in their country. Below, we provide relevant information on the country context and a description of the basic features of each of the national regulator entities included in the analysis. These countries were selected based on convenient sampling and a more detailed description of the work and characteristics of each regulator can be found in Annex 5.A.

Canada

The National Energy Board (NEB) is Canada’s energy and safety regulator. It makes regulatory decisions and recommendations that represent the interests and concerns of Canadians. In doing so, the NEB factors in economic, environmental and social considerations. The NEB oversees safety and environmental protection for the full life cycle of a project – from approval to construction, operation, abandonment and works with communities, sharing the goal of making energy infrastructure as safe as it can be. The NEB also monitors aspects of energy supply, demand, production, development and trade which the federal government controls. The NEB reports to parliament through the Minister of Natural Resources.

Mexico

Created in 2015, the Agency for Safety, Energy and Environment (*Agencia de Seguridad, Energía y Ambiente, ASEA*) is a technical regulator responsible for industrial and operational safety and environmental protection in Mexico’s hydrocarbons sector. It oversees activities throughout the hydrocarbons value chain, from exploration and

extraction to midstream and downstream transformation, production and storage as well as distribution and retail at the petrol station level. ASEA's aims are mapped under five dimensions (clients; industry; process; organisation and learning; and financial resources) and within each of these dimensions, there are medium- to long-term visions.

Ireland

The Commission for Regulation of Utilities (CRU) has responsibility for safety in the energy sector in three broad sectors:

- Regulating the activities of natural gas and liquid petroleum gas (LPG) undertakings with respect to safety under the Energy (Miscellaneous Provisions) Acts 2006 and 2012. This is carried out under the Gas Safety Framework, which covers shipping, supply, storage, transmission, distribution and use of natural gas, as well as certain specified LPG undertakings.
- Regulating upstream petroleum safety, including offshore safety under the Petroleum (Exploration and Extraction) Safety Acts, 2010 and 2015. This is carried out under the Petroleum Safety Framework (PSF) Requirement of the Petroleum Safety Framework (CER/16/023).
- Designation and oversight of the safety supervisory bodies charged with monitoring natural and liquid petroleum gas installers and electrical contractors doing domestic gas and electrical works respectively, with respect to safety under the Energy (Miscellaneous Provisions) Acts 2006 and 2012.

Oman

The Authority for Electricity Regulation (AER) is responsible for regulating the electricity sector and some aspects of the water sector. It was established by Article 19 of the Law for the Regulation and Privatisation of the Electricity and Related Water Sector promulgated by Royal Decree 78/2004 on 1 August 2004 and Amended by Royal Decree 59/2009 and 47/2013 ("the Sector Law"). The authority is a financially and administratively independent organisation and reports directly to the Council of Ministers. The authority's duties under the Sector Law are to protect the interests of its three main stakeholders: electricity customers, electricity sector companies, and the Government.

Understanding the system – Shared responsibility, awareness of safety culture, complacency

When discussing the scope of the research project with members of the committee from the countries of study, a few key concepts were highlighted as most important. We define them below and describe the approach with which the study addresses them.

Shared understanding of responsibility between regulators and regulated entities

It is important for regulators and regulated entities to be on the same page regarding shared responsibilities for safety culture. Indeed, collaborative contexts (rather than adversarial ones) are likely to lead to safer environments. Safety culture is deemed "strong" when safety attitudes and perspectives are positive but also shared among staff. Thus, it is important that the regulator and regulated entities share a similar perspective on their shared responsibilities as well as the state of safety culture in their sector. We

measure the extent to which safety culture perspectives are shared by asking both regulator workers and regulated entity workers to provide their perspective on safety culture in their sector. We describe the perspective as “divergent” when we are able to demonstrate a significant difference between these perspectives and “shared” when we cannot.

In practice, regulators and regulated entities have different perspectives on safety. Regulated entities, for example, have a very detailed understanding of safety in their organisation and relatively little understanding of safety in other organisations, whereas regulators have less of an understanding of any specific organisation’s safety (presumably less than the organisation itself does) but a good understanding of safety in across the sector. Differing perspectives, such as these, may impede a shared understanding of the situation and potentially cause conflict regarding shared responsibility.

Awareness of safety

Although the field of behavioural insights is more about shaping contexts than raising awareness, regulators have identified the importance of awareness of safety practices and the need to address it in their contexts. While it may be difficult to know all the ways in which safety can be compromised (this is the reason that prescriptive safety policies often fail), it is important for workers in safety-critical industries to exercise caution in their work and decision-making. Workers need to understand that there are many “known unknowns” as well as “unknown unknowns”, which are sometimes referred to as “black swans” (Taleb, 2007). This frame of thinking is an essential element of a strong safety culture. For example, safety management systems that are less prescriptive rely on the workforce having appropriate safety and risk awareness. We measure respondents’ awareness of safety culture with questions like “people in regulated entities understand how others’ jobs contribute to safety” and “voicing concerns about safety is encouraged”. While these questions are not usually considered psychometrically valid measures of safety culture awareness, they provide valuable insights into workers’ perception of their environment.

Complacency

Maintaining safe working environments is an ongoing challenge – employees need to always be vigilant of situations where safety might be compromised. But maintaining vigilance itself is a difficult task and can lead to fatigue. It is in situations like these that vigilance is most important. When operators have not faced a major incident in a number of years or when they are using what they understand to be safe technologies for the first time, complacency can negatively impact safety culture. We measure the presence and magnitude of complacency in participant entities with items like “everyone perceives that safety is their personal responsibility” and “people are committed to safety”.

Organisational structure, management and workers

There are studies documenting how understanding national cultural dimensions is important for safety culture (Mearns and Yule, 2009). There are several national cultural dimensions that have been documented to vary between nations: power distance, uncertainty avoidance, collectivism, masculinity-femininity and short-term orientation (Hofstede et al., 2010). These are understood to occur through national, cultural and educational institutions shaping shared values and behaviours with respect to the way nation-members think of and approach different cultural concepts.

Power distance, for example, refers to the way nation-members think of and approach hierarchies and power in interpersonal contexts. Uncertainty avoidance is the degree to which nation-members approach or avoid situations with uncertain social consequences. Collectivism is the tendency and acceptance for nation-members to act predominantly as members of a group or as individuals. Masculinity-femininity refers to a societal balance of “masculine” values (e.g. competitiveness, power) over “feminine” values (e.g. relationships, quality of life). Short-term orientation refers to the societal importance placed on the future (e.g. saving, adaptation) or past and present (e.g. respect for tradition, fulfilling social obligations).

There are documented relationships between these national cultural dimensions and elements of safety culture. For example, power distance has a negative relationship with safety culture: high power distances discourage speaking out and correction of superiors, create an unwillingness to challenge authorities and create asymmetric communication streams between management and frontline staff. Equally, uncertainty avoidance has a negative relationship with safety culture: high uncertainty avoidance restricts innovation, leads to more rigid rules and regulations, and leads to an over-reliance on procedures that cannot be applied to all contexts.

Collectivism also has a negative relationship with safety culture, as high collectivism means that group cohesion is prioritised over speaking up (or challenging group norms), thus increasing embarrassment for errors and self- or career-defensive behaviours. Masculinity is another cultural factor with a negative relation to safety culture: it can create competitive environments that obstruct collaboration. Finally, short-term orientation has a negative relationship with safety culture: high short-term orientation creates pressure for short-term gains at the expense of long-term planning and inhibits the development of safety from a holistic systems perspective.

From a policy standpoint, it may be important to acknowledge these differences amongst countries to understand the related differences in safety culture.

Methodology

Understanding context and fine-tuning the design

A key step to applying BI is a detailed understanding of the context in each of the countries studied. This process began through the OECD Network of Economic Regulators (NER) where the country representatives involved in the project as well as the broader community of economic regulators could provide their inputs into the initial stages of the research agenda. Following these discussions, this project and themes were discussed by representatives from participant countries in a small setting to gain a detailed understanding of each in each country context. These themes were then explored in a scoping literature review and discussed with academics at the London School of Economics (LSE).

Following this initial scoping, contact points within each regulator were established and numerous informal discussions were conducted to explore the practical elements of the project as well as an understanding of safety culture, the potential application of BI and these three themes in a more nuanced way. These discussions helped to inform the selection of the BI principles studied (i.e. messenger, social benchmarking, reciprocity and social norms/speaking up). This was followed by a survey to each contact point to gain a more detailed understanding of specific behaviours that could be addressed in the research.

Following discussions with the contact points and the responses from the behaviour survey, informal focus groups were conducted with two or three representatives from each regulator and regulated entities. Overall, there were a total of eight informal focus groups. The aim of these discussions was to gain a detailed understanding of the safety culture and behavioural science principles we were planning to test, as well as detailed feedback on the safety culture and experiment questions themselves (e.g. suggestions on language and terminology, etc). Feedback was provided on various iterations of the questionnaires, often at multiple times. Hypotheses about the most impactful interventions in each experiment were also collected from representatives of regulators and regulated entities.

A number of academic experts and practitioners from the wider safety culture and BI communities were also engaged to gain feedback on the experiment design and application of BI principles. Ethical approval was obtained from the LSE and the researchers followed OECD principles of confidentiality and ethics. The study was also pre-registered after data collection, but before data analysis, on the Open Sciences Framework (Tear and MacLennan, 2018).

Experimental design: Questionnaire on safety culture and behavioural scenarios

Once identified the main research questions and hypotheses, we designed a computer-based questionnaire and distributed it as a link to respondents, who completed it in their own time.

Administering the questionnaire

Emails were sent to respondents in regulators and to contact points within regulated entities, who then passed the questionnaire to frontline staff, managers and senior management, including contractors. The questionnaire was not sent to everyone who works in the regulator and certain individuals were excluded from the experimental sample (Annex 5.A). As we did not want to collect any personal/identifiable information, we wanted to ensure that the questionnaire was only sent to people who would be able to provide responses as we intended.

The email with the link to the questionnaire came from either a regulator email address or an OECD email address, with an initial email provided by the regulator to ensure that the message would not be lost as it would be received from an unfamiliar email address. In addition, the email was designed in line with the relevant literature on how to increase response to electronic questionnaires.

Language

For Canada, Ireland and Oman, the email and experiments were in English. For Mexico, they were both translated into Spanish. However, respondents had the option of responding to the questionnaire in English or Spanish. Feedback was provided from the contact points in Mexico (Mexico's Agency for Safety, Energy and Environment and the Mexican Association of Hydrocarbons) who read through the translated documents and agreed with the translation.

Timeframe

The questionnaire was sent in an email in August 2018. Responses were collected for six weeks and a reminder message was sent in early September. This was particularly important as it was highlighted that some staff work on five-week rotations and likely would not receive the initial email.

Design

The experimental questionnaire that was sent to subjects was divided into four main sections: 1) demographics; 2) safety culture; 3) behavioural Scenarios; and 4) qualitative questions. The complete questionnaire can be found in Annex 5.A.

1. Demographics

The aim of this section was to collect basic, unidentifiable data which could inform our analysis in the following sections. This section collected only the regulated entity/regulator of the respondent and what level they were in the organisation (i.e. frontline staff, manager, senior manager). In regulated entities, we additionally asked participants to specify whether they were a contractor. In initial discussions with senior contacts within study country regulators, it was discussed that regulated entity size, level of safety culture maturity and the level of contractors would likely impact safety culture. The information in this demographic section enables us to keep these areas under consideration when conducting the data analysis. We purposefully asked minimal questions in this section so that individuals could not be identified and their privacy would be protected. This was also crucial to ensure that participants felt they were not giving away a substantial amount of personal data and could, therefore, be honest in their responses.

2. Safety culture/climate questions

In discussions at the April 2018 NER as well as in scoping discussions with safety culture academics and contact points in the regulators of study countries, an emphasis was placed on understanding the extent to which views on safety culture differ between individuals in regulated entities and in regulators. Driven by this practical question and an understanding of the literature in this area, we opted to ask a number of related questions to respondents in both regulators and regulated entities.

We first asked participants to report how much they agreed with a number of questions on various dimensions of safety culture. The safety culture questions were derived from the existing literature (Reader et al., 2015) and slightly adapted to include a few novel questions on reciprocity between regulated entities and the regulator. Examples of questions include: “On average, in regulated entities, information about safety-related changes is clearly communicated to staff” and “On average, in regulated entities, people are committed to safety”. Participants then responded through a 7-steps Likert scale ranging from Strongly Agree to Strongly Disagree. The safety culture questions were asked in the same order for all respondents.

3. Behavioural scenarios

After gauging the general perception of safety culture in different entities, we asked individuals from regulators and regulated entities to respond to questions testing the application of BI principles through vignettes or scenarios.

Although placing individuals in vignettes/scenarios is common in the behavioural insights literature, this is not often done in the safety culture literature. Furthermore, in discussions with the regulator contact points, we ascertained that safety culture surveys using vignettes are very uncommon. Therefore, this project not only contributes value-added in terms of applying BI in an area which is not commonly studied – dimension of strong safety culture – but also, in terms of the methods used.

As described above, the experiments (i.e. vignettes/scenarios testing different behavioural science principles) were designed based upon an initial literature review, consultations with academics, discussions with contact points within regulators as well as informal focus groups with representatives from regulators and regulated entities. Ultimately, the behavioural principles chosen for the analysis were: messenger, benchmarking, reciprocity and social norms (speaking up).

It was important to ensure that the scenarios would resonate within each country context, but we also wanted them to be specific enough such that the respondents would be able to picture/understand each as it relates to their work. The specific examples in the vignettes were chosen through discussions with contact points and then discussed with members of the informal focus groups (we received iterative feedback from two to three people in regulators as well as at least one regulated entity, piloting the vignettes and associated questions). The criteria for the more specific vignettes were that they were generalisable across countries and sectors (from oil and gas to electricity) but also provide some level of specificity.

After careful examination, we determined that we were most interested in observing how participants would react to the following three scenarios:

1. The introduction of a new guideline regarding Personal Protective Equipment (PPE).
2. Reports of bad lost-time injury rate.
3. A situation where a supervisor asks a worker to carry out a task in an unsafe manner.

These three scenarios were then manipulated to test the application of the four behavioural principles mentioned above. For example, we tested whether workers would behave differently as a function of whether the PPE regulation was introduced by a manager or a peer (messenger effect). We acknowledge that these examples are not reflective of safety culture as a whole and have drawbacks in terms of PPE not reflecting wider aspects (e.g. near misses, etc.) and lost-time workers often gaming because the indicator gets too much attention, however individuals from frontline workers to senior management and those working for regulators are familiar with these concepts and find them important to some degree. Careful vignette design and bolding were used to place emphasis on the behavioural insights' principle and not on the context of the specific vignette chosen.

In terms of designing the questions following each scenario, we asked about three main aspects. First, we asked about attention and salience to the information in the vignette, as this is an important concept in BI in terms of understanding effectiveness. We hypothesised that different formulations of the same scenario would have a different saliency in the eyes of participants and that behaviourally-informed information would be perceived as more salient. Second, the questions asked about what the respondent in the survey (regulated entity) would do themselves or what a person in an entity (regulator) would do in a given scenario, which allowed us to get an understanding of how they viewed their behaviour/the behaviour of someone in an entity.

The third section of questions was focused on the organisational management specifically and asked participants how their managers would react to the scenario (regulated entities) or how the entity as a whole would react (regulators). The intention behind this question is linked to safety culture being about what you think of others/not yourself; asking about managers' behaviour also allowed us to investigate participants' perceptions of authority. This question was of particular interest as behavioural science literature demonstrates that people are better at predicting how others will behave as opposed to how they themselves will behave.

For the process described above, we also ask what respondents think should/would be impactful. This enabled us to compare a prediction about a factual matter with a judgement about the implications for what is ultimately the right thing to do. In other words, to gauge the perception of both descriptive and injunctive social norms.

Importantly, each respondent was asked all of the vignettes and the experiment did not involve random assignment to treatment. However, the scenarios were presented in a randomised order and we ensured that the same vignettes were not asked together as a group.

4. Qualitative

At the end of the questionnaire, we provided a place for respondents to write in the reasons why they selected what they did for each experiment (e.g. Why did you think a certain messenger would be more impactful compared to the others?). People are often poor predictors of why they make certain decisions and it may be due to rules of thumb/heuristics, but this opportunity for respondents to tell the narrative back to themselves may provide useful additional information which can help to interpret the results of the experiments.

Results and discussion

Results

The following section describes the results of the study. For both the safety culture question as well as the BI experiments, three main hypotheses were tested:

1. **Regulator vs. regulated entity:** Are there differences between the regulator and regulated entities in terms of the effectiveness of the scenarios? This is an exploratory hypothesis.
2. **Country differences:** Are differences between regulator and entities driven by national context? Can national culture account for the differences?

3. **Frontline vs. manager differences:** In nations where power distance is high (Oman, Mexico to a lesser extent), are there differences in responses between frontline staff and management staff? Is this related to the safety culture of the organisation?

Sample size

We conducted a number of data quality checks before completing analyses. First, we took the original dataset ($N = 1\,366$) and removed the data marked as potential spam (a feature of the online survey software). We then conducted a missing data analysis and created new samples based on the number of safety culture items and vignettes items participants responded to. Those who responded to <50% of the safety culture items were removed from the safety culture sample and those who responded to <50% of the vignettes items were removed from the vignettes sample. This left us with $N = 1\,033$ for the safety culture sample and $N = 885$ for the behavioural vignettes sample. Note, that these were not independent samples – it was possible (and likely) that participants would appear in both samples.

It is worth noting here that we do not have a statistically robust sample from Canada ($N = 28$). We can only guess at the overall response rate here. Given the number of Canadian organisations that had the opportunity to participate ($N = 96$) and the approximate number of workers who could have participated, we must conclude that a sample of 28 is very low and potentially susceptible to selection bias. We encourage caution in interpreting conclusions made about the Canadian sample as inferring conclusions about safety culture is difficult with a small sample size (Pronovost and Sexton, 2005).

As for the other nationalities, we had a total of 92 Irish respondents; 409 Mexican respondents and 504 from Oman.

Safety culture questions

The following section describes the results of the safety culture assessment. In many of the countries studied, this type of safety culture assessment had not previously been completed by regulators. The safety culture survey comprised several dimensions: perceived management commitment to safety, perceived regulator commitment to safety, collaboration for safety, reporting culture, communication for safety, colleague commitment to safety, safety support and perceived relationship between the regulator and regulated entities. These dimensions were then collated into a summary safety culture scores, which was the target outcome variable for these analyses. The target for the safety culture items was the safety culture in regulated entities. Thus, our data reflect the regulator and entity workers' perceptions of the safety culture in the average sector entity. Safety culture items were rated on a scale from 1 (strongly disagree) to 7 (strongly agree). Respondents were excluded from analyses if they answered less 50% of the safety culture items.

Before presenting the results, we must discuss a caveat for interpreting safety culture survey scores. It is difficult to determine exactly what a safety culture score represents. An element of strong safety culture is the ability for workers to criticise and question the decisions of senior management. Of course, if workers are criticising the decisions of senior management, then that may actually manifest itself as poorer safety culture scores (e.g. “senior management do not take action on safety issues when raised”). Another example concerns organisations whose workers have little experience of safety culture –

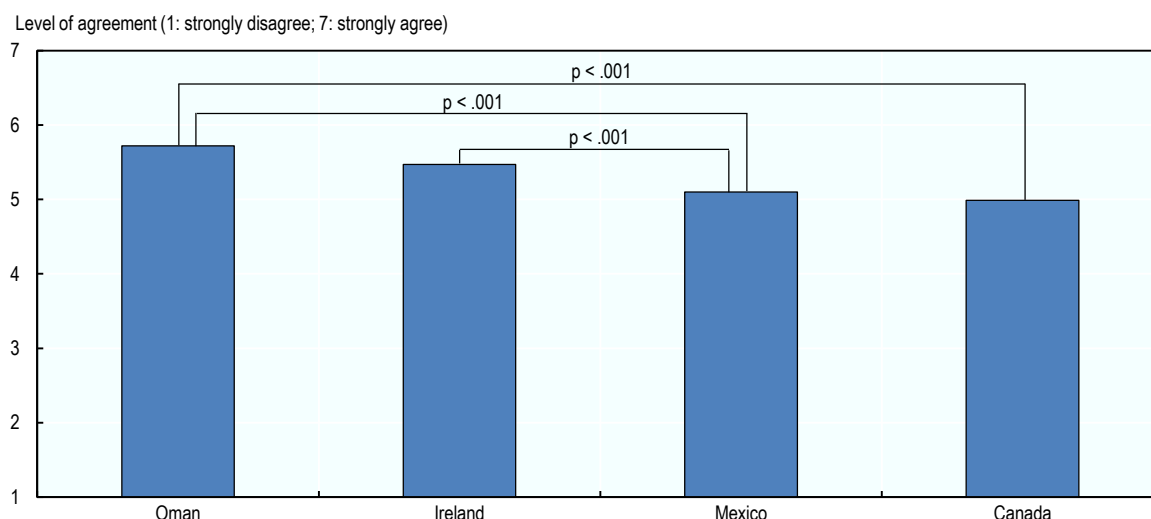
if they have not seen it before, then can they really know what strong or weak safety culture actually looks like?

For this reason, it is difficult to compare safety culture across organisations because the organisational context will inevitably vary. A comparison that makes more sense, however, is against earlier measurements within the same organisation (e.g. Organisation X at Time 1 and then again at Time 2).

National differences

When comparing how different countries responded to the safety culture questions, we detect significant national differences in the perception of safety culture. Figure 5.1 summarises the cross-national findings and their relative statistical significance.

Figure 5.1. National differences in safety culture perception



Note: For the Canadian data, due to the small sample size inferring conclusions about safety culture cannot be made.

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

As can be seen from the graph, the Omani sample displayed the most positive perception of safety culture ($M = 4.99$, $SD = 1.15$), followed by Ireland ($M = 5.47$, $SD = 0.78$) and Mexico ($M = 5.10$, $SD = 1.00$). Canada ranked the least positive ($M = 4.99$, $SD = 1.15$), perhaps due to the small number of respondents we were able to gather from Canada ($N = 28$). We also observed that the Canadian sample was overrepresented by regulator workers by a factor of 6 to 1. Thus, the mechanism driving the differences in safety culture perceptions between workers from regulators and regulated entities may be driving the result for the Canadian sample.

Three of the cross-national relationships were found to have statistical significance at the $p < .001$ level: respondents from Oman perceived the safety culture of the average regulated entity as more positive than respondents from Canada ($p < .001$) and Mexico

($p < .001$); and respondents from Ireland ranked significantly higher than the Mexican sample ($p = .001$).

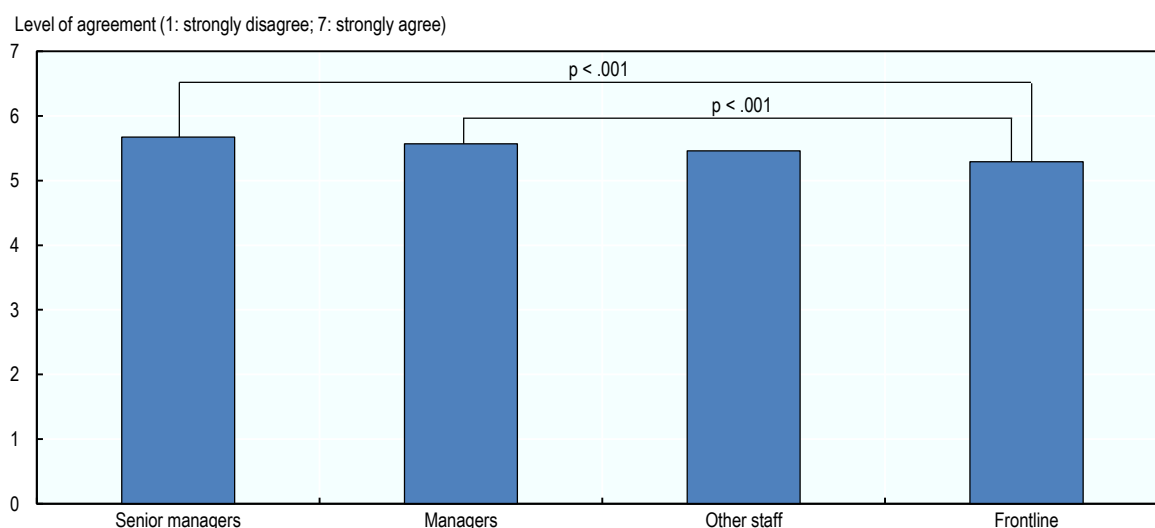
All other cross-national comparisons were statistically similar or only approaching statistical significance, as in the comparison of Ireland and Oman ($p = .076$).

Role differences

Our data on professional role differences replicated a common finding in the literature – frontline staff have a more negative perception of safety culture than management do. Here we observed data from 472 frontline staff, 217 respondents who identified themselves as managers, 107 senior managers and 237 other members of staff.

As displayed in Figure 5.2, senior managers displayed the most positive vision of safety culture in their organisation ($M = 5.67$, $SD = 0.76$), followed by managers ($M = 5.57$, $SD = 0.72$), other staff ($M = 5.46$, $SD = 0.90$) and frontline ($M = 5.29$, $SD = 1.01$). Additionally, we find that both managers and senior managers have a significantly higher perception of safety than frontline ($p = .001$).

Figure 5.2. Role differences in safety culture perception



Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Behavioural insights vignettes

The following section describes the results of the experiments. They are provided according to the BI principles tested in the experiments.

Overall

While we were able to demonstrate several significant differences in safety culture perception, it should be noted that the average scores were always positive (above 4) and the differences were never in excess of half a Likert point.

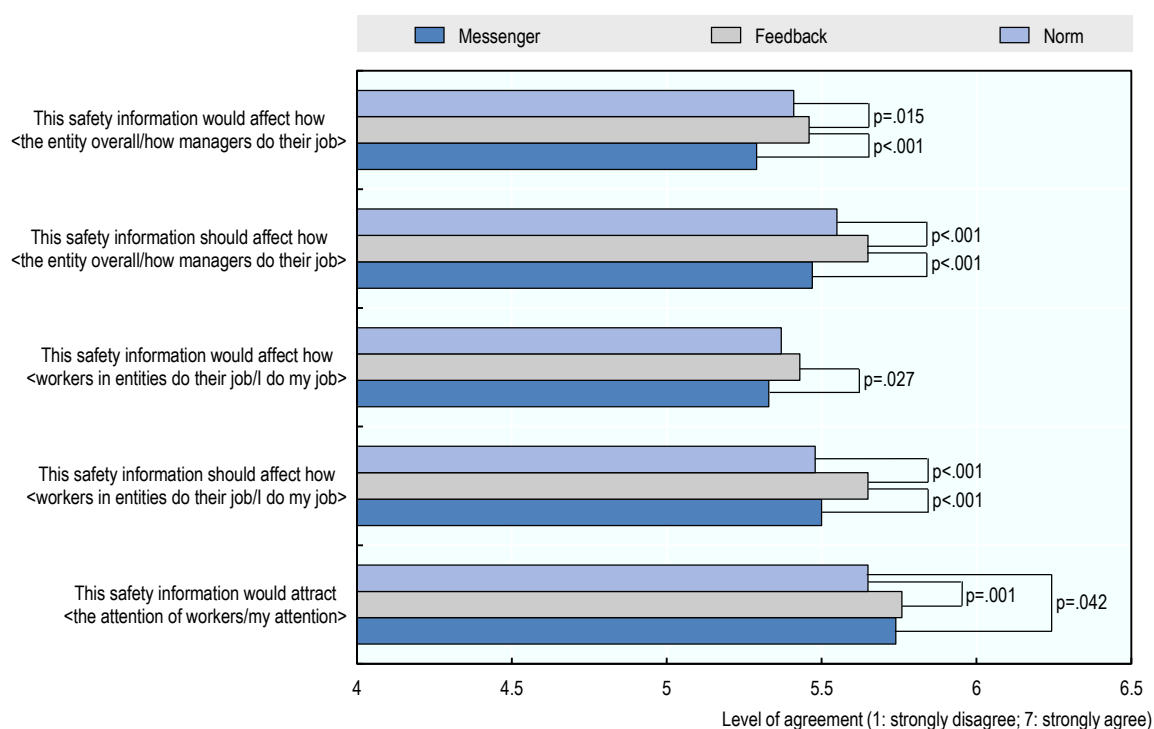
Regulator vs. regulated entity

Workers from regulated entities were more favourable to the vignettes than were the regulator worker. This is interesting when taken with the safety culture data, where regulator workers have more negative perceptions of safety culture and here, they are less receptive to behavioural interventions to address safety culture.

Behavioural principle

In general, respondents responded **most favourably to the feedback vignettes** compared to the messenger and norm vignettes. Vignettes informed by the messenger effect were perceived as the second most impactful. Interestingly, respondents felt like the norm vignettes would be the least effective at attracting the attention of workers or themselves. Figure 5.3 unpacks the differences in perceived effectiveness between the principles.

Figure 5.3. Comparing behavioural principles



Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Feedback effects

While overall, feedback was the most effective behavioural principles, responses to the five items after reading the different feedback vignettes (e.g. social benchmarking vs. feedback, vs. control) were statistically similar. That is, respondents believed that there was no difference between the levels of the feedback variable.

Norm effects

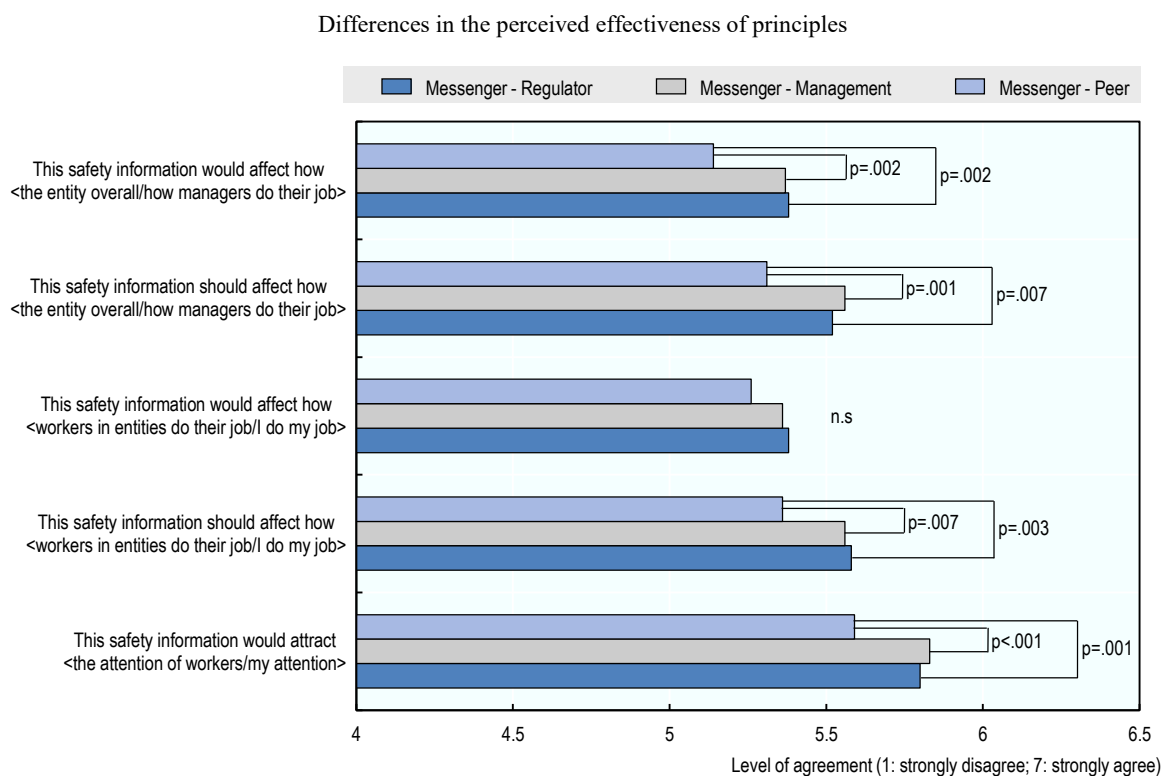
Norm conditions were roughly equivalent across each of the five effectiveness items, with the key descriptive and injunctive levels being statistically indistinguishable from the control.

There were some differences, however, with respondents rating the descriptive norm vignette more positively than the control vignette for two items: i) “this safety information would attract the attention of workers/my attention”; and ii) “this safety information would affect how workers in entities do their job/I do my job”.

Messenger effects

Respondents believed that messenger effects for regulators and management were statistically similar. They felt, however, that **the peer messenger effect would be the least effective**. This is interesting given there is ample evidence for the influence of peer messengers. Qualitative feedback indicated that peer-initiated safety direction would likely only have influence if it was picked up by the senior management, in which case the messenger becomes the organisation’s senior management.

Figure 5.4. Messenger effects



n.s. = no significance

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

International lens and understanding the results

We conducted several analyses of the behavioural principles data to see how their effectiveness might be affected by national cultural contexts.

Messenger

The messenger vignettes were responded to most positively by respondents from Ireland, followed by Omani respondents, and then Canadian and Mexican respondents approximately equal.

Feedback

The feedback vignettes were responded to most favourably by respondents from Oman, followed by Irish respondents, with Canadian and Mexican respondents alternating for least favourable responses.

Social norms

Again, Irish respondents were the most favourable to the norm vignettes, although Mexico and Oman responses were often quite favourable. Canadian responses were the least favourable.

To understand what might be driving these differences, we next looked at responses to each level of the principles by country. We observed that there are statistically **significant effects of social norms in Mexico** – indicating that the Mexican sample perceives descriptive norms as more effective than control and injunctive norms. In addition, we found that, **in both Oman and Ireland, the peer messenger was deemed to be less effective than the management messenger** on some items.

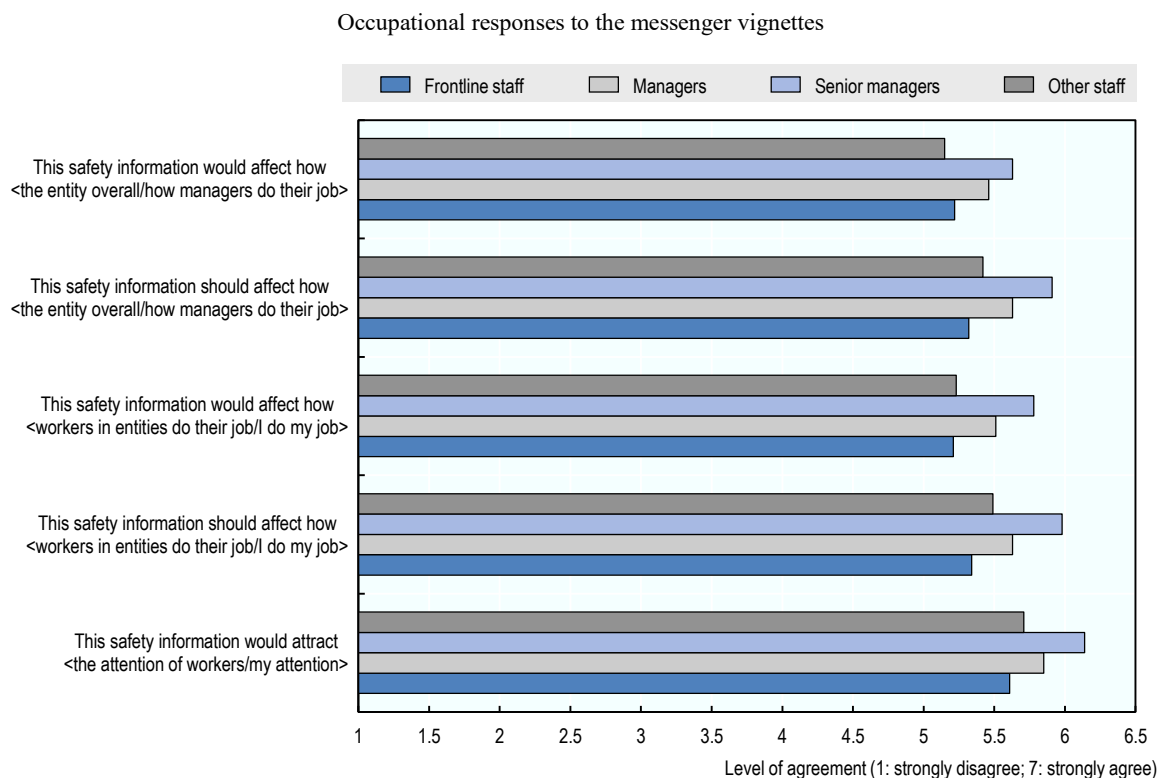
For the Canadian sample, no differences were found in how respondents perceived different behavioural cues. There were no differences in the perceived effectiveness of the different messenger types (regulator, management, peers); nor for different feedback types (control, simple control, social benchmarking, reciprocity); nor different norm types (control, descriptive, injunctive). This may be due to the small sample size.

Lens of an occupational role for understanding the results

We conducted further analyses of the behavioural principles data to see how their effectiveness might be affected by the **occupational role of respondents**.

Messenger

The messenger vignettes were responded to most positively by senior managers, followed by managers, and then frontline staff and other staff respondents responded approximately equally. As can be seen in Figure 5.5, the pattern is consistent across different items.

Figure 5.5. Occupational response to the messenger effect

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Feedback

The feedback vignettes followed the same general pattern as messenger vignettes, where senior managers responded most favourably, followed by managers, and then frontline staff and other staff respondents responded approximately equally.

Social norms

Responses to the social norm vignettes were slightly different. On Item 1, senior managers and managers were indistinguishable and significantly more positive than frontline staff and other staff (who were indistinguishable). For Item 2, only the difference between frontline staff and managers was statistically distinguishable (managers more positive). For Item 3, the only statistical difference was between other staff and managers (managers more positive). There were no differences between the occupational roles on Item 4. On Item 5, managers and senior managers were significantly more positive than other staff but only managers were more positive than frontline staff.

To understand what might be driving these differences we next look responses to each level of the principles by occupational role. By doing so, we found no differences in the perceived effectiveness of feedback type (control, simple control, social benchmarking, reciprocity) or norm type (control, descriptive, injunctive). However, the data revealed

that for frontline staff, managers and other members of staff (excluding senior managers), the peer messenger was deemed to be less effective than messages from regulators/managers.

Lens of an organisational background for understanding the results

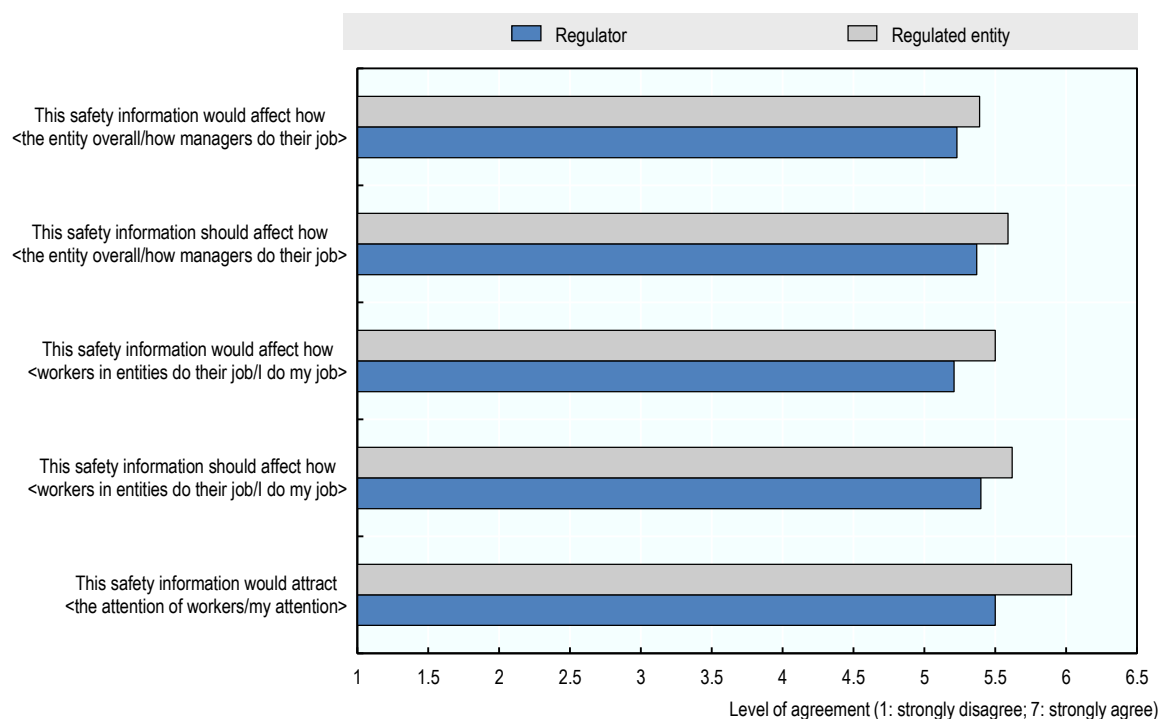
We conducted several analyses of the behavioural principles data to see if there were how their effectiveness might be affected by whether the respondents are from regulators or regulated entities.

Across all principles, respondents from regulated entities responded more favourably than respondents from regulators. This is exemplified by Figure 5.6, which displays how regulators and regulated entities responded to the messenger formulation of the vignettes. As can be seen, the perception of regulated entities (lighter shade) is consistently higher than that of regulators (darker shade) across all items.

We conducted further analyses into the principles to see if there were differences in the perceived effectiveness of the vignettes.

While there were no differences in the perceived effectiveness of the different feedback types or norm types, we did find that, for both regulators and regulated entities, the peer messenger was deemed to be less effective than manager and regulator messengers on several items, thus confirming the presence of a messenger effect.

Figure 5.6. Regulator vs. entity responses to messenger vignettes



Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Summary of behavioural results

Messenger

Overall, the messenger of safety instructions seems to mostly only matter in Ireland and Oman. In those countries, the peer messenger was perceived to be the least effective messenger vehicle. From a Hofstedian perspective, Ireland and Oman do not share many cultural similarities. Where they do share similarity is in the rate of regulator worker to entity worker responses (Ireland 1:3; Oman 1:4). Perhaps this majority of entity worker responses explains why the peer messenger was perceived least favourably. Where there is a messenger effect, its direction is such that messages from managers and regulators are deemed more effective than messages from peers.

Feedback

While feedback was overall the most impactful behavioural principle among the ones tested, A deeper investigation into the responses to the feedback vignettes failed to reveal any differences at the country, occupational or organisational level.

Social norms

In general, norms were found to be the least effective behavioural principle overall. However, cross-national comparisons revealed some interesting trends. In particular, the Mexican sample was the only nationality for which there were statistically significant results with regards to norm type. Mexican respondents were found to react more strongly to descriptive norms than control or injunctive messages.

Descriptive norms differ from injunctive norms in that they describe what people actually do, whereas injunctive norms describe what people ought to do (motivation may be unclear). From a Hofstedian perspective, Mexican samples score high on uncertainty avoidance, meaning that the clear signal from descriptive norms may be preferred over the motivationally unclear nature of injunctive norms. Mexican samples also score low on individualism, meaning that they may be more susceptible to group norms in general.

Limitations of the study

There are a number of limitations to this study such as varying degrees of English language abilities (particularly in Oman) and that we relied on contact points within the regulated entities to pass along the information. Where possible, efforts were made to counteract these limitations.

While it was possible that regulator workers have a more accurate perception of the state of safety culture in the sector, our data cannot support such a conclusion. We do not provide any kind of objective benchmark against which to compare workers' safety culture perceptions. Nor do we test whether regulator workers: i) do have access to more data about the state of the sector's safety culture; or ii) whether their perceptions are subsequently more accurate.

While we tested the perceived effectiveness of three distinct BI principles, there is likely a degree of overlap between them. For example, the feedback vignettes often incorporate a particular messenger (i.e. the regulator), which we have found elsewhere is an effective messenger. Our feedback vignettes also incorporate elements of social norms (e.g. being in the "bottom/worst 25%" is social information that respondents could compare against).

Part of the difficulty in interpreting the feedback vignettes is that they all share these common overlaps with other principles.

Conclusions

Policy lessons

Overall, the project constitutes one of the first applications of behavioural insights through online experiments to the study of safety improvement and elements of safety culture. It is intended to serve as a stepping stone towards a more frequent integration of the field of BI and safety. A number of key policy lessons emerge from the research.

- **To avoid unintended negative consequences, it is important for regulators to take into consideration differences in perception within and between actors when designing new safety regulations or policies.** The study found that the closer one is to the front line, the lower one's perception of safety culture. From a system perspective, the study showed that regulators have a more negative perception of safety culture in the regulated entities than the entities themselves, perhaps due to their position overseeing the sector. Moreover, results show that senior managers reacted most favourably to the behavioural principles (i.e. feedback, messenger effects and social norming) than other occupations, indicating that there are differences of perceptions within entities (not only between entities and the regulator) regarding safety culture. When developing policy, it is important to take these differences in perception into consideration to ensure policies are targeted for different audiences.
- **The study found that some feedback is better than no feedback but the results are inconclusive as to which type of feedback or benchmarking is best.** Results show that respondents reacted most favourably to feedback vignettes, compared to messenger and norm vignettes, generally speaking. From a policy perspective, this highlights the importance of considering providing workers with some form of feedback. However, which form of feedback is most effective and whether feedback is beneficial in every context needs to be studied in further detail.
- **The source of safety messages (messenger) still matters, which highlights the need to ensure regulators and senior managers in regulated entities are working together to encourage a culture of safety.** Results showed that respondents reacted similarly to messages on safety coming from a regulator as well as senior management of the regulated entity. However, messages from peers were considered the least effective, which runs counter to conventional thinking about the use of norms in nudging.
- **Social norming was perceived as the least effective across the sample, which requires more research to determine the benefit for policymakers of using social norms to encourage a culture of safety.** Results for all social norm vignettes were statistically indistinguishable from the control. However, some differences were noted for respondents rating the descriptive norm vignette more positively than the control vignette, giving some possible avenues for future research.

- **Differences between countries highlight the need for policymakers to take a location-based approach to strengthening elements of safety in their own contexts.** While the above notes the trends for each behavioural insight tested, between-country differences were notable. For feedback, there were no statistically significant differences at the country level; however, results did show respondents from Oman reacted most favourably, followed by Irish respondents, and then Canadian and Mexican respondents alternating for least favourable responses. Caution should, however, be taken inferring results from the Canadian results due to small sample size. For the messenger effect, it seems that this really mattered most in Ireland and Oman, perhaps due to a similar regulator worker to entity worker response rate. For social norming, the Mexican sample was somewhat responsive to descriptive norms, though Irish and Omani responses were also favourable. Canadian responses were least favourable.

Potential next steps/areas for further work

Although the results of this study point toward useful areas of potential attention, there is a substantial amount of merit in further research. Context is very important and, in order to have any policy recommendations/a toolkit, it is necessary that further work be carried out. This work may include the following:

- **Qualitative follow-up:** A key component of creating a toolkit would be a deeper understanding of the nuances in the context in each country setting. In order to obtain this level of understanding, it would be important to carry out focus groups and/or individual interviews in each country context.
- **Additional survey experiments:** There is scope for additional online experiments which could dive deeper into these behavioural principles, test different principles, focus on certain types of respondents, etc.
- **Randomised controlled trials:** Field experiments can be conducted to test the application of behavioural science principles in real-world contexts.
- **Behavioural lens on existing policies:** Taking a closer look at each country context and providing more tailored recommendations.

Complementary field work

Behavioural insights as a field can provide a great deal of understanding and can be complemented by data analytics and other academic disciplines that take a deep dive perspective. For example, qualitative research and design thinking can play an important role in facilitating, augmenting or strengthening and evaluating the impact of BI.

References

- Antonsen, S. (2009), “Safety culture assessment: A mission impossible?”, *Journal of Contingencies and Crisis Management*, Vol. 17(4), pp. 242-254.
- Ariely, D., G. Loewenstein and D. Prelec (2003), “Coherent arbitrariness: Stable demand curves without stable preferences”, *The Quarterly Journal of Economics*, Vol. 118(1), pp. 73-105.
- Avineri, E. (2014), “Nudging safer road behaviours”, Tel Aviv Academic College of Engineering Afeka/Acitrail <https://pdfs.semanticscholar.org/346b/dd0b382aa0bc15530d4d8b1d0411bcfcf1e1.pdf> (accessed on 28 October 2018).

- Branson, C. et al. (2012), *Acceptable Behaviour: Public Opinion on Behaviour Change Policy*, Ipsos MORI, <https://www.ipsos.com/ipsos-mori/en-uk/acceptable-behaviour> (accessed on 28 October 2018).
- Bicchieri, C. (2006), *The Grammar of Society*, Cambridge University Press, Cambridge.
- Breman, A. (2011), “Give more tomorrow: Two field experiments on altruism and intertemporal choice”, *Journal of Public Economics*, Vol. 95(11-12), pp. 1349-1357.
- Choudhry, R.M., D. Fang and S. Mohamed (2007), “The nature of safety culture: A survey of the state-of-the-art”, *Safety Science*, Vol. 45(10), pp. 993-1012.
- Christian, M.S. et al. (2009), “Workplace safety: A meta-analysis of the roles of person and situation factors”, *Journal of Applied Psychology*, Vol. 94(5), pp. 1103-1127.
- Cialdini, R. (2005), “Don’t throw in the towel, use social influence research”, *Observer*, Vol. 18(4-4), www.psychologicalscience.org/observer/getArticle.cfm?id=1762.
- Clark J. et al. (2013), “Evaluating the message or the messenger? Implications for self-validation”, *Persuasion, Personality and Social Psychology Bulletin*, Vol. 39(12), pp. 1571-84.
- Clarke, S. (2000), “Safety culture: Under-specified and overrated?”, *International Journal of Management Reviews*, Vol. 2(1), pp. 65-90.
- Cooper, M.D. (2000), “Towards a model of safety culture”, *International Journal of Disaster Risk Reduction*. Vol. 36(2), pp. 111-136, [https://doi.org/10.1016/S0925-7535\(00\)00035-7](https://doi.org/10.1016/S0925-7535(00)00035-7) (accessed on 21 March 2019)
- Cox, S. and R. Flin (1998), “Safety culture: Philosopher’s stone or man of straw?”, *Work and Stress*, Vol. 12(3), pp. 189-201.
- de Groot-Mesken, J. and W.P. Vlakveld (2014), “A nudge in the right direction: Safe traffic behaviour”, R-2014-13, SWOV Institute for Road Safety Research, Den Haag.
- DeJoy, D.M. (2005), “Behavior change versus culture change: Divergent approaches to managing workplace safety”, *Safety Science*, Vol. 43(2), pp. 105-129.
- Dekker, S. (2001), “Disinheriting Fitts and Jones ’47”, *International Journal of Aviation Research and Development*, Vol. 1(1), pp. 7-18.
- Dolan, P. et al. (2012), “Influencing behaviour: The mindspace way”, *Journal of Economic Psychology*, Vol. 33(1), pp. 264-277.
- Eckel C. and H. Gintis (2010), “Blaming the messenger: Notes on the current state of experimental economics”, *Journal of Economic Behavior & Organization*, Vol 73(1), pp. 109-119.
- Fehr, E. (2002), “Strong reciprocity, human cooperation, and the enforcement of social norms”, *Human Nature*, Vol. 13(1), pp. 1-25.
- Fleming, M. et al. (2012), “Cultural disasters: Learning from yesterday to be safe tomorrow”, *Oil and Gas Facilities*, Vol. 1(3), Society of Petroleum Engineers, Houston, Texas.
- Flin, R. et al. (2000), “Measuring safety climate: Identifying the common features”, *Safety Science*, Vol. 34(1), pp. 177-192.
- Goldstein, N.J., R.B. Cialdini and V. Griskevicius (2008a), “A room with a viewpoint: Using social norms to motivate environmental conservation in hotels”, *Journal of Consumer Research*, Vol. 35(3), pp. 472-482.

- Goldstein, D.G. et al. (2008b), “Nudge your customers toward better choices”, *Harvard Business Review*, pp. 99-105.
- Guldenmund, F.W. (2000), “The nature of safety culture: A review of theory and research”, *Safety Science*, Vol. 34(1-3), pp. 215-257.
- Haines, M.P. (1996), *A Social Norms Approach to Preventing Binge Drinking at Colleges and Universities*, Higher Education Center for Alcohol and Other Drug Prevention, Newton, MA.
- Hansen, P.G. and A.M. Jespersen (2013), “Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy”, *European Journal of Risk Regulation*, Vol. 4(1), pp. 3-28.
- Health and Safety Commission (1993), *Third Report: Organising For Safety*, ACSNI Study Group on Human Factors, London.
- Hofstede, G., G.J. Hofstede and M. Minkov (2010), *Cultures and Organizations: Software of the Mind (Third)*, McGraw-Hill.
- Holden, R.J. (2009), “People or systems? To blame is human. The fix is to engineer”, *Professional Safety*, Vol. 54(12), pp. 34-41.
- Karlan, D.S. and J. Zinman (2009), “Expanding credit access: Using randomized supply decisions to estimate the impacts”, *The Review of Financial Studies*, Vol. 23(1), pp. 433-464.
- Krause, T.R., G. Sellers and C. Horn (2001), “Moving to the 2nd generation in behaviour-based safety”, Symposium Series, No. 148, IChemE XVI-Paper-57, pp. 731-738.
- Kvalheim, S.A., S. Antonsen and S. Haugen (2016), “Safety climate as an indicator for major accident risk: Can we use safety climate as an indicator on the plant level?”, *International Journal of Disaster Risk Reduction*, Vol. 18, pp. 23-31.
- Langford, D., S. Rowlinson and E. Sawacha (2000), “Safety behaviour and safety management: Its influence on the attitudes of workers in the UK construction industry”, *Engineering, Construction and Architectural Management*, Vol. 7(2), pp. 133-140.
- Lindhout, P. and G. Reniers (2017), “What about nudges in the process industry? Exploring a new safety management tool”, *Journal of Loss Prevention in the Process Industries*, Vol. 50, pp. 243-256.
- Linkenbach, J. and H. Perkins (2005), “MOST of us don’t drink and drive campaign. A social norms strategy to reduce impaired driving among 21-34 year-olds”, U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Linkenbach, J. and H. Perkins (2003), “MOST of us wear seatbelts: The process and outcomes of a 3-year state wide adult seatbelt campaign in Montana”, presented at the National Conference on the Social Norms Model, Boston, MA, July 17.
- Mearns, K. and S. Yule (2009), “The role of national culture in determining safety performance: Challenges for the global oil and gas industry”, *Safety Science*, Vol. 47(6), pp. 777-785.
- Nolan, J. et al. (2008), “Normative social influence is underdetected”, *Personality and Social Psychology Bulletin*, Vol. 34(7), pp. 913-923.
- OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

- OECD (2017), *Behavioural Insights and Public Policy: Lessons from Around the World*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264270480-en>.
- Oullier, O. et al. (2009) “Improving public health prevention with a nudge”, *Improving Public Health Prevention with Behavioural, Cognitive and Neuroscience*, Centre for Strategic Analysis, Paris, pp. 38-46.
- Perkins, H.W. et al. (2010), “Effectiveness of social norms media marketing in reducing drinking and driving: A statewide campaign”, *Addictive Behaviour*, Vol. 35(10), pp. 866–874.
- Pronovost, P. and B. Sexton (2005), “Assessing safety culture: Guidelines and recommendations,” *BMJ Quality and Safety*, Vol. 14(4), pp. 231-233.
- Rand, D.G., E. Yoeli and M. Hoffman (2014), “Harnessing reciprocity to promote cooperation and the provisioning of public goods”, *Policy Insights from the Behavioral and Brain Sciences*, Vol. 1(1), pp. 263-269.
- Reader, T.W. et al. (2015), “Safety sans Frontières: An international safety culture model”, *Risk Analysis*, Vol 35(5), pp. 770-789, <https://doi.org/10.1111/risa.12327>.
- Reason, J. (1990), *Human Error*, Cambridge University Press.
- Selinger, E. and K. Whyte (2011), “Is there a right way to nudge? Practice and ethics of choice architecture”, *Sociology Compass*, Vol. 5(10), pp. 923-935.
- Singer, S. et al. (2009), “Relationship of safety climate and safety performance in hospitals”, *Health Services Research*, Vol. 44(2p1), pp. 399-421.
- Smith, A. et al. (2009), “Safety culture advice and performance”, Report for IOSH Research Committee, http://www.behavioral-safety.com/articles/cardiff_safety_culture_report.pdf.
- Sunstein, C.R. (2006), *Infotopia: How Many Minds Produce Knowledge*, Oxford University Press.
- Talabi, B.O., F.T. Edum-Fotwe and A.G.F. Gibb (2015a). “Construction actor safety behaviour: Antecedents, current thinking and directions”, in S.D. Smith and F. Sherratt (eds.), *Proceeding of ARCOM Doctoral Workshop: Health, Safety and Wellbeing*, 11 February 2015, University of Edinburgh, pp. 9-20.
- Talabi, B.O., F.T. Edum-Fotwe and A.G.F. Gibb (2015b). “Behaviour-based safety (BBS): A construction industry’s perspective”, in M. Behm and C. McAleenan (eds.), *Proceedings of the CIB W099 Benefitting Workers and Society through Inherently Safe(r) Construction*, 9-11 September 2015, Belfast, pp. 181–190.
- Taleb, N.N. (2007), *The Black Swan: The Impact of the Highly Improbable*, Random House, New York.
- Tear, M.J. and M. MacLennan (2018), “OECD behavioural insights for safety culture”, *Open Science Framework*, <https://osf.io/r482d/>.
- Thaler, R.H. and C.R. Sunstein (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Yale University, New Haven.
- TRB, (2016), “Strengthening the safety culture of the offshore oil and gas industry”, Transportation Research Board, Washington, DC.
- Zohar, D. (1980), “Safety climate in industrial organizations: Theoretical and applied implications”, *Journal of Applied Psychology*, Vol. 65(1), pp. 96-102.

Annex 5.A. Additional information and sample survey

Additional details on sample and participants

Annex Table 5.A.1. Additional information on regulators by country

	Canada	Mexico	Ireland	Oman
Sector	Oil and Gas	Oil and Gas	Oil and Gas Exploration and Extraction Petroleum Safety Framework (PSF) Gas Transmission and distribution, and LPG distribution networks Gas Safety Framework (GSF) Gas and electrical installers Safety Supervisory Bodies (SSB)	Electricity
Year regulator established	1959	2015	Economic regulation 1999 Safety Function 2007 GSF 2011 PSF	2004
Number of regulated entities/companies	99	18 200	5	28
About the regulator	We regulate pipelines, energy development and trade in the Canadian public interest.	Mission: To guarantee an individual's safety and environmental integrity with legal, procedural and cost certainty in the oil and gas sector. Vision: To be the agency that takes the Mexican oil and gas sector to be the cleanest and the safest worldwide. Values: 1) Professionalism - We are ethical, knowledgeable and experienced. 2) Transparency - What we do is public and accessible. 3) Impartiality - Decisions are made based upon objective criteria.	Performs three major functions: regulates, gas and electricity consumer and wholesale markets; regulates gas/petroleum industry with respect to safety, including upstream activities, transportation and downstream activities. The CRU also regulates the Irish Water utility provider.	The authority is responsible for regulation of the electricity and related water sector and has a statutory duty to secure the provision of electricity and related water services in all parts of Oman, including rural customers and a duty to protect the interests of customers.

	Canada	Mexico	Ireland	Oman
Duties and responsibilities	<p>The National Energy Board (NEB) regulates:</p> <ul style="list-style-type: none"> - the construction, operation, and abandonment of pipelines that cross international borders or provincial boundaries, as well as the related pipeline tolls and tariffs - the construction and operation of international power lines and designated inter-provincial power lines - imports of natural gas and exports of crude oil, natural gas liquids, natural gas, refined petroleum products, and electricity, oil and gas exploration and production activities in specified areas that are not regulated under joint federal/provincial accords. 	<p>Legal mandate: regulate; authorise; supervise.</p> <p>Responsibilities: safety; environmental protection.</p> <p>ASEA has responsibilities throughout the hydrocarbons value chain: from upstream exploration and extraction to midstream and downstream transformation, production and storage, as well as distribution and retail at petrol station level, making it a globally unique technical regulator.</p>	<p>Regulate gas network safety including transmission and distribution systems.</p> <p>Regulate petroleum (oil and gas) safety, including exploration, extraction and decommissioning (onshore and offshore) safety.</p> <p>Design and oversee safety supervisory schemes for electrical contractors and natural gas and liquid petroleum gas (LPG) installers.</p> <p>Set programme of audit and inspection of regulated entities.</p> <p>Issue safety permits for petroleum activities and safety licences to LPG undertakings.</p> <p>Promotion and public awareness of electrical and gas safety issues.</p>	<p>Regulate electricity and some aspects of the water sector.</p> <p>Secure the provision of electricity and protect the interests of customers particularly customers who have limited income, the sick and elderly.</p> <p>Secure and develop the safe, effective and economic operation of the electricity sector and to enhance the safety of the public</p>
Behavioural assets and behavioural needs				
Identified main safety culture behavioural asset	Personal accountability	Respectful work environment (Gas); Safety leadership commitment (GSF)	Respectful work environment (Petroleum); Environment for raising concerns (Gas)	Safety leadership commitment (Transmission); Work processes (Generation)
Identified main safety culture behavioural need	Enquiring attitude	Personal accountability (Gas); Work processes (GSF)	Continuous improvement (Petroleum); Safety leadership commitment (Gas)	Enquiring attitude (Transmission and Generation)

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Annex Table 5.A.2. Canada National Energy Board

Invited to participate	NOT invited to participate
All Operations Staff (Field Operations and System Operations)	Various Support Staff (Administrative Assistants)
All Energy Adjudication Staff	Communications and Engagement Staff (media relations, webmaster and staff, graphic artists, printing shop staff, translation)
All Executives and Board Members	Legal Services
All Regulatory Policy Staff	People and Knowledge (Human Resources and IT)
	Corporate Performance and Results
	Integrated Energy Information and Analysis (monitoring of energy trade date, supply, imports, etc.) (other than Reg Policy staff who are included)

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Annex Table 5.A.3. Mexico: Agency for Safety, Energy and Environment (ASEA)

	Total no. of people
Invited to participate	350
Directive Group	7
Supervision, Inspection, and Surveillance Unit	90
Permits and Authorisations Unit	130
Regulation and Legal Standards Unit	83
Planning, Processes and Strategic Stakeholder Engagement Unit	33
Executive Direction	7
NOT invited to participate	115
Finance and Administration Unit	
Legal Affairs Unit	
Administrative Assistants	
Press and Communication Office	

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Annex Table 5.A.4. Ireland: Commission for Regulation of Utilities

	Total no. of people
Invited to participate	42
Senior management	6
Frontline staff	16
Managers	4
Others (includes staff who previously worked in Safety Division, Legal advisor on safety and human resources staff with responsibility for inhouse safety)	8
External consultants who participate in inspections	8
NOT invited to participate	58
Economic regulatory managers/analysts	
Finance, IT, Communications, and administrative staff	

Note: The contact points in the companies are the safety managers. We included some economic regulatory managers and analysts in the group we planned to send surveys to. The ones left out are those with responsibility for water and energy markets.

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Survey on behavioural assets and needs

Using numbers 1-9, can you please rank the following safety culture dimensions with respect to your regulated entities (not your own organisations)? A one (1) corresponds to what you perceive as the strongest safety culture dimension in your regulated entities; a nine (9) corresponds to what you perceive as the weakest safety culture dimension in your regulated entities.

	Safety leadership commitment
	Respectful work environment
	Environment for raising concerns
	Effective safety and environmental communication
	Personal accountability
	Enquiring attitude
	Hazard identification and risk management
	Work processes
	Continuous improvement

For the strongest and weakest safety culture dimension (denoted by 1 and 9 above), can you please identify what you believe to be *behavioural assets and needs*. By “behavioural *asset*” we mean things that people do to keep their organisation safe – behaviours that need to be protected and promoted. If you placed a 1 next to “environment for raising concerns”, then an example behavioural asset could be “workers challenge their colleagues when they see colleagues’ unsafe behaviour”.

Use this box to describe 2-3 behavioural assets in line with what you perceive to be the strongest safety culture dimensions in your regulated entities...

By “behavioural *need*” we mean behaviours that must occur in order for the organisation to become safe. If you placed a 9 next to “work processes”, then an example behavioural need could be “workers need to let management know when the written processes do not match how work is done”.

Use this box to describe 2-3 behavioural needs in line with what you perceive to be the weakest safety culture dimensions in your regulated entities...

Finally, please use the space below to provide any other context you think is important for us to understand these behavioural assets and needs? Can you think of a reason why these behaviours exist (or don't)? Do these behaviours have flow-on effects to larger problems in the organisations?

A large, empty rectangular box with a thin black border, intended for the respondent to provide additional context or answers to the questions posed above.

Safety culture items

Box 5.1. Safety culture survey

A manager has people under them in the hierarchy and is responsible for directing their subordinates' work.

On average, managers in regulated entities...

- ...are committed to safety.
- ...take action on safety issues when raised.
- ...would always provide support if there is a concern about safety.

On average, the regulator...

- ...is committed to safety.
- ...has a positive influence on safety.
- ...takes action on safety issues when raised.

On average, in regulated entities...

- ...people understand how others' jobs contribute to safety.
- ...people who raise safety issues are seen as problematic.
- ...there are people whom others do not want to work with because of their negative attitude to safety.
- ...involvement of staff in safety activities is sufficient.

On average, in regulated entities...

- ...people who report safety-related incidents, near misses, hazardous conditions or occurrences are treated in a just and fair manner.
- ...voicing concerns about safety is encouraged.
- ...timely feedback is provided on the safety issues raised.

On average, in regulated entities...

- ...information about safety-related changes is clearly communicated to staff.
- ...lessons are learned from safety-related incident or occurrence investigations.
- ...there is good access to information regarding safety incidents or occurrences.
- ...there is good communication up and down the organisation about safety.

On average, in regulated entities...

- ...everyone feels that safety is their personal responsibility.
- ...there is confidence in other people.
- ...people are committed to safety.

On average, in regulated entities...

...there is sufficient staff to work safely.

...people share safety-related information.

There is a good relationship between the regulator and the regulated entities.

In regulated entities, good safety behaviour is acknowledged.

Behavioural vignettes items

Annex Table 5.A.5. Behavioural vignette items

Behavioural science principle	Level	Vignette
Messenger	Regulator	You hear that employees in the entities are learning about a new Personal Protective Equipment (PPE) guideline introduced by the regulator .
	Management	You hear that the workers in the entities are learning about a new Personal Protective Equipment (PPE) directive introduced by their senior management team .
	Peer	You hear that workers in the entities are learning about a new Personal Protective Equipment (PPE) direction introduced by their peers and colleagues , through word of mouth.
Feedback	Simple benchmarking	A regulated entity in your sector scored in the bottom/worst performing 25%* for lost-time injury rate (LTI) performance.
	Social benchmarking	A regulated entity in your sector scored in the bottom/worst performing 25%* of organisations in similar business areas for lost-time injury rate (LTI) performance. An entity in your sector scored in the bottom/worst performing 25%* of lost-time injury rates (LTI) performance. However, the entity exerted effort to improve LTI rates over the previous period and the regulator acknowledged this progress/work toward progress with a letter from the Director of the regulator.
	Control	A regulated entity in your sector scored in the bottom/worst performing 25%* for lost-time injury rate (LTI) performance and emphasises that lost-time injury performance is a known concern in your industry/organisations in your industry acknowledge the importance of the lost-time injury rate.
	Reciprocity	An entity in your sector scored in the bottom/worst performing 25%* of lost-time injury rates (LTI) performance. However, the entity exerted effort to improve LTI rates over the previous period and the regulator acknowledged this progress/work toward progress with a letter from the Director of the regulator.
Social norms	Control	A supervisor** in an entity instructs a worker to take a short-cut in procedures in order to speed up the completion of an important task. You know the entity's rule is to report risky behaviour.
	Injunctive	A supervisor** in an entity instructs a worker to take a short-cut in procedures in order to speed up the completion of an important task. You know that the entity's rule is to report risky behaviour and that 9 out of 10 workers believe that they ought to report their supervisor if they suspect the supervisor's actions are negatively affecting safety.***
	Descriptive	A supervisor** in an entity instructs a worker to take a short-cut in procedures in order to speed up the completion of an important task. You know the entity's rule is to report risky behaviour and that 9 out of 10 workers in regulated entities report their supervisors if they suspect the supervisor's actions are negatively affecting safety.***

* This 25% is used for illustrative purposes.

** A supervisor is in charge of the process and around 10-20 workers.

*** The "9 out of 10" figure is used for illustrative purposes.

Source: OECD (2018), "Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture", Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

Sample of behavioural vignettes

Your **organisation's senior management team** notices a safety issue and introduces a new directive around Personal Protective Equipment (PPE).

Annex Table 5.A.6. Sample behavioural vignette

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
This safety information from this source attracts my attention.							
This safety information from this source should affect how I do my job.							
This safety information from this source would affect how I do my job.							
This safety information from this source should affect how managers do their job.							
This safety information from this source would affect how managers do their job.							

Source: OECD (2018), “Behavioural insights and safety improvement in the energy sector: Experimental evidence from Canada, Ireland, Mexico and Oman on strengthening dimensions of safety culture”, Unpublished, Prepared by M. MacLennan and M. Tear for discussion at the 11th Meeting of the Network of Economic Regulators, 26 November 2018, Paris.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

Delivering Better Policies Through Behavioural Insights

NEW APPROACHES

As behavioural insights (BI) become more widely used, countries are looking to expand the application of the methodology to new frontiers of policy making. This report tests behaviourally informed solutions to complex policy problems in the fields of competition, consumer protection, energy consumption and safety through a variety of theoretical and experimental approaches. The results give policy makers new ideas for tackling policy problems related to individual behaviour, as well as for changing the behaviour of organisations.

Consult this publication on line at <https://doi.org/10.1787/6c9291e2-en>.

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases. Visit www.oecd-ilibrary.org for more information.

