

Dulleck, Uwe

Working Paper

Financial Decisions and Financial Regulation: Three Concepts of Performance Based Regulation

CESifo Working Paper, No. 8006

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Dulleck, Uwe (2019) : Financial Decisions and Financial Regulation: Three Concepts of Performance Based Regulation, CESifo Working Paper, No. 8006, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/215008>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Financial Decisions and Financial Regulation: Three Concepts of Performance Based Regulation

Uwe Dulleck

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

www.cesifo-group.org/wp

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

Competitively Elected Women as Policymakers

Abstract

This chapter discusses the basis for and concepts of implementing performance based measures in financial regulation. Drawing on empirical methods and insights generated in the field of behavioural economics theoretical considerations and alternative measurements of services and disclosure performance are presented. The behavioural economics inspired approaches rely on control and treatment group comparisons and come in three shapes: concepts analysing existing administrative and financial data counting the number of dominated choices; data from laboratory experiments simulating consumer decisions given provided services counting the frequency of dominated choices and products and data from randomized controlled trials/field experiments in the form of mystery shopper experiments, again counting dominated choices. Each of these are discussed in detail. These concepts differ from alternative, more traditional, approaches that would include economic analysis of incentives and or direct measures of confusion created by information provided, such as computational linguistics, to capture financial service performance.

Uwe Dulleck
Centre for Behavioural Economics,
Society and Technology (BEST)
Queensland University of Technology
Brisbane / Queensland / Australia
uwe.dulleck@qut.edu.au

This report benefited from extensive discussions and comments from Nicola Howell and Benno Torgler, providing their insights in to legal issues and the behavioural economic approach respectively. I am grateful to Jeremy Webb and Harriet Smith for editorial feedback. I would like to thank CESifo for its support and hospitality in December 2019 that allowed me to finish this work.

Contents

A	Economics and behavioural economics and the performance of financial services and products	3
	Introduction	3
	Behavioural economics: Theory and data	4
	Dominated choices – measuring confusion.....	6
	Behavioural economics	8
	Empirical methods	9
	Three sources of data – three concepts.....	10
	Alternative conventional measures	11
B	Three Concepts of Performance based regulation	14
	Concept 1: Using existing data – What have customers chosen in the past?	14
	Concept 2: Laboratory experiments – simulated choices using real world communication and online services...	18
	Concept 3: randomized controlled trials or field experiments – mystery shoppers.....	21
C	Illustration and Discussion: credit cards, savings and insurance products	25
	Credit cards	25
	Savings products	26
	Detailed case study: savings products.....	27
	Insurance products	29
	Discussion	30
	What works for which market	31
	Naming confusion audits	32
	Regulation and confusion audits	32
D	Conclusions	33
	References	34

A Economics and behavioural economics and the performance of financial services and products

Introduction

This chapter develops concepts of performance based on financial services consumer protection regulations and building on Lauren Willis' (2015, 2017a, 2017b) proposal to rely on performance based measure in financial regulation.

Regulators of financial services aim to ensure, that customers 'know before they owe' (Sunstein, 2013; Franklin, 2012). That is, markets operate such that consumers have all the information available and therefore, in principle, they are able to make informed decisions.

The challenge relates not only to whether relevant information is available (see, e.g., Jadad and Gagliardi, 1998) but also to whether this information can be processed by consumers (Simon, 1996). In particular, this applies to all consumers - not only those with specialized knowledge or training. If confusion² or 'information or choice overload'³ play a role and are abused by service providers, it is unlikely that consumers will make appropriate decisions⁴ - or at least decisions they would feel comfortable with when these decisions are analysed for them. In addition to information overload, there are, of course, a range of additional factors relevant to consumer decision making which may play a role. They include context, biases and cognitive load. In addition, on the supply side, actions can range from unintentional to intentional manipulation.

This chapter outlines regulatory approaches that do not rely on mandated content and/or mandated forms of the financial service provided. Instead of assuming behaviour of consumers to be of either rational, in the sense of the homo economicus model, or reasonable, in the sense of the reasonable person assumption in law, the approaches focus just on the choices made by consumers. The *confusion*

² Confusion can be defined as the combination of information and choice overload (Cohen 1999, Walsh et al. 2007).

³ For a meta-analysis on choice overload see Scheibehenne et al. (2010).

⁴ See for example Lee and Lee (2004), who found that online, overload leads to less satisfied, confident and more confused consumers.

audits proposed are not asking whether consumer reasoned or could have reasoned appropriately but looks instead at the choices made. This is designed to capture the performance of a financial service provision in terms of whether observed choices of consumers are transitive. It also captures whether they would stand by their choice when having access to an analysis of their decision.

The proposed approaches rely on choices made by consumers. They use data on consumer decisions measuring the extent to which financial services consumers chose are not dominated by an option from the same service provider.

The approaches chosen rely on empirical methods used in the field of behavioural economics. Specifically, they involve an experimental approach which identifies the effects of different means used to provide a service using a control and a treatment group. The appropriate choices of these groups made under the different concepts will be discussed.

Behavioural economics: Theory and data

Behavioural economics combines the normative backbone of economics with insights from the behavioural sciences (Kahneman 2011). This combination allows situations to be identified in which consumers and decision makers in general deviate from the economic benchmark of rational behaviour. To discuss a bias or deviation, we first need to define what determines a benchmark of rational behaviour. Such a discussion will also explain where the traditional economic argument that governs current financial services regulation comes from.

Basic economic theory considers choices as rational if they can be shown to fulfil a set of basic assumptions. While authors vary on the exact composition of this set and the specificity, usually the set includes: completeness and transitivity.⁵

Completeness simply says that consumers are able to make a choice between any pair of products offered to them. Transitivity requires a consistency of choice, i.e. if one option is preferred over another, then whatever other options are added, this relationship remains. Economic models may be

⁵ Any microeconomic textbook can provide an overview of these assumptions or axioms – the terminology academic economists use for these. See for example Varian 2014, Mas-Colell, Whinston, & Green 1995 or Bowles 2004.

more specific in their characterization of optimal behaviour of individuals. But in almost all cases they assume that (optimal) choices of decision makers need to be complete and transitive to be called rational.

Gilboa (2010, p. 5) is one of the few authors with a less demanding definition of rational behaviour. In his view behaviour “is rational for a given person [the decision maker] if this person feels comfortable with it, and is not embarrassed by it, even when it is analysed for him [or her]”. This definition is helpful as it is evident - and can be shown empirically - that simple choices with assumptions such as completeness and transitivity are the determinants of ‘comfort’ or ‘non-embarrassment’ of decisions that we try to justify to ourselves or others.

In law, the concept of rationality has a close counterpart in the concept of a reasonable person⁶. While a reasonable person may not have the capacity to understand and analyse all the information provided, he or she is expected to use all the information that is provided to him or her reasonably. Both the rational and the reasonable decision maker will not be confused by overprovision of information or by the use of unusual concepts.

For many choices, rational or reasonable behaviour would seem to be an acceptable point of comparison for a ‘good’ decision. Simple choices such as choosing between apples and peaches for an afternoon snack, or between a family holiday and a new TV, fall into this category. However, many financial products present a far more involved choice. Financial products do, nevertheless, have one advantage. Some characteristics of these products are easily comparable as they can be represented in monetary terms. The decision is thus cast in the simple terms of ‘more is better than less’. That is, people prefer to have as much money as possible.

Traditional theory of regulation assumes that rationality applies to these decisions as well⁷. But here the definitions lead to different interpretations of rationality. A fundamental assumption-based definition will claim that, as long as individuals have all the information, they must make a rational or

⁶ For a discussion of these two concepts and their relationship, see Sibley (1953).

⁷ For a growing literature that discusses judgement errors of individuals in the context of law and regulations, see, e.g., Sunstein (2000) and Jolls (2007).

reasonable choice. Whereas a rational individual in the sense of Gilboa's definition, will say he is comfortable with many choices given the fine print is too much for him or her to have considered in making the choice at hand.

When data relating to financial decisions are examined we ask questions about choices made by consumers and try to benchmark them against optimal choices. While this may be difficult with more complex financial products - for example superannuation funds or hybrid securities - it may be more straightforward in the case of simpler products, such as credit cards or simple saving accounts. We will look for cases where simple dominance relationships between financial product offerings can be found in a category. That is, situations are sought where the returns and benefits are at least as good as for one option compared to the other and the costs of this option are lower. Most consumers, when the two options are properly explained, would use the low cost/same or higher benefits option over the alternative option. If we observe choices to the opposite, we will see this as confusion or underperformance. Note, that we define confusion in this case as choices that cannot be rationalized easily. We do not necessarily identify the reasons for these choices. In two of our three concepts proposed, the measures can be complemented with instruments to identify reasons for these decisions. These approaches do not consider whether consumers are able to identify specific characteristics of products. Are consumers able to correctly understand product information provided? To provide a simple illustration of the case not covered by the proposed approaches, consider the 2011 floods in Queensland, Australia. Many consumers were caught out by not being aware of the difference between inundation – water raising from below - and floods – water rushing through a property. While many affected residents were not covered for inundation, they were for floods. For most people, however, these two phenomena were essentially the same. In other words, what is referred to as Brisbane flood, was largely an inundation event, against which the majority of consumers were not insured against. This type of confusion is not covered by the present proposals.

Dominated choices – measuring confusion

The three main concepts proposed in this chapter and described in Sections B, are based on violations of transitivity. In respect to financial services, a simple explanation of transitivity lies in the

observation that customer choices for financial products should follow the rule that under the assumption that all non-monetary characteristics of a product 'are equal' the consumer chooses the product that leaves more money in his or her pocket – i.e. a simple 'more is better than less' heuristic applies. That is, if the benefits of a product are the same, we expect a customer to choose the cheaper product – for example fees on a credit card or interest paid on a savings product. If the costs are the same - fees or interest on money owed - we expect a customer to choose the product that offers more benefits - interest earned on savings or or additional services such as insurance offerings. This is a quite restrictive measure as it is rare that benefits in the first case, or the costs in the second, are exactly the same. Nonetheless, it would apply when the benefits are better or the same and the costs are lower, or when costs are the same or lower and services are better.

More importantly, the concept can also apply to a wider range of the dimensions of costs and benefits.

Figure 1 illustrates the case.

INSERT FIGURE 1 NEAR HERE

Each dot in the diagram represents an offering for a financial product. For now, ignore the colours and thus we consider products that vary in only two dimensions. They are a benefit dimension varying from low to high shown on the horizontal axis and a cost dimension varying from low to high shown on the vertical axis. Our assumption for a rational choice is based on customers' preference for more benefits and lower costs, all things being equal. In this case, product A is a dominated choice given product B offers higher benefits. These could be assumed to be in the form of higher interest on savings as well as lower costs in the form of annual fees. If only these two dimensions matter when a customer chooses product A over B, it is very likely that this customer would want to change this choice if he or she has the advantages of B explained. If A and B are offerings from the same financial service provider, we could question whether the service, including advice provided, performs in the interests of the consumer given it is most likely this customer was not directed to option B. Given the

offers are from the same financial service provider and no other factors are relevant, access to branches and other service variables cannot matter.

To illustrate the restrictiveness and limitation of this approach, no such statement could apply to choices A and C. C offers lower benefits at lower costs compared to A. Thus there will be customers that prefer C to A while others will prefer the converse. Our dominance argument applies only when options are to the right *and* above of the option considered. In this diagram, C would also be dominated by B as well as by the option to the north-west of point C.

How does this argument change with additional dimensions? While each additional characteristic takes away some of the power to classify offers as dominating or dominated, the principle argument still applies. Consider the colouring of the dots. Assume that the red and the blue dots represent different options and, for now, assume red is better than blue. In this case A is still dominated by B but C cannot be compared to B. Similarly, if the third dimension has no ordering, only blue and red dots could be compared but the same basic principle would apply.

The argument made relies on the fact that many financial products have large financial dimensions – in terms of benefits, interest earned, etc. - as well as in cost, fees and interest paid. These benefits and costs can be ordered in a way that transitivity should apply. Customers who pay more than necessary or take less of a financial benefit than is available, are likely to be confused or act in a non-rational manner. That is, as Gilboa (2010) describes it, they would very likely have changed their mind if they received proper analysis and advice. This is the fact that concepts 1-3 rely on to identify performance of financial services.

In section C of this chapter we illustrate with examples that such dominance relationships seem to be frequent enough for financial products to use our measure.

Behavioural economics

Behavioural economics is an approach that applies behavioural insights – that is cognitive biases – to help people make better decisions for themselves (Thaler and Sunstein, 2009). Performance-based

regulation⁸ aims to ensure that financial service providers apply these insights to help their customers make better decisions for themselves. However, performance-based regulation can only present a case for regulating the outcomes of these decisions. On the other hand, can we apply the protocol of behavioural economics to ask financial service providers to provide evidence that their service and communication channels perform with respect to customer decision making?

In his book, ‘Simpler: The Future of government’, Sunstein (2013) suggests that government policy should always require reporting on the success of a policy intervention using an experimental – control vs. treatment group – approach. In this report it is argued that performance-based regulation should operate on the same premise. That is, financial service providers should be asked to provide evidence of the way they provide their services which enable consumers to make appropriate decisions. If not, at the very least they should provide evidence that they do not contribute to confusion of consumers. Thus, the concepts proposed in this report rely on data about choices and behaviours which have been observed. That is, they are firstly based on actual past choices made by consumers of financial service providers and which have been documented in administrative data. They are also choices made by participants in computer-based laboratory experiments that have simulated real problems. And finally, they are choices made by mystery shoppers interacting with financial service providers.

Empirical methods

Behavioural economics relies on identifying where behaviour deviates from a rational or reasonable benchmark. In particular, it relies on evidence supporting certain behavioural interventions which, in this case, can demonstrate that such interventions lead to consumer behaviour that is likely to be closer to the consumers own interest (Thaler and Sunstein 2008; Camerer et al. 2003).

The scientific method employed is one that asks for a comparison of different policies or strategies mainly in terms of both an existing ‘business as usual’ model and a new intervention. Comparing these two allows identification of changes in a policy or strategy which leads to better behaviour.

⁸ For an overview on performance-based regulation, see May (2010).

The performance-based regulation concepts presented in this report rely on this approach. It is argued that the changes to financial services regulation based on the performance concepts presented can improve the effectiveness of the system. This would be done by requesting financial institutions to provide evidence that changes they intend to make with respect to their services offered will lead to better consumer decisions as measured by these concepts.

While such an approach may at first seem unusual in terms of financial service regulation, it is not unusual in other parts of consumer law. In most countries health regulation includes provisions for asking providers of services and products to furnish evidence of safety for consumers. This extends to those covered by public health insurance and even to requiring evidence of the benefits of a medication over existing therapies. In the regulation of vehicles, not only do public regulators rely on specifications but also on the producers of these vehicles to provide evidence that vehicles are safe in expected use situations. This extends to evidence that vehicles operate within predetermined environmental limits. ‘Safety’ aspects in relation to financial products do exist as well in the regulation of financial products, in particular with respect to vulnerable consumers. (see s31A, 32A National Credit Code). The approaches presented here thus propose an evidence based approach to be applied to a broader range of financial products.

Three sources of data – three concepts

Data to inform measurements of performance can come from three sources: existing administrative reporting or accounting data; observed decisions in a controlled or (computer) laboratory environment and observed decisions in a field experiment.

Concept 1 is based on the use of existing data. (Behavioural) economists refer to this as natural experiments if some quasi-experimental variation can be observed in the conditions under which choices were made. The experiment we can employ are the choices consumers have made with one service provider. We can compare these choices among all the institution’s consumers and identify how often a consumer chose a product sub-optimally. This analysis will rely on identifying situations where the available choice set of a consumer has dominance relationships that allows such a measurement. An experimental perspective applied to this data relies on changes to the system or the

environment that are – in the best case – exogenous to decision making by the financial service provider. For example, unpredicted shocks, such as an international financial crisis (see, e.g., Puri et al. 2011) or a news report that changes demand behaviour of consumers, can serve as a quasi-natural experiment.

Concept 2 is based on simulated decision situations in a controlled environment such as a computer lab (Kagel and Roth 1995). In this situation we can apply the communication⁹ and service strategy of a financial service provider and observe the choices made. We can also study whether consumers make decisions that they understand. To incorporate an experimental approach we compare a business as usual scenario as a control group with an alternative new strategy put forward by the financial service provider. Another alternative to illustrate any deviation is to compare choices of a highly financially literate group as a control to a more representative consumer group.

Concept 3 relies on interactions in the field (Harrison and List 2004). We can vary the demand that consumers - in this case mystery shoppers (participants that are asked to shop for a certain product category) - have when they interact with a service provider. We can then observe whether the behaviour that providers show leads to appropriate choices by the consumers. Different service policies – where there are different providers - can be observed. Our experimental variation could, for example, be based on different scenarios, including the demanded financial product or background information about potential financial needs in the near future, and education levels of mystery shoppers. A natural benchmark may be the choices made by a highly financially literate customer group compared to a more normal audience.

Alternative conventional measures

How are these measures or concepts different from more traditional approaches? In discussing this question, a comparison and contrast can be made to approaches and measures that rely on economic reasoning and which explicitly cover the incentives of service providers or on measures that rely on explicitly measuring the potential creation of confusion. Confusion may be created by

⁹ See also ASIC Report Dulleck et al. (2015), <http://download.asic.gov.au/media/3040748/rep428-published-18-march-2015.pdf>

communication¹⁰ or the way services are provided. Regulation in this context would focus on capturing incentive structures or ways to evaluate potential confusion created by communication and service channel design. Such measures are not unknown: regulations now forbid the use of bonus payments to financial advisors. This is an example of an intervention informed by the analysis of incentives. Similarly, with respect to the creation of confusion, style guides for disclosure documents try to regulate using a similar philosophy to what will be discussed as a comparison in this chapter. The theory of credence goods (Darby and Karni 1973; Dulleck and Kerschbamer 2006) provides a framework with which to analyse the incentives of financial service provider employees to recommend appropriate products. An index measuring a ‘culture to confuse’ would look for indications in the incentive structure of these employees that may encourage them to confuse consumers - that is, that lead to choices not in the best interest of consumers. While such an approach could be effective, it may require more onerous regulatory interventions. A simpler approach would be to ask financial service providers to provide direct evidence of appropriate choices by consumers. Another alternative for a confusion audit would be to ask financial service providers to implement new measures of clarity, simplicity of communication and disclosure for customers. In recent times computational linguistics (Miller 2010; Vajalla et al. 2016) have been employed to achieve greater readability of documents. Such an approach could be added to content and form regulation and explicitly be used to complement consumer surveys in order to ensure consumer communication and disclosure are not confusing. If this approach takes the length of a document into account it may even provide some protection against an information overload strategy (Moy et al. 2018). This approach would ensure that the information not only meets the condition of the rational or reasonable consumer but also allows for it to be absorbed without specialist knowledge. However there is no mechanism to ensure that the information is used appropriately by the consumer. The measure would capture the performance or non-confusion of communication or disclosure provided but not the performance of

¹⁰ For an interesting case study showing how consumers cope with confusion derived by overload in information and choice, see Kasper et al. (2010).

non-confusion of the decision maker. To take this into account would still require an outcome measure
- i.e. whether this leads to better decisions by consumers.

B Three Concepts of Performance based regulation

Concept 1: Using existing data – What have customers chosen in the past?

Description of the concept

Customers continuously choose among financial service provider product offers. These choices will depend on both their specific needs and wants (i.e. their preferences) and on the way they are served by employees of the institution.

Concept 1 relies on a systematic analysis of the choices a specific institution's customers make. In particular, the aim is to measure how often consumers choose a product from this provider that is dominated either in terms of the cost to the consumer without offering additional benefits, or in terms of the benefit - without any additional costs - by another option the financial provider offers. By restricting the offers made by an individual provider, it cannot be argued that sale channels, other services of this provider or the location of branches, make a difference for customers. The only difference is that some consumers may not be aware or find a better option. It can be expected that good performance of financial services would ensure that consumers are able to choose the option that is most financially advantageous to them.

More formally, consumers may purchase a financial product - which may be a savings product or a credit card - and have several options on offer that differ in a limited number of characteristics. They might be fees, interests owed or earned and limitations on access to funds. In most cases these characteristics can be clearly ordered: higher interest earned on savings is better than low interest, higher interest owed on debt is worse than low interest owed on debt or low fees are better than higher fees. A rational or reasonable person will always prefer the better option. In terms of economic theory, if they choose the opposite they violate the assumption of transitivity as they prefer less over more.

Concept 1 counts the cases of customers choosing dominated options among the portfolio offer by a financial provider in a specific class. Looking at all consumers making a choice, **the ratio of consumers making suboptimal to optimal choices can be used as a confusion index.**

Natural experiments: How would ‘confusion audits’ work?

Financial service providers would be asked to report on consumer decisions among their product classes as well as to report on what options the financial service provider had available at the moment in time the consumer made the decision. This data would be available to financial service providers given the concept does not rely on the set of options from which consumers chose. Rather, it relies on the set of all options that would have been available from the provider at the time. Financial reporting data is likely to be available to satisfy regulatory requirements and to provide appropriate support to customers.

These decisions can then be aggregated and the confusion ratio can be calculated and compared across institutions.

The control and treatment group in this case is stylized. Thus it can be said that a neutral environment – for example a consumer picking his or her product via a well organised market survey such that provided by third party ratings – allows for optimal choices. Against this benchmark the identified ratio will show the level of deviation due to the financial provider’s particular strategy used for the majority of consumers and employing the services of the institution. Alternatively, recent developments using methods of artificial intelligence and empirical approaches to so called Big Data could be employed to identify dominance relationships in existing administrative data of financial service providers.

What would a ‘confusion auditor’ look like?

The auditor could be internal or external to a financial market regulator. The index or ratio depends purely on truthfully reported data. The auditor needs to aggregate this data (most likely treat it as commercially in confidence) and calculate the ratios, and only the aggregated data would need to be made public.

To be able to calculate the ratios, the auditor needs to classify products and identify dominance relationships. This should be doable for those with a high level of financial literacy. A university qualified auditor with some experience in empirical research will be able to conduct such an analysis.

Given that the methods and data are well defined, the audit could be easily verified by regulated entities as well as by the interested public. As for academic research, anonymous data and the code used in the analysis can be published to ensure credibility and transparency.

Observations on limitations, risks and challenges

One limitation that applies throughout the three behavioural economic approaches presented in the chapter is the reliance on the dominance relationship between different offers in a product class. Even simple products, such as a credit card or a savings contract, may have dimensions such as non-direct transfers or extras, which are of high value to the customer. The most obvious examples are frequent flyer points with airlines or other consumer loyalty programs as well as insurance components. If this is the case, the group of dominated products may become very limited. The following presents three solutions to address this issue.

First, the regulator can opt to provide an unedited audit that takes even small variations in products as separate dimensions and thus limiting the number of dominance relationships that can be identified. This will reduce the measured value of the confusion index. However, given that this measure is compared to a value set by the regulator, the regulator would be able to increase or decrease this required value to partially address this limitation. A stronger limitation is the likely strategic reaction of financial service providers. Providers are likely to issue a greater number of extras and a much wider variety in order to create seemingly non-comparable options.

Second, instead of considering all consumers and all offers in a product class, the regulator can opt to consider only consumer decisions in product classes that are clearly comparable and where dominance relationships exist. This would substantially reduce the number of data points that can be considered. However, given that this measure attempts to capture confusion that the provider may intend to create for consumers, the index should provide a reliable measure. The measure in this case would always show a lower level of performance compared to the original concept as the measure is defined as a ratio and the number of cases considered is reduced to cases where a dominated option exists. Again, the regulator can adopt the critical value that a provider needs to achieve.

Third, based on either informed experts or consumer research, some of these extras being offered can be classified as valuable and relevant or as comparable. These changes open up criticism of the choice of what still constitutes a dominance relationship that is at the heart of the approaches presented. But, as long as the auditor is transparent about the decisions made about dominance relationships, consumers will receive valuable information. Classification of non-direct transfers can be easily evaluated by a skilled auditor. This could be carried out by conducting a simple survey that classifies each extra on some scale that can be converted to a monetary equivalent. Adding to this method, a more targeted approach would include determining a consumer's willingness-to-pay for the product extra. This context sensitive method uses a decision framework in either a direct or indirect setting to derive a value for the non-monetary good. Such a technique is extensively used in microeconomic analysis (see, for example, Kopp and Pommerehne 1997).

The direct risks of this technique to customers are minimal as only past decisions are studied. The main risk may be political pressure from financial service providers, in particular large financial service providers like major banks. This is because, the larger the institution, the more likely it is that some products that are frequently sold are dominated by other products offered by the same institution. The index would make such large organisations likely to receive lower scores. An index could be scaled to acknowledge this effect.

Another limitation is that targeted regulation is required to ensure financial service providers grant access to the relevant data. For some financial products this may already be required for other aspects of financial regulation. For others, additional safeguards may be needed.

Apart from further development of this concept their dynamic limitation needs to be considered. If such a regulation is implemented it is likely to lead to strategic behaviour by financial service providers. The main challenge is that, while an index based on this concept would provide a good indication of behaviour in the past, it will lead institutions to ensure that offers do not fulfil the dominance relationships indicated. In this case, further regulation may be needed to ensure that products offered are standardized to the extent that competition takes place only on the limited set of

characteristics. The long term effect of the policy could be that financial service providers avoid offering clearly dominated products. This should be seen as a successful outcome for the index.

Concept 2: Laboratory experiments – simulated choices using real world communication and online services

Description of the concept

Laboratory experiments in economics are situations where participants make decisions in a controlled and computerised environment. Economists put a lot of emphasis on how participants are rewarded for their efforts. In particular, the decisions made should influence the rewards in a way that participants would expect (Friedman and Sunder 1994; Smith and Walker 1993; Hertwig and Ortmann 2001).

Thus, the controlled environment simulates real world decisions in term of the effects on payments, albeit on a much smaller scale. That is, it is more or less as it would happen with larger sums in the real world.

In the case of a performance index based on this concept, participants representing usual customer groups of a certain financial product, can be asked to make a decision from options the financial providers offer. The easiest illustration may be a situation where participants are asked to use the internet interfaces of financial providers to choose a product for a specific use. From these choices it can be determined whether the internet interface leads participants to the right option or not.

This concept can test for two aspects of financial service performance. Similar to concept 1, it can identify whether choices are optimal in the sense that no dominated options should be chosen. In this case, an index or ratio similar to concept 1, can be calculated based on data collected within the laboratory environment and which produces controlled decisions.

Second, whether participants understand what they chose can be part of the empirical strategy - i.e. whether information provided is interpreted correctly can be checked. Here, an index could measure the number of correctly identified characteristics of a marketing study. Participants could be enrolled either from a specific pool maintained by the auditor, or from a market research sample or those found

through an advertising campaign. They would then be asked to complete the choice experiments online or in a computer lab. Choices of the participant group will have an effect on audit results. But, at least over time, the changes in behaviours of financial service providers will be easily identifiable with this method. In general, evidence shows that the exact size of an effect is difficult to identify in a laboratory experimental setting. However, the qualitative effect - i.e. the changes that result from a change in the relevant environment - carries over, independent of other factors affecting the participant pool. After making decisions about which products to use, participants can be surveyed regarding their understanding of the product, (if this is part of the index) and, as such, can be tested for behavioural biases.

Organising the data collection, the logistics, for this index are – naturally – higher than in the first case but still within a reasonable range.

Extension: surveying participants' understanding of products

In the context of a confusion index, the choice environment of the computer laboratory allows the auditor to conduct a survey of consumers' understanding of his or her decisions. This survey would be conducted after the computer based simulated choice has been carried out and would take the form of a quiz on the characteristics of the options chosen by the participant. The questions could range from simple product characteristics: "What are the annual fees of the service you have chosen?" or "How high is the interest you earn on the invested sum?", to more specific questions asking for reasons why the participant preferred his or her choice over an alternative. The latter would help to identify whether choices were made for the correct reasons.

The survey based index can be read in a similar manner to a school test. It can be rated by the number of mistakes that are made, but alternatively by identifying critical mistakes. An example would be consistent evidence that participants confuse interest earned on credit and interest to be paid on debt for an account.

Such a survey, given that it is conducted in a simulated, computer based environment, would be ethically acceptable as the choices are part of an experiment, i.e. choices made are non-binding.

Participants can receive feedback, providing a learning opportunity.

The limitation of this survey-based measure would be that only specific product classes will be considered instead of a larger set of products.

What would a ‘confusion auditor’ look like?

Concept 2 could be implemented by a confusion auditor who is either internal or external to the financial regulator. To use the choice-based index (similar to concept 1’s index), the auditor would need financial literacy to identify cost and/or benefit dominance relationships. Some programming capacity would be needed as would a computer lab or an online facility to conduct the experiments. It would be recommended participants be paid based on their choices to ensure validity of the results.

A performance index based on measuring an understanding of products involved would require expertise to develop. It would require creation of surveys that provide insights into what participants know or understand after they have made decisions about which products to choose. Once such knowledge of the products is obtained the index again becomes an aggregate of the level of understanding the participants displayed in the survey. This would be similar to the results of a standardized exam, used in the context of education.

Observations on limitations, risks and challenges

Concept 2 will have to rely on a limited set of products to be considered at any audit. The products need to be those for which the service provision can be simulated in a laboratory environment. This may work particularly well to represent some customers whose journey to a choice is a pure online journey. For the broader public, the journey may be more extensive. In their case the experimental setting focuses only on the online offering and audits its performance. While this approach is an easy way to audit online offerings, cases where more direct, personal advice is involved, would be much more difficult to implement in an audit of this kind. Engagement with online and traditional offline communication channels, including disclosure documents, would lend themselves to this type of audit. Such engagement could be generated in virtual environments, for example by video recordings. Risks in this type of audit are limited. Most participants will feel only minor discomfort in participating and may benefit from acquiring financial knowledge.

The index is calculated, similar to that of concept 1, by counting the number of choices where the lab participant chooses sub-optimally and comparing this number to the overall number of choices.

Financial service providers may again have an incentive to artificially avoid products where such dominance relationships exist. This in itself may be more of an advantage of the concept than a challenge. Additionally, it can be asked whether such decisions correctly represent the attention consumers give to these decisions in the field. In this regard concept 2 is richer in the information it provides about service performance but lacks the real world decision making that is captured by concept 1.

The problem of identifying dominance relationships discussed in the previous section applies here as well. To some extent, this limitation can be mitigated by the product class chosen for the experiment. Auditors could clearly specify that extras are not to be considered.

Concept 3: randomized controlled trials or field experiments – mystery shoppers

Description of the concept

Randomized controlled trials or field experiments vary specific parameters in decisions where behaviours are observed (Harrison and List, 2004)¹¹. Given that a performance index or a confusion audit is about measuring the performance or behaviour of financial service providers, the experiment providing information for this audit should include a variety of customer cases. Concept 3 will achieve this by using mystery shoppers, i.e. agents, actors or simulated customers, which are instructed to shop for a certain product and engage with the financial service provider. Mystery shopping¹² is a qualitative method for observing a service interaction in a real market setting. The advantage of this method is that researchers can control the information flow on one side of the interaction and monitor

¹¹ For a practical guide to running randomized control trial evaluations, see Glennerster and Takavarasha (2013).

¹² Depending on the industry environment participants are also referred to as pseudo-clients, simulated consumers, mystery buyers, or standardized patients.

the consequent behaviour of the financial service employee on the other side¹³. This method is used extensively worldwide as a diagnostic tool and in all types of service markets.

Unlike other methods of evaluation – such as satisfaction surveys – mystery shopping captures the process of the decision making by providers (Wilson, 2001). Note, these agents could be simply briefed consumer representatives told to shop for a particular product or actors with a prepared case. The trade-off here is between more control to reveal the operation of financial service providers, versus greater insight into how these services affect information acquisition and, potentially, the nature of representative consumers' confusion. These agents will be asked to make decisions based on this interaction. If the agents are representative consumers, their decisions can then be compared to benchmark decisions made by professionals or well-trained financially literate individuals.

Similar to concept 2, these audits can be complemented with surveys or questionnaires on product characteristics now known by the agent. If agents are well informed, a memory protocol could be employed to determine whether relevant information was provided. But relying on memory protocols will have its own set of shortcomings and potential biases. These are likely to be unrelated to the audit measure. Another alternative to overcome the memory problem would be to focus on interactive online channels, chat functions, or on telephone services and keep a record of the conversations that took place.

This concept relies on agents acting as consumers and engaging with the service channel of financial service providers. Depending on the number and training of agents, decisions can be seen as a direct measurement of service performance or as measurement of the extent to which information is provided reliably. Furthermore, it is an assessment mechanism of which offers were actually considered or recommended by the financial service provider or its employees.

¹³ For further discussion on practical application of mystery shopping in financial markets see Lubin (2011, p80-89). Smith et al. present a mystery shopping experiment on performance of service provision by pharmacies.

Field experiments: how would ‘confusion audits’ work?

Concept 3 relies on a set of mystery shoppers interacting with financial service providers. A likely annual audit might pick a small set of products (say three to five), and ask participants to interact with financial service providers to inquire about the products. While the ‘needs’ of customers would be well specified in a brief these mystery shoppers receive, training them on what to look for in the interaction with the financial service provider could be avoided. If the audit is conducted this way, concept 3 could employ the strategies of concepts 1 and 2 in order to identify the ratio of optimal choices. It could also survey and document the correct understanding of characteristics of the products chosen. If untrained participants are chosen as mystery shoppers, the extension of survey questions can still be applied. In this case it may be worthwhile to employ quantitative as well as qualitative surveys. To note: there is also a discussion on this extension in Section C of the report.

Alternatively, the mystery shoppers could be actors trained to perform certain conversations. These conversations should lead to the recommendation of specific offers of the financial service provider. A brief in this case would try to prepare for a full conversation with an advisor or agent from the financial service provider. In this way, the audit focusses on how close the final recommendation is to the identified best option and whether the conversation meets performance criteria.

What would a ‘confusion auditor’ look like?

In terms of the design used to implement this concept, this would be the most resource intensive of all audits outlined. Clearly, it would need to be conducted by a specialized unit within the financial regulator or by a skilled independent agent.

While the data collection process would require detailed preparation, data aggregation would be of similar complexity to the other concepts. Thus, the preparation of briefs would require educated professionals with good to very good market expertise. Data could be collected by either trained actors or by agents that are representatives of the general population or financial services customers. In the case of actors, they would be prepared with standard briefs and a set of questions and answers to be exchanged with the financial service provider and are asked to record recommended services.

Alternatively if data is collected by representative customers, these should not receive particular

training but instructed to seek a certain product and report back what was recommended (and potentially discussed). In both cases the observations can be compared to the benchmarks.

Observations on limitations, risks and challenges

The main limitation of this process is that financial service providers may prepare for the exact design of the mystery shopper engagement given that the potential stakes in the audit are high and the observation size will be limited. This implies special safeguards will be needed to keep the exact audit cases secret and so avoid strategic behaviour by financial service providers. If this is possible, this process may provide the most effective performance measure as it studies real world cases and allows mystery shoppers to engage with the full suite of communication services provided. As indicated above, the costs of this audit may be a limiting factor. The industry backlash may be substantial and involve raising ethical concerns. Service personnel will not be aware of being subject to an audit and there will be imposed, but not planned for, costs of the mystery shopper interactions. These concerns are important with respect to the costs of the regulation. The auditing will not only require time and effort of the regulator but also of the financial service provider, taking potentially away resources that would have otherwise been devoted to customer service. These costs require a judgement on the benefit that can be derived from such a process and it can be argued that the benefits outweigh these cost if the number interactions is limited. Another concern is that the, exact data on interactions can potentially lead to consequences to provider personnel. This should be addresses with strict confidentiality arrangements to limit ethical risks.

There would necessarily be an annual change in the cases audited to prevent a learning effect. This would pose a challenge given that it limits audit comparability over time.

C Illustration and Discussion: credit cards, savings and insurance products

Credit cards

A quick review of credit card offerings of the major banks in most countries, immediately raises suspicion that some of these products must dominate others. Indeed rating agencies, see for example *Canstar* in Australia¹⁴, usually have more than one offer from a specific institution on their lists comparing similar products, even when one keeps other product characteristics largely identical. These type of lists that can be used to identify performance of consumer choice across financial provider products.

Using existing data, regulation would require the type and characteristics of cards, accounts or insurance products offered by a financial service provider at any point in time and the choices made by consumers signing up to an offer at that time. Offers that are seen as dominant, i.e. lower fees and interest rates on debt or higher benefits (holding other components of the product constant or dominating in other dimensions as well) serve as the benchmark. The performance index captures the share of new consumers that chose these ‘optimal’ offers over the total number of new credit card consumers.

An even more demanding analysis could additionally look at legacy consumers. If new products are offered which dominate old ones, this could be captured by looking at whether or not the financial institution succeeds in switching the consumer to the new and better product, or vice versa if new products are dominated by legacy products, do consumers remain with the old product.

¹⁴ Canstar is one of Australia’s biggest financial comparison side, see <https://www.canstar.com.au/> for examples and more information, including how this particular side aims to solve conflict of interest problems.

The application of concept 2 - decisions are made in a simulated computer-based controlled environment – involves asking real world participants to make choices in a laboratory environment. The performance of the online service a financial provider offers is the ratio of the participants in these experiments that choose the dominant product over the other products. Additionally, a survey of the characteristics of the product participants decide on, could measure to what extent the information provided is absorbed and understood.

The field experiment for this concept would require mystery shoppers to engage with a financial institution after being briefed about some product inquiry. For example, a case may involve a mystery shopper being told that he or she will always repay their credit card on or before the due date and to approach a financial institution with this information. Again, after the decision a survey could help to illuminate some of the understanding of these mystery shoppers about product characteristics. If a benchmark for good consumer interaction exists whether this is fulfilled can be examined.

Savings products

The offering of financial institutions to consumers which involves choosing among simple savings options is at least as large and confusing as offers for credit cards. In this case, bonus-rates for the first weeks or months, early withdrawal fees or interest payments (or fees) depending on minimum monthly savings, lead to a large set of choices. Naturally, some of these will be dominated and the methodology proposed in this report would apply.

Similar to the case of credit cards, the data needed to measure performance of a financial provider is the information regarding the choices available at the time and the decisions made by consumers. If we assume rational consumer behaviour, an index could even match ex-post saving behaviour to the option chosen. This version of the index would be more demanding and, without well-defined parameters, could overestimate the level of confusion at the time of decision making. That is, it would be based on the assumption that consumers have perfect foresight about behaviours in the time period considered.

In laboratory experiments as above, interaction with the web interface of the financial service provider would be an option to see whether participants choose reasonably in the sense that they avoid

dominated options. The necessary brief for participants could be the availability of a certain amount of funds to be saved over a fixed amount of time.

Similarly, mystery shoppers could be equipped with a clear savings plan and a protocol of how to answer questions on the likelihood of withdrawal. A post interaction survey could test awareness of hidden costs and risks involved in the choices recommended by the financial service provider which relate to uncertainty regarding interest rate payments.

Detailed case study: savings products

To illustrate the three approaches discussed in this report, the following will consider a simple savings decision. This involves offers for a cash deposit with a major Australian Bank (NAB). The following table summarizes the three products, as offered by the bank in April 2018.

Table 1: Basic Savings Products by one Australian Bank in April 2018.

Name	Base interest	Conditional interest	Fees	Conditions
Reward Saver	0.5%	2%	0	Bonus interest only paid with monthly deposit, Branch access, No ATM, No EFTPOS
iSaver	0.8%	1.1%	0	Bonus Interest only new customers, first 4 months, No branch access, No ATM, No EFTPOS
Cash Manager	1%	0%	0	Branch access, ATM, EFTPOS, only product with cheque facility
High Interest (NABtrade)	2.15%	0%	0	No branch access, No ATM, No EFTPOS

This example illustrates some of the difficulties but also the potential of the approach. All products can be compared through interest rate levels. But apart from the bottom two interest rate levels, their form of product offering can confuse consumers. Conditional interest, which may depend either on a deposit or on a new contract, make it difficult to compare products. While fees play a role for many products, in this case they are nil in all four cases.

Among this set of options, it is obvious that the NAB Trade High Interest account offer dominates the iSaver option with and without bonuses. Both accounts are the same in terms of accessing the accounts via automatic teller machines (ATM) or via vendor payment terminal access (referred to as EFTPOS in Australia). Adding assumptions, or observing consumer deposit behaviour, allows the auditor to find similar relationships for other options. For example, a consumer making a one-off deposit should choose the Cash Manager over the Reward Saver if he or she values branch access. If branch access is not important, the iSaver should be chosen if the expected time the money is left in the account is less than 2 years. For a period of more than 2 years a calculation of the interest received reveals that the Cash Manager dominates the iSaver. While the comparison of these options needs to be made carefully, the choices observed by consumers can be compared - either based on previous behaviour in a simulated/laboratory environment (concept 1), or by a mystery shopper. In this way it becomes obvious that dominance relationships between the products exist.

By limiting the considered choices only to one financial provider, we consequently have a confusion audit for this provider. In principle, an advisor of this provider should know all the options the provider offers and would inform consumers about them. Furthermore, access to branches and services offered - on and offline - are comparable, limiting arguments about differences resulting from different providers.

Survey questions for participants in simulated/laboratory choices and participants in the mystery shopper experiment (i.e. untrained individuals), can be posed regarding the interest rate they are paid for a one off deposit of \$1000 over a 12 months period. Indicators for confusion would be if consumers in the survey instrument state that base plus conditional interests is the long term interest rate in the case Reward Saver as well as the iSaver; or if they state for any product but the Cash Manager offer that ATM and EFTPOS access is included as a service.

This example is conservative in that it restricts the products considered only to those that allow daily withdrawal of the funds. Term deposits could be added as an example in order to consider the option of a one-off deposit over a 12 months period. Even more advanced financial options could be

considered, but such considerations would then clearly lead to questions as to whether products with respect to their other characteristics are comparable.

Insurance products

Insurance products are the most complex product in the examples discussed in this section. To that extent, they may serve as an example of the difficulty to generalize about these proposed measures. In this case it may be harder to identify dominance relationships, particularly given that services provided may differ between different brands offered by the same insurance company. Nonetheless, it is likely that such relationships exist and, if that is the case, the concepts presented can be applied. More importantly, these limitations apply particularly to the use of existing data where some assumptions may need to be made when specific characteristics can be seen as the same. As an example, the availability of a telephone channel to lodge a claim may be an important aspect of service delivery. This would be especially so when online and branch lodgement are offered for two products that otherwise involve the same service and coverage. For concepts 2 and 3, the brief prepared and the choices made, can be compared relatively easily. In this case we control for what consumers were asked to purchase and compare whether the chosen option is the best of that which the insurance provider has on offer at the time. Within the mystery shopping concept, even the communication can be planned and described allowing for further control of variables.

Where existing data is used, the establishment of clear dominance relationships between products becomes difficult. That is because the way in which products can differ in the insurance space are much more numerous than in the case of credit cards or savings products. This diversity will lead to fewer cases of clearly dominated products and there may be instances where a ‘sort of’ dominance can be identified. This product class, as with many more complex products, may hence require one of the other concepts to be considered. In the case where a clear dominance is present, for example where the risks covered are exactly the same but fees differ, the concept still applies and can provide some insight into financial service provision. The main problem arising here is that the number of observations may be significantly lower and hence the index would lose some of its reliability.

Laboratory experiments face a larger problem whereby services decided upon are not so much derived from a web interface or a stylized choice situation, but from an environment that relies on some form of consultation and advice. Where this advice is provided through an online channel, such as a chat function on a webpage, the concept can still be applied. However, it is not clear to what extent generalizing from these findings would be possible.

On the other hand, field experiments/randomized controlled trials using mystery shoppers do not struggle with the above limitations. Given that mystery shoppers would be briefed with what they are asked to shop for, a follow up of choices made available by the financial provider would provide a means of determining the quality of advice.

Discussion

The examples discussed in this section provide some evidence as to why the concepts are substitutes as well as complements. The more standardized the product, as in the case of savings products and to some extent credit cards, the easier it will be to derive a confusion or performance index based on existing large data sets. The more complex the product offering, the more an auditor will be required to make assumptions about consumer needs and the comparability of specific aspects of a product.

Where this applies to experimental concepts – either laboratory or field – complexity can be controlled for by carefully designing cases or briefs for participants. For the concept based on existing data there is a trade-off between the availability of consumer data and the extent to which assumptions are made about products offered by financial service providers.

Good financial services should be measured by the personalization of financial solutions for consumers. Thus, experimental methods (concepts 2 and 3) are more likely to allow for measurements of more specialized aspects of financial services. These two concepts would also allow the auditor to keep the measurement up to date in the sense that the index would depend each year on specific products to measure service performance. While this keeps the measure current, it reduces comparability over time.

What works for which market

As already indicated in discussion of the examples, the type of products will inform the concept best suited to the needs. The discussion also revealed that the use of existing data as a comparable measure for highly standardized products may be the most cost effective and reliable over time. Performance regulation in these cases can be conducted based on consumer decisions in the past and can be compared across different periods of time. For more complex and specialized solutions there is a need to control the interaction with financial service providers using experimental techniques. These will provide quite targeted feedback on performance in specific cases.

Considering the requirement of regulating based on such measures, it may be necessary to determine how comparable the measure or index needs to be over time. Regulation could require minimum performance or even improvement of performance. This may imply that cases and products to be considered have to be specified and the measures need to be fixed. While this may well be possible for some products, it will be more challenging for others, especially if we allow for strategic behaviour of financial service providers. Keeping a constant protocol over time is very likely to lead to strategic behaviour of financial service providers in the cases and product classes used for analysis. This can be seen as a learning effect. However, the index may be misleading in the sense that financial service providers may avoid confusion in the measure implemented but do not change behaviour in other classes.

Some of these shortcomings can be accommodated for by varying the products studied but normalizing performance relative to the market. That is, while a different product may lead to other index results, the requirement could be relative to the industry. For example, it may be determined that performance should be above half of the median of the performance ratio, or where feasible, above a certain minimum value.

As an extension, the concepts could also be applied to other service providers relevant to the industry. Price comparison services, such as the *Cannex* service, could be evaluated using these concepts with or without fees or commissions paid to them.

Naming confusion audits

The confusion audits discussed vary in the way the data is collected. But they all rely to a large extent on a benchmark of a rational or reasonable choice. To this extent the name *confusion index* is quite appropriate as the concepts discussed indicate that services provided yields behaviour that many consumers would want to change if they receive all information and a reasonable analysis of their decision. More importantly, the indices could easily allow creation of subindices e.g. the *savings confusion index*, the *credit card confusion index*. This would allow financial service providers to differentiate with the type of service they excel in.

The complementary information collected through concepts 2 and 3 could be well named the *consumer understanding index*. If the focus is on minimum clarity, a star system could be used where each star or tick indicates aspects of a product that a sufficient number of consumers need to understand. All of these measures would allow financial service providers to compete on the quality of their services and not just on the financial aspects of products that currently seem to be easily used to confuse consumers. Bonus interest rates for limited periods of time are a good example of the intentional use of confusion. This can be captured by an index showing the level of understanding consumers have of chosen products.

Regulation and confusion audits

The confusion audits or indices can serve as a basis for financial service regulation. This would be similar to the requirement that a newly registered car needs to fulfil certain fuel efficiency requirements. In the case of new financial products and services, they would need to pass a test which shows consumers are able to choose these products correctly and indicates the level of understanding required by regulation. The operationalisation of the indices would allow for the setting of minimum standards and the setting of protocols governing testing conditions.

D Conclusions

This chapter introduced concepts to regulate the performance of financial services. The concepts rely on the one hand on empirical approaches – natural experiments, laboratory experiments and randomized controlled trials – in the form of mystery shopper experiments - taken from Behavioural Economics, on the other hand they apply a simple concept of dominated options in choice set – which we argue applies particularly well when financial products are considered.

All of these concepts are straight forward to implement. Considering the Australian case, where financial services are regulated by the Australian Securities and Investments Commission (ASIC), the regulator would be well able to conduct such an analysis.

With respect to first concept, ASIC has powers to require financial services providers to produce books about financial products (ASIC Act s31) or financial services (ASIC Act s32A). Such powers can be used for the performance or exercise of any of ASIC's functions and powers under the corporations legislation (ASIC Act s 28(a)). Thus, using its information gathering powers to collect data for creating performance benchmarks, would be consistent with ASIC's functions and powers in relation to promoting consumer protection. Similarly, wide ranging information gathering powers are included in the credit and corporations legislation. For example, under the National Consumer Credit Protection Act 2009 (Cth) ('NCCPA'), ASIC can require a credit licensee to provide a written statement containing specified information about the credit activities engaged in by the person (NCCPA, s49).

With respect to third concept, involving the collection of data using mystery shopping or shadow shopping. ASIC has previously conducted or commissioned a number of shadow shopping exercises in the course of its investigations (see for example, ASIC Report 279 Shadow shopping of retirement advice (March 2012)). A shadow shopping exercise can be completed (or commissioned) by ASIC without needing to rely on its information gathering powers. Instead, ASIC could secure this information directly from shadow shopping participants (whether they be genuine consumers or

professional actors). This could include any relevant documents created for the purpose of the transaction (for example Statements of Advice) or provided to the ‘consumer’ during the transaction. In both cases ASIC is subject to confidentiality obligations in relation to information that it collects or receives in carrying out its functions. However it is permitted to disclose summaries of information or statistics that do not identify individual consumers (s 127(1A) ASIC Act). In addition, ASIC can disclose information with the permission of the relevant person (s 127(3A) ASIC Act).

While this discussion is limited to the Australian context, it illustrates that legal barriers to such approaches to regulation are likely not to be insurmountable and in most jurisdictions – Australia included – they could be implemented even without a change of legislation.

For an economist, such performance based regulation would allow the regulator to enable competition, by providing consumers with information on the performance of financial service providers. Even without regulating certain minimum performance levels, providing ratings on the extent to which a service provider avoids its customers to buy inferior products – in particular when limited to the set of products offered by this service provider, enables customers to exert the right level of caution when a service provider claims to be interested in the customers best interest. Compared to a regulation of incentives, informed for example by the theory of Credence Goods (Dulleck and Kerschbamer, 2006, Dulleck et al., 2011) customers are not only informed about the providers incentives for its agents but about the actual outcome for customers. Such a measure could be used by services providers to incentivise its executive to align it with – an often claimed – customer focus.

References

- Bowles, S. (2009). *Microeconomics: behavior, institutions, and evolution*. Princeton University Press.
- Camerer, C., Issacharoff, S., Loewenstein, G., O'donoghue, T., & Rabin, M. (2003). Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism". *University of Pennsylvania Law Review*, 151(3), 1211-1254.
- Cohen, M. (1999). Insights into consumer confusion. *Consumer Policy Review*, 9(6), 210-213.

Darby, M. R. & Karni, E. (1973). Free Competition and the Optimal Amount of Fraud. *The Journal of Law and Economics*, 16(1), 67-88. <https://doi.org/10.1086/466756>

Dulleck, U., & Kerschbamer, R. (2006). On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods. *Journal of Economic Literature*, 44(1), 5-42. <https://doi.org/10.1257/002205106776162717>

Dulleck, U., Kerschbamer, R., & Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review*, 101(2), 526-55.

Dulleck, U., Koessler, A.-K., Schaffner, M., & Torgler, B. (2015). Improving communication with directors of firms in liquidation, ASIC Report 428.

Gilboa, I. (2010). *Rational choice*. Cambridge, Mass: MIT Press.

Glennester, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Franklin, E. H. (2012). Mandating Precontractual Disclosure. *U. Miami L. Rev.*, 67, 553.

Friedman, D., & Sunder, S. (1994). *Experimental methods: A primer for economists*. Cambridge University Press.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009-1055.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists?. *Behavioral and Brain Sciences*, 24(3), 383-403.

Jadad, A. R., & Gagliardi, A. (1998). Rating health information on the Internet: navigating to knowledge or to Babel?. *Jama*, 279(8), 611-614.

Jolls, C. (2007). Behavioral law and economics (No. w12879). National Bureau of Economic Research.

Kagel, J. H., & Roth, A. E. (Eds.). (1995). *The Handbook of Experimental Economics*. Princeton University Press.

Kasper, H., Bloemer, J., & Driessen, P. H. (2010). Coping with confusion: The case of the Dutch mobile phone market. *Managing Service Quality: An International Journal*, 20(2), 140-160.

Kopp, R. J., Pommerehne, W. W., & Schwarz, N. (Eds.). (1997). *Determining the value of non-marketed goods: Economic, psychological, and policy relevant aspects of contingent valuation methods* (Vol. 10). Springer Science & Business Media.

Lee, B. K., & Lee, W. N. (2004). The effect of information overload on consumer choice quality in an on-line environment. *Psychology & Marketing*, 21(3), 159-183.

Lubin, P. C. (2011). *Protecting Main Street: Measuring the customer experience in financial services for business and public policy*. New York: Routledge.

- Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. New York: Oxford University Press.
- May, P. J. (2010). Performance-based regulation. jprg Paper, (2).
- Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *Accounting Review*, 85(6):2107-2143.
- Moy, N., Chan, H. F., & Torgler, B. (2018). How much is too much? The effects of information quantity on crowd funding performance. *PloS one*, 13(3), e0192012.
- Puri, M., Rocholl, J., & Steffen, S. (2011). Global retail lending in the aftermath of the US financial crisis: Distinguishing between supply and demand effects. *Journal of Financial Economics*, 100(3), 556-578.
- Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3), 409-425.
- Sibley, W. M. (1953). The rational versus the reasonable. *Philosophical Review*, 62(4), 554-560.
- Smith, V. L., & Walker, J. M. (1993). Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, 31(2), 245-261.
- Simon, H. A. (1996). *The sciences of the artificial*. MIT press.
- Smith, H., Whyte, S., Chan, H. F., Kyle, G., Lau, E. T., Nissen, L. M., ... & Dulleck, U. (2019). Pharmacist compliance with therapeutic guidelines on diagnosis and treatment provision. *JAMA network open*, 2(7), e197168-e197168.
- Sunstein, C. R. (Ed.). (2000). *Behavioral law and economics*. Cambridge University Press.
- Sunstein, C. R. (2013). *Simpler: The Future of Government*. New York: Simon & Schuster.
- Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.
- Varian, H. R. (2014). *Intermediate microeconomics: A modern approach* (9th ed.). New York: W.W. Norton & Co.
- Vajjala, S.; Meurers, D.; Eitel, A.; and Scheiter, K. (2016), "Towards grounding computational linguistic approaches to readability: Modeling reader-text interaction for easy and difficult texts" (2016). *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, 38–48.
- Willis, L. E. (2015). Performance-based consumer law. *The University of Chicago Law Review*, 82(3), 1309-1409. Retrieved from <http://www.jstor.org.ezp01.library.qut.edu.au/stable/43575200>
- Willis, L. E. (2017a). Performance-based remedies: Ordering firms to eradicate their own fraud. *Law and Contemporary Problems*, 80(3), 7-41. Retrieved from <http://link.galegroup.com.ezp01.library.qut.edu.au/apps/doc/A503308788/LT?u=qut&sid=LT&xid=7d93c831>
- Willis, L. E. (2017b). The consumer financial protection bureau and the quest for consumer comprehension. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 3(1), 74-93. 10.7758/rsf.2017.3.1.04
- Wilson, A. M. (2001). Mystery shopping: Using deception to measure service performance. *Psychology and Marketing*, 18(7), 721-734. <https://doi.org/10.1002/mar.1027>

Figures

Figure 1: Dominated choices – Product B offers higher benefits with lower costs – thus dominates products A and C. If different colours represent other product characteristics, such as brand or non-financial aspects like accessibility of services, we may argue that B only dominates A because B and C cannot be compared.

