Essays in Products Complexity, Shaping Effects And Experimental Methodology

Stefania Sitzia

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

University Of East Anglia

School of Economics

August 2010

Abstract

My doctoral thesis comprises three essays.

The first essay is an experimental investigation on how complexity affects experimental markets. Individual choice experiments show that subjects, violating expected utility theory, are complexity averse. Firms may try to exploit consumers' confusion brought about by complex products in order to charge higher prices. I run a posted market experiment using lotteries as products and I am only able to find that consumers tend to buy more when prices are high. I also find preliminary evidence for shaping effects: consumers' preferences are shaped by prices.

This second essay describes a follow-up experiment aiming: a) to check the robustness of complexity aversion; b) to test for shaping effects. The experiment consists of two parts: 18 binary lottery choices; an individual choice task with a posted offer market setup consisting of twenty periods and different pricing strategies (high-low and low-high). We find that complexity aversion is robust to individual choices but not to market changes. We identify evidence of shaping effects; therefore it may be more profitable for firms to first price high then decrease prices.

The third essay is a methodological appraisal of those particular kinds of experiments that aim to test models by implementing as many assumptions as possible. However they turn out to be tests of one behavioural assumption. My analysis shows that, in order for experiments of this kind to be informative about the model target domain, the lab environment should resemble it in some relevant respects. Conversely, if interpreted as tests of a single assumption, some experiments are too complicated and are characterised by lack of control. Some others however, even if they implement as many of the model assumptions as possible, are simple enough that they can be used as a framework to study individual behaviour.

Table of Contents

Abstract *2*

Table of Contents *3*

List of Tables *6*

List of Figures *7*

Acknowledgements *8*

CHAPTER 1: Introduction ***9***

CHAPTER 2: Does Products Complexity Matter in Experimental Retail Markets? ***19***

1. Introduction 19

2. Experimental Design 23

2.1 Products Employed 23

3. Complexity or Framing? 26

4. Experimental Structure and Implementation 31

5. Predictions in the market treatments 34

6. Experimental Results 37

7. Discussion and Conclusion 48

CHAPTER 3: Products Complexity and Shaping Effects An experimental Investigation ***50***

1. Introduction 50

PART 1: THE BINARY CHOICE TASK 51

1. Introduction 51

2. Sonsino et al.'s experiment 52

3. Binary choice task 55

3.1 Lotteries/products used 55

3.2 Experimental design. 57

4. Results 58

4.1 Overview 58

4.2 Differences between lotteries 61

4.3 Complexity aversion or noisy decision process? 62

4.4 Analysis of the noise in the decision process 66

5. Brief note on the incentive system 68

6. Summary 69

PART 2: THE INVIDUAL CHOICE TASK AND SHAPING EFFECTS 69

1. Introduction 69

2. A close-up: Anchoring 72

3. A close up: Shaping 75

4. General Outline 77

5. Experimental design 78

5.1 Earnings 80

6. Experimental Results 80

6.1 Sales and Expenditure 80

6.2 Regression analysis 84

6.2.1 Description of the models 84

6.3 Profitability Analysis 92

7. Summary 95

PART 3: CONCLUSIONS 95

CHAPTER 4: Models and Experiments – A Methodological Analysis 98

1. Introduction 98

PART 1: MODELS IN ECONOMICS AND THEIR METHODOLOGICAL STATUS 99

1. Models in economics 99

1.1 The instrumentalist approach 100

1.2 The realist approach 110

1.3 The fictionalist approach 115

1.4 Summary 120

2. Two model: Varian's model of sales and Bikhchandani et al's model of fads. 120

2.1 "A model of sales" by Hal R. Varian (1980) 121

2.2 "A model of fads, fashion, custom and cultural change as informational cascades" by S. Bikhchandani, D.Hirshleifer and I. Welch (BHW) 125

3. Which methodological approach best describes what is an economic model is? 129

PART 2: EXPERIMENTS AND MODELS IN ECONOMICS 131

1. Experiments which implement models 131

2. Methodological theories and consequences for experimental testing of models ……………………………………………………………………….........132
3. Experimental tests that implement models – two examples 135
3.1 An experimental study of price dispersion – Morgan, Orzen, Sefton............136
3.2 “Information cascades in the laboratory” by Anderson and Holt 140
4. What do we learn from experiments that implement models?.......................143
5. Experiments that implement real world features 148
5.1 “An experimental imperfect market” – E.H. Chamberlin..............................148
5.2 Schelling’s focal points experiments 150
6. General implications 152
PART 3: CONCLUSIONS 159
APPENDIX A ESSAY1 - EXPERIMENTAL INSTRUCTIONS......................................161
APPENDIX B ESSAY1 – PREDICTIONS FOR THE MARKET TREATMENT............180
APPENDIX C ESSAY2 - EXPERIMENTAL INSTRUCTIONS......................................190
APPENDIX D ESSAY2 – DETAILED DISCUSSION OF THE EXPERIMENTAL RESULTS..193
List of references..197

List of Tables

CHAPTER 2 ***19***

Table 1: Simple Products 24

Table 2: Complex Products 25

Table 3: Marginal Cost Function for each Unit of the Product Sold by Sellers 33

Table 4: Experimental Design and Number of Independent Observations 37

Table 5: Average Values of Key Variables 42

Table 6: Regression Analysis 43

CHAPTER 3 ***50***

Table 1: Sonsino et al.s Binary Choices 54

Table 2: Lotteries Used in the Binary Choice Task 56

Table 3: Summary of the Lotteries Used 57

Table 4: Choices in the Binary Task 58

Table 5: The Variable Difference 61

Table 6: Results of χ^2 Tests on Switches from Risky to Safe (Noise) and Safe to Risky (Complexity Aversion) 66

Table 7: Example Product 79

Table 8: Market Task: Summary of Treatments 81

Table 9: Average Sales by Treatment, Phase and Product 84

Table 10: Average Expenditure by Treatment, Phase and Product 85

Table 11: Phase 1 Random Effects Regressions Results (n=3840) 88

Table 12: Phase 2 Random Effects Regressions Results (n=3456) 91

Table 13: Estimated Average Profits per Treatment, Market size and α Values 95

Table 14: Mann-Whitney P Values of Greater Profits with a 'High-Low' Pricing Strategy 96

List of Figures

CHAPTER 2 ***19***

Figure 1: Mean Price Dynamics 38

Figure 2: Demand Schedules Scatter plots for Simple and Complex Products 39

Figure 3: Average Number of Units Bought by Treatment 40

Figure 4: Average Number of Units Bought when Prices are Between 75 and 95 47

CHAPTER 3 ***50***

Figure 1: Summary of Binary Choices for the Entire Dataset 59

Figure 2: Summary of Binary Choices for Group 1 60

Figure 3: Summary of Binary Choices for Group 2 60

Figure 4: Summary of Binary Choices for Group 3 60

Figure 5: Average Quantity Bought Broken down by Phase 80

Figure 6: Average Expenditure across Treatments Broken down by Phase 82

CHAPTER 4 ***98***

Figure 1: Schelling's Map of a Coordination Game 150

Figure 2: Structure of Experiments that fully Implement Models 153

Figure 3: Structure of Experiments that Implement some Features of the Real World and Some Features of Models 157

Figure 4: Structure of Experiments that are also Models of the Real World (Schelling's Experiments 159

Acknowledgements

The product of four years work is summarised in this thesis. This work would not have been possible without the help, support and guidance of my two supervisors, Daniel John Zizzo and Robert Sugden, that I would warmly like to thank. I would also like to thank Francesco Guala and Jean-Robert Tyran for their helpful comments on how to improve my work. I am also grateful to the school staff for always being kind, willing to listen and helpful whenever asked. It is a pleasure to thank my office mates and in particular those who started this long research journey with me, namely James Watson, Matt Aldrich, Maya Elliot, Oindrila De and David Rojo-Arjona. All of them have been a source of ideas and discussions but also laughter and jokes. I am also grateful to Tullio Usai and Vittorio Pelligra that encouraged me to undertake this doctorate. I thank all of my friends here and in Italy, in particular Tiziana, Andrea, Francesca, Chiara, Virna and Giuliana and finally a big thanks to all my family who have supported and put up with me since I was born, and I hope will keep doing so.

CHAPTER 1: Introduction

Since the early ‘40s economics have traditionally been, and thought to be, a science based on theoretical and econometric models. It was felt that the world was too complicated that the economist did not have the opportunity to reproduce complex economies and in general economic phenomena in a controlled environment, like for example the physicist has had. In the late ‘40s however, starting from Chamberlin (1948), experiments began making their way into economics. Since the ‘80s experimental economics has gained, if not mainstream status within economics, more and more relevance and is widely used as an investigative tool.

My thesis comprises three separate essays, two of them are experiments and the last one is a methodological appraisal of a certain kind of experiments. I will first discuss what is the main contribution in the field and then briefly discuss each essay in order. Finally, I will apply my methodological analysis to the experiments I have run and conclude suggesting possible ways in which my research can be extended.

The three essays of my thesis contribute to the more general field of study of behavioural economics that mainly uses as an investigative method experimental economics. Experiments have been used to study markets, individual behaviour and so on. Since experiments made their way through the more classical analysis of economic phenomena based on theoretical models, the behaviour of economic agents has been shown to be inconsistent to the rational principle of choice widely used in most of economic theoretical models. In particular, models assume that agents have well-behaved preferences and are self-interested. Behavioural economists and psychologists as well (e.g. Smith, 1976, Sugden, 1984 among the economists, and Kahneman et al., 1991, and Kahneman and Tversky, 1978, among the psychologists) have shown that this model is inconsistent with subjects' behaviour observed in the lab. For example experimental studies (e.g. Dubourg et al., 1994, Morrison 2000, etc.) show that there is a gap between what an individual is willing to pay to buy an object and what she/he is willing to accept in order to sell it. The difference observed in experiments should be less than what is theoretically assumed (i.e. Hanemann, 1991). The preference reversal phenomenon is another violation of the rational model. This anomaly has been first discovered by psychologists (i.e. Lichtenstein and Slovic, 1971) and then replicated by many economists (e.g. Grether and Plott, 1979). Other studies show that subjects take into account in addition to their own self-interest also altruism (e.g. Andreoni, 1990), fairness, reciprocity (e.g. Fehr and Gächter, 2000) and so on. Part of this research has been purely investigative; part of it however has led to the development of models as an alternative to the standard theory of choice used in economics (e.g. Fehr and Schmidt, 1999).

The first two essays contribute to the experimental study of economic behaviour and in particular to violations of rationality. In particular they analyse how subjects' react to complexity and to past prices. Standard economic theory does not take into account in fact either complexity or shaping effects. However, among others, as Sonsino et al. show complexity may have a relevant influence on subjects, choices. Similarly, Loomes et al. (2003) show that preferences may not be well defined as assumed by standard theory but can in fact be shaped by past experience.

The aim then of the two experiments I present in the thesis is to explore further both complexity and shaping effects. In particular the main contribution of the first essay is an investigation of complexity in a market setting, given that previous research has mainly focused on individual choices. The second experiment explores further complexity aversion in a binary choice setting and shaping effects in a market setting and the implications of such an anomaly for firms' pricing strategies.

Although the third essay seems not strictly connected to the other two, and partly this is correct, its more general content makes it relevant not only for experimental economics as a discipline but also for the experiments I ran. Here is a brief description of the essays.

The first essay is an experimental posted offer market that uses lotteries as products. The main aim of the chapter is to analyse whether complexity has any relevance in experimental retail markets. Previous research, i.e. Sonsino et al.

(2002)[1], shows evidence that subjects are complexity averse, they prefer simpler and riskier portfolios to more complex but safer ones, violating expected utility theory. However, complexity may be relevant for competition and consumer policy, it is therefore worthwhile to analyse its effects in an experimental retail market as we did although as a preliminary study we do not study the effects of competition among sellers. Our experimental markets are indeed monopolies.

Complexity may have two main effects on consumers' behaviour. Firms may try to increase the complexity of their products in order to make the consumers' valuation of the real value, real value to the consumers, more obscure. As a result of this consumers may buy at a higher price and higher quantities (*consumer exploitation effect*). Given the higher complexity of products, consumers may also react differently. This would lead them to avoid complexity (as Sonsino et al. found and Huck and Weiszäcker, (1999) among others) and therefore give rise to lower prices and quantities sold (*complexity aversion effect*). Finally we may also expect the elasticity of demand for complex products to be greater than for simpler one. The fact that complexity makes difficult for the consumer to figure out their WTP, may make them more prone to rely on the only piece of information they have, that is, the price.

Given the relevance of complexity for markets and the fact that no one has studied it in a market setting we decided to run two experiments: experiment 1 is a posted market offer market with 4 buyers and a human seller; experiment 2 is a posted offer market with a computerised seller. We used lotteries as products (the reasons why we do that will be explained later on in the relevant chapter) with different complexity. Two simple products with only 3 outcomes and the same expected value, and two complex ones derived from the simple product, with the same expected value but with 27 outcomes. As Sonsino et al. did, we measure complexity in terms of outcomes and our procedure to generate complex products is the one that Sonsino et al. used.

In the experiment with the human seller, buyers were endowed with experimental points and had the chance to buy the product on sale at the price stated by the seller. The product could have been either the simple product or the complex

[1] Later we shall refer to this paper as Sonsino et al.

one or both depending on the treatment. Each subject played with two products: a simple product and a complex one. The experiment consisted of two phases of 20 periods each. In the first phase the simple product was on sale, in the second phase we used the complex one, and vice-versa. We also ran a treatment with the simple product and the complex product simultaneously on sale.

Experiment 2 focuses on buyers' behaviour. Similar to the other experiment the products on sale were either one or both depending on the treatment. However, the purpose of this experiment is that we wanted more consistent pricing strategy to analyse whether this made any difference, in terms of complexity on buyers' behaviour. The experiment involved a computerised seller and two pricing strategies, chosen according to the prices observed in the previous experiment. One pricing strategy involved prices chosen randomly from a uniform distribution with a range from 75 to 95 (the high pricing strategy). The other pricing strategy was different in that we used a different range, from 45 to 65 (the low pricing strategy).

Overall, we do not detect any complexity aversion effects, however we do find evidence of complexity exploitation in the treatments where both products where on sale simultaneously. That is subjects tend to buy more of the complex products at higher prices than they do when the simple products are on sale. We also find preliminary evidence that subjects' preferences are shaped by past prices. Finally we find preliminary evidence that demand elasticity for complex products is greater than for simple ones.

The second essay is a follow up study of the first experiment. One of the purposes of this experiment, given the relevance of complexity for competition policy, is to understand why in our previous experiments we are not able to detect any relevant complexity aversion effects while Sonsino et al. find it in their individual choice experiment. There may be several reasons that may explain that.

In particular, complexity aversion may depend on the kind of lotteries used or on the kind of task faced by the subjects. In the first experiment subjects are simply asked how many units they are willing to buy if any at a given price. Sonsino et al. on the other hand present the subjects with a binary choice that involves lotteries of different complexity. The difference in the tasks may explain why we do not detect any complexity effects in the first experiment while Sonsino et al. do in theirs. There is also the possibility that complexity aversion is sensitive to the kind of lottery used. In the previous experiments we use not only different products with respect to Sonsino et al. but also our complex lotteries have 27 outcomes, while Sonsino et al. use a complex lottery with 6 outcomes.

In this experiment, we employ 9 lotteries with three different levels of complexity in both tasks. This allows us to check whether complexity aversion is robust to change in complexity and type of lottery used. More specifically, we use 3 groups of lotteries and three different levels of complexity. The simple lotteries are the same used in the previous experiment although scaled up to match Sonsino et al.'s lotteries expected value and Sonsino et al. simple lottery. We then employ three lotteries with 6 outcomes; one is Sonsino et al. ones, the other two are obtained from the simple lotteries we use in the previous experiment. Finally we employ three lotteries with 27 outcomes: two are the ones that we use in the previous experiment; the third one is obtained from Sonsino et al.'s simple lottery. Sonsino et al. in fact only use one level of complexity, that is, a complex lottery with 6 outcomes.

The first task of the experiment is a binary choice task with 18 choices. Subjects have to choose one lottery. The lotteries used in each choice, as explained before, have a different level of complexity. We replicated Sonsino et al.'s pattern of choices. It has to be noticed that they explain part of the results by complexity aversion and part by noise. That is, the more complex the lotteries involved, the more noise results from subjects' decision process. Our results show that overall noise does have a relevant effect but this does not allow claiming that the results are driven by complexity aversion, which can in fact be explained just by random errors. We also find that our results are robust to changes in the type of lottery used.

The second task is an individual choice framed as posted market setup with computerised sellers, which is a clear test for shaping effects that we detect in the first experiment.

The experimental design of the second part is relatively simple. There are two phases of 20 periods each. In each period subjects have to decide how many units, if any, they are willing to buy, at the stated price. In order to test systematically for shaping effects, the pricing strategies are different across treatments. In the second phase we use the same pricing strategy. In the first phase however we use different pricing strategies where prices are in some treatments lower and in others higher than in the second phase. Our results show clear evidence of shaping effects. In particular, quantities bought and expenditure are significantly higher when the pricing strategy in the first phase is lower than in the second phase and vice-versa. Our regression analysis supports these results. We conclude then that a pricing strategy, starting with high prices and then decreasing them is more profitable than the opposite strategy. Our results are robust to changes in the lotteries used.

Similarly to the results obtained in the first experiment we are not able to detect any complexity aversion effects. And this gives us more confidence to the hypothesis that maybe noise has a bigger impact on choices than complexity. Although in this experiment we do not have two products on sale simultaneously as in the previous one, where we find evidence of complexity exploitation effect. Contrary to the results obtained in our first experiment, demand elasticity for more complex products does not differ significantly from that for simpler ones.

Third essay. Given my interest in experiments, I started wondering at some point of my doctoral studies what we can actually learn from experiments and what they tell us about the real world. My focus in this essay is not on all types of experiments but only on experiments that test theoretical models by implementing almost completely their assumptions. Usually the assumptions that are not implemented are the behavioural ones: the experiment uses real people rather than theoretical entities that represent economic agents in the model implemented. The chapter is organized in two parts. The first part is a philosophical discussion of models. The second part discusses the contributions that model-implementing experiments give to our knowledge.

In order to understand what experiment-implementing models teach us about the real world (the model target domain) we first need to understand what a model is. Therefore the chapter first discusses three methodological accounts of what a model in economics is. These accounts are not exhaustive, but represent three extreme cases that provide us with a useful benchmark. The instrumentalist account sees assumptions as false hypothesis that are being used just to create models and the predictions that have to be confronted with reality. The realist account sees assumptions as false in the sense that they are not an accurate description of reality (thus idealise and abstract reality) but are meant to represent reality. The models then are also representations of reality. We then confront to reality the assertions the model does, that are claimed to be true about the target domain. If they are confirmed then the model is a good representation of the of the world. The fictionalist view sees models as idealised systems or counterfactual worlds that are created by the modeller. Assumptions are neither true nor false. What we confront to reality is the relationship of similarity between the target domain and either parts of the model, or the entire mechanism the model describes.

I then focus on two influential models: Varian's model of sales (1980) and Bikhchandani et al.'s model of fads (1992). Not surprisingly for theoretical economic models, these authors say little about the relationship between their models and the real world. They only provide a vague definition of the target domain and vague empirical claims. It is not surprising then that the philosophical theories examined can provide a reasonable account of what models do. The reason is that the modellers say so little about the model and the assumptions that there are gaps left to the interpretation of the philosopher. The consequence of this is that all three accounts provide a good account of what a model is.

The second part of the chapter explores the experimental strategy of strictly implementing theoretical models. At a close inspection all three philosophical theories discussed in the first part provide the same suggestion: if the experimental test has to teach us something about the target domain, there has to be some relevant similarities between the lab and the target domain. I focus in this part of the chapter on two experiments that are tests of the models described in the first part: Varian and Bikhchandani et al.'s. These two experiments, Morgan et al.'s experimental study of price dispersion (2006) and Anderson and Holt test of informational cascades (1997), closely implement the models they test. These experiments turn out to be to be just tests of very standard behavioural economic assumptions: MSNE and Bayesian rationality. They are not informative about the target domain but could, in principle, be informative about the behavioural assumption they test. I will argue however that, if the model implemented is so complicated that the test is not a clean test of MSNE, then it would be better to test a generic component of a model in isolation. However there are cases, such as Anderson et al.'s test of Bikhchandani, that because of the simplicity of the model they implement, that makes it test a good way to test Bayes' rule. Moreover the model, and therefore the experiment, provides a simple framework that can, and in fact has, be used to study experimentally other economic issues.

I conclude the chapter with a discussion of other experimental investigative strategies. I do this by using two case studies: Chamberlin's experiment on imperfect markets (1948) and Schelling's experiments on focal points (1960). Chamberlin implements the Walrasian auction model but does that implementing also some features of real markets. Schelling runs experiments on focal points without having a specific model to refer to. In this sense his experiments can be thought of models themselves. I will argue that those experiments are informative about the real world by virtue of relevant similarity they implement of the target domain.

At this point an interesting question that I will try to answer is whether I follow in my experiments the methodology I propose in my third essay.

The first thing to notice is that both experiments do not strictly implement any particular model. They analyse whether subjects behave rationally or their choice are affected by anomalies such as complexity aversion and shaping effects, so they can be interpreted as a test of a generic component, rationality, used in many models in economics. It could be argued that: a) the first experiment and the second part of the second experiment do implement a model, a posted offer market; b) or the model of rational choice. That may be true for the former case. However as I explain in my methodological analysis, there is nothing wrong in using a model/framework to test a principle such as rationality, especially if it allows the experimenter to have a clean test of what she/he is interested in and the framework (or model, or design) is used widely used as a framework to study economic behaviour and. Let us start by discussing firstly points a) and b) and then we will see whether the tests we run allow for a clean interpretation of the results.

Posted offer market experiments that have been widely used in the discipline to study for example the effect of fairness on price increases (e.g. Franciosi et al., 1995), the effect of advertising and quality on the efficiency of the market (e.g. Holt and Sherman, 1990) and so on. The advantage of using the same framework to study economic phenomena is that it allows for a more systematic study of the subject and also for comparability across different studies. Thus, I see the use of a posted offer market as a strength of our experimental designs. Regarding the latter case, that is, that our experiments can be seen as tests of the rationality model, my opinion is that a test of rationality can be implemented using models as frameworks, like the one that Bikhchandani et al. do. And since rationality is a general component of models, its validity is in fact assumed almost in any theoretical model, it can be tested in any target domain, including our lab environment.

A weakness of the first experiment is however that it does not allow for a clean test of complexity aversion. There may be in fact two competing effects at work; one is complexity aversion the other one is complexity exploitation. We are therefore only able to check whether there is a net complexity aversion effect. The second experiment is also a test of complexity aversion and in addition we also want to check whether subjects are affected by shaping effects. As a test of complexity aversion the same comments may apply to this posted offer market, however it is a clean test of shaping effects.

A remark that could be raised is that our lab environment is too abstract and does not reproduce any features of the target domain. The first thing to notice is that, however abstract the lab environment is, the posted offer market reproduces some features of real retail markets. The first one is that sellers decide the price for the products they want to sell and buyers decide whether or not they want to buy.

The second remark that could be raised regarding the abstractness of our design is that we used lotteries as products for several reasons (this will be explained in more detail later in the thesis). Lotteries are not similar to many products that we can find in the target domain, however, since we are analysing a generic component of many models, it can be argued that (and this is what I do in the methodological chapter) the target domain can be in principle anything, and this is because the generic component (rationality) is assumed to hold for any economic phenomena and therefore for lotteries as well. This therefore applies to the second experiment where we also employ lotteries.

The argument provided so far regarding the first experiment and the second part of the second experiment also applies to the first part of the second experiment, which is a binary choice task. This is obviously not a test of a model. It uses a design that has been used to study differences in WTP/WTA disparities, preference reversal and so on. So the same comments for the posted offer market apply here. Similarly, the same comments apply to the abstractness of the lab environment of our binary choice task.

To conclude, I maintain that our experiments follow the methodological suggestions I give in the last chapter. What we investigate is a generic component of models, the target domain can also be the abstract lab environment, we use designs/frameworks that are widely used in the discipline allowing for a systematic study of economic issues and for comparability of results with others.

Regarding possible extensions of my research, complexity aversion can be further explored using multi-period lotteries that Sonsino et al. use both in posted offer markets or binary choices. Similarly products that represent tariffs can be used instead of lotteries. Regarding further work on shaping effects, it would be interesting to analyse different pricing strategies to understand which one is the more profitable from firms' point of view. Another possible extension would be that of using different products simultaneously, or real products like chocolate bars and the like. Regarding the methodological analysis, the next step will be for me to focus on the nature of models in economics and how they help us understand the real world. The different methodological accounts I will discuss in the third essay (e.g. Maki, 1992, Sugden, 2000) can be a good description of what models do in economics, and the reason is that economics modellers are too vague in many respects about their models. In particular nothing is said about the role of the assumptions, whether they should be considered "as if" assumption or representations, albeit abstract and idealised, of reality. The target domain is only vaguely defined. The empirical claims are vague as well and are supported by casual evidence as to convince the reader of the real-world relevance of their models. I would explore models in more detail to understand whether they can be considered just interpreted mathematics.

CHAPTER 2: Does Products Complexity Matter in Experimental Retail Markets?[1]

1. Introduction

This paper presents a first experimental study of whether product complexity matters for consumers' decisions and of whether it should matter for competition and consumer policy. It is often said that consumers are confused by the complexity of products such as modern cars (Rouse, 2008), broadband (Kerven, 2001), electronic products (Bostrom, 2005) and financial services (Hughes, 2007). In the words of an IDC consumer market analyst (reported in Bostrom, 2005), "imagine replacing your TV… today it's digital or analog; 4:3 versus 16:9; direct view, rear projection or a flat. If you go for a flat: a plasma or an LCD? And the resolution: standard, enhanced, or high definition?"

Along analogous lines, due to the different combinations of product features and add-ons, the president of a consulting company who conducted a marketing study in the Phoenix area has noted that "an apple-to-apple comparison of multichannel entertainment and related products… is challenging at best and often nearly impossible for the consumer" (cited in Kerven, 2001). Similarly, modern mobile phones have a large number of features the combination of which provides some level of utility to consumers: as a result, in buying a mobile phone, consumers need to reason in terms of expected utility and distribution of possible utility values from buying a particular model of mobile phone.

Complexity may affect individual choices on how much to consume of certain goods, how much to pay for them, how much to save and how much to work. Kotlikoff and Rapson (2006) maintain that the complexity of the U.S. tax system is such that it is extremely difficult for U.S. citizens to understand how much to save or to work. The complexity of the tax systems may lead to fiscal illusion that can be exploited by policy agents. The phenomenon of fiscal illusion was first discussed by Amilcare Puviani (1903) and subsequently developed by Buchanan (1967) and Wagner (1976). The main idea behind is that governments

[1] This paper is almost completely based on Sitzia and Zizzo (2010)

purposely obscure, by making it more complicated, the tax system. Their purpose is to try to confuse tax payers in order to achieve their goals that it would not be possible if the tax system was transparent. So for example confused tax payers, that misjudge both the tax burden and the revenues collected by the government, may be induced to vote for a certain fiscal measure that would not be voted for if they were not confused. Many empirical studies have found evidence for it. So for example Clotfelter (1976) and Munley and Greene (1978) have found evidence for the revenue-complexity hypothesis. That is, the complexity of the revenue system is increased so to obscure the amount collected by the fiscus. The first experimental study testing fiscal illusion it the one carried out by Sausgruber and Tyran (2005). They test the hypothesis that individuals can correctly calculate the tax burden when taxes are direct while the burden from indirect taxation is underestimated because this form of taxation is less transparent. Their experimental results confirm this hypothesis. Complicated mobile phone pricing schedules, or even the discounts offered on ordinary goods such as milk, may be a source of confusion for the consumers that may buy more or at higher prices than they would otherwise. Liebman and Zechauser (2004) suggest that complexity, driven by nonlinear pricing, schedule complexity or frequent revisions of schedules may lead to a phenomenon that they label “schmeduling”. Not to perceive the actual pricing schedule is called ‘schmedule’. Schmeduling is then “the act of behaving as if one were facing a schmedule rather than the true schedule”. Sometimes firms have an incentive, especially when they face boundedly rational consumers, to “shroud” some charges, such as shipping costs, telephone fees charged by hotels and so on. Brown et al. (2010) show evidence that firms’ revenues increase when shrouded shipping charges increases. Confusion may also arise as a consequence of money illusion as noticed by Fehr and Tyran (2005). When agents are affected in their decisions by nominal values rather than real ones they are said to be subject to money illusion. Money illusion has been documented among others by Shafir et al. (1997) and Fehr and Tyran (2001).

We have seen so far that complexity is present in many situations that have economic relevance. More complexity may mean more confusion, and therefore the likelihood that consumer may exploited or at least exploitable by firms increases. A recent report by the U.K. Office of Fair Trade notes that the complexity of decisions may affect consumer choices and that as a result “firms may have an incentive to make consumer tasks more difficult" (Garrod et al., 2008, p. 56). This *consumer exploitation effect* has been modeled by Spiegler (2006). In this context, firms may try to confuse consumers with the purpose of charging higher prices inducing them to buy higher quantities than those they would have were they able to easily identify the value of each product in advance. Engaging in complicated descriptions of products (Ellison and Ellison, 2004) or in complicated tariff structures, such as those observed in the U.K. retail electricity market (Wilson and Waddams Price, 2007) is another form of exploitation.

We have seen so far that complexity may lead to exploitation, however complexity could lead consumers to buy less than they would otherwise. It is known that subjects in experiments dislike ambiguity in outcomes (Camerer, 1995), and Sarin and Weber (1993) found evidence that ambiguity aversion replicates in market settings. Complexity may be considered related to ambiguity insofar as the inability by the consumer to understand the value of a complex product induces ambiguity in the decision setting. Sonsino et al. (2002) found that, when faced with a choice between a simple and a complex product, in the form of a lottery, subjects tended to prefer the simple product. Their interpretation is that subjects are complexity averse. While this *complexity aversion effect* has been replicated in other individual choice experiments (Huck and Weiszacker, 1999; Sonsino and Mandelbaum, 2001), it has never been tested in a market setting, which is of course the one most relevant for consumer decisions.[2]

It is an open empirical question, therefore, whether a *net* consumer exploitation effect or the complexity aversion effect dominates in the presence of complex products, and if there is a sense in which consumers are more exploitable in the presence of complex products. A net consumer exploitation effect, or at the least the concrete possibility of consumer exploitability, would suggest scope for consumer and competition watchdogs such as the Office of Fair Trade to take product complexity into account in their investigations. Conversely the role of product complexity would seem to be overstated if we found no net evidence of any

[2] One implication of complexity aversion is that firms may find it optimal to *simplify* products. An example of this may be the success of the mono sound and as simple as it gets Tivoli Model One radio (see Triano, 2001, for a review).

effect or evidence of a net complexity aversion effect. Thus, our question has policy relevance.

An additional reason why we might expect markets for simple products and markets for complex products to differ is in the elasticity of demand. We would expect markets for complex products to have a more elastic demand than markets for simple products. The reason is straightforward: if consumers have less to rely on because information about the quality features of the product is fuzzy, then they are more likely to rely on the piece of unequivocal information which is available, namely on prices.

A final motivation of this paper is to look at whether, if consumers see a sequence of prices, their willingness to buy is a function of past prices. It has been argued that agents have unclear preferences and so their willingness to buy may be affected by anchors provided either artificially or through the operation of auction mechanisms (Ariely et al., 2003, 2006; Loomes et al., 2003). Following Loomes et al. (2003), we label these psychological mechanisms *shaping effects*. We hypothesize and test the prediction that, if buyers have experienced lower prices for a given product, they may believe that the value of the product is low and as a result they may be less willing to buy the product. If this is true, consistently pricing high is then a better strategy for firms to try to exploit consumers' uncertainty about their preferences than following a strategy with more variability in terms of mix of low and high prices.

An experimental methodology is especially useful to address these topics for two reasons. First, finding a metric for product complexity is difficult in comparing products that change over a variety of dimensions, whereas in the experimental laboratory we can precisely and unequivocally identify which product is more complex than another, and provide evidence that they are considered such based on the behavior of subjects (namely, their response time in making decisions, which we measure). We rely on the methodology by Sonsino et al. (2002) to identify separate products, in the form of lotteries in keeping with the existing research we are benchmarking our work against. These products are differentiated in complexity by a procedure that multiplies outcomes and scrambles the order they are presented in. Second, a key way product complexity is achieved in retail markets is by adding product features complicating the set of possible utility outcomes, but product features often are themselves a source of utility for consumers, and may also be a source of strategically useful horizontal product differentiation. In the same way in which in an economic model we can investigate the role of a given economic factor by controlling for other factors, in the experimental laboratory we can control for potential confounds and try to isolate a net complexity exploitation (or exploitability, or complexity aversion) effect, if such an alleged effect does exist, independently of additional factors such as the tastes of consumers or inter-firm rivalry. Controlling for the tastes of consumers is the other and more important reason (apart from comparability with the existing experimental literature) for which the choice of lotteries as products is useful in a first experiment on this topic, while we control for inter-firm rivalry by having a single seller. Achieving this can also enable us to more clearly identify eventual shaping effects.

To anticipate our key results, we find no evidence of a net complexity aversion effect, while there is evidence of shaping effects and qualified support for potential consumer exploitability if sellers play a consistent intertemporal strategy in choosing prices. Section 2 describes the experimental design, section 3 presents the results and section 4 concludes.

2. Experimental Design

2.1 Products Employed

As discussed in the introduction, we largely modeled our procedure to identify pairs of (simple, complex) products on Sonsino et al. (2002). In brief, the procedure is based on deriving compound lotteries (products) from simple lotteries (products) using small payoff perturbations in a way detailed below, and on presenting the resulting compound products using a scrambled order format. The procedure enables large changes in product complexity – due to the additional outcomes (27 rather than 3) combined with order scrambling – while making the riskiness of the products indistinguishable, and therefore controlling as much as possible for differences in preferences between products for reasons other than complexity.

Define $p_i > 0$ and $\sum p_i = 1$ as the probabilities attached to outcomes x_i. The simple lotteries, or products, L_{sj} have three possible outcomes and associated probabilities:

$$L_{sj} = (p_1, x_1; p_2, x_2; p_3, x_3)$$

We used two such lotteries (S1 and S2) as products in the experiment (Table 1).

Simple Product 1 (S1)			Simple Product 2 (S2)		
Outcomes	Probability	Results	Outcomes	Probability	Results
1	0.5	10	1	0.5	3
2	0.2	65	2	0.2	66
3	0.3	140	3	0.3	151

Table 1: Simple Products[3]

Complex products can then be generated by deriving, for any given L_{sj}, a compound lottery L_{cj} that assigns weights α, β and (1 - α - β) to the outcomes of 3 draws of L_{sj}. That is,

$$L_{cj} : L_{cj} = \alpha\, L_{sj} + \beta\, L_{sj} + (1 - \alpha - \beta)\, L_{sj}$$

which is to say that each complex product L_{cj} can be obtained by making three draws of L_{sj}, assigning weights α, β and (1 - α - β) to the outcomes of each draw, and summing up the three weighted payoffs to obtain the L_{cj} payoffs structure. To make a simple example, if L_{sj} were just a flip of a coin with 50% chance of getting 12 pounds (x_1) and 50% of getting 0 ($x_2 + x_3$), and if $\alpha = \beta = 1/3$, then L_{cj} would correspond to the compound lottery obtained by flipping the coin three times with a 50% chance of getting 4 pounds each time.

Generally speaking, L_{cj} has $3 \times 3 \times 3 = 27$ possible outcomes, and, although in principle some outcomes may yield the same payoff and so may not be separable, the simple products we chose were such that this did not occur in practice, and so there were 27 differentiated outcomes in the complex product (as opposed to the 3

[3] *Note*. Results are in experimental points.

of simple products). While the procedure might in principle mean that L_{cj} is perceived as having a different degree of riskiness relative to L_{sj}, this potential problem can be addressed as follows:

Complex Product 1			Complex Product 2		
Outcomes	Probability	Results	Outcomes	Probability	Results
1	7.50%	13.9	1	7.50%	7.44
2	3.00%	63.4	2	3.00%	64.14
3	12.50%	10	3	12.50%	3
4	2.70%	140	4	2.70%	151
5	7.50%	19.1	5	7.50%	13.36
6	2.00%	61.15	6	2.00%	61.59
7	1.80%	137.75	7	1.80%	148.45
8	3.00%	130.85	8	3.00%	140.61
9	2.00%	15.5	9	2.00%	9.3
10	4.50%	136.1	10	4.50%	146.56
11	1.80%	72.5	11	1.80%	74.5
12	3.00%	128.65	12	3.00%	138.09
13	5.00%	11.65	13	5.00%	4.89
14	5.00%	59.5	14	5.00%	59.7
15	7.50%	127	15	7.50%	136.2
16	4.50%	130.9	16	4.50%	140.64
17	1.80%	134.75	17	1.80%	145.05
18	1.20%	70.25	18	1.20%	71.95
19	1.20%	67.25	19	1.20%	68.55
20	1.20%	132.5	20	1.20%	142.5
21	0.80%	65	21	0.80%	66
22	4.50%	23	22	4.50%	17.8
23	2.00%	63.35	23	2.00%	64.11
24	3.00%	17.75	24	3.00%	11.85
25	3.00%	68.6	25	3.00%	70.06
26	5.00%	13.85	26	5.00%	7.41
27	3.00%	20.75	27	3.00%	15.25

Table 2: Complex Products[4]

(a) by choosing α and β small enough as to imply just a small payoff perturbation while still multiplying the number of outcomes; in our experiment, we chose $\alpha = 0.03$ and $\beta = 0.07$, and so our products were defined as follows

$$L_{cj} = 0.03\, L_{sj} + 0.07\, L_{sj} + 0.9\, L_{sj}$$

[4] *Note*. Results are in experimental points.

(b) as in Sonsino et al., by scrambling the order of presentation of the 27 outcomes (see Table 2),[5] thus further helping make the products undistinguishable to the buyers in terms of risk while at the same time being markedly different in terms of complexity.[6] Complex products C1 and C2, derived in this way respectively from S1 and S2, are presented in Table 2.

3. Complexity or Framing?

In the previous section we show how simple lotteries have been manipulated in order to increase their complexity. It is worth discussing how this manipulation relates to framing.

A rational agent when choosing between different options considers the outcomes (or consequences) of each option and the probabilities attached to them. The way these outcomes and probabilities are presented is irrelevant to her decision. This amounts to assume that the choice (or preference) does not change if the problem is presented in a different but equivalent way (i.e. principle of invariance). A substantial body of experimental evince has shown however that subjects, when presented with the same decision problem that is framed differently, tend to reverse their preferences. One famous example is the Asian disease problem used for the first time by Tversky and Kahneman (1981):

"Problem 1 [*N* = 1521]: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

---

[5] That is, outcomes were not presented from lowest to highest (or vice versa) but instead in a random order.

[6] This is confirmed by the fact that we did not receive any debriefing feedback suggesting that subjects perceived products different in terms of risk. A by-product of our procedure, also entailing additional complexity, was that, while the outcomes and probabilities were integer (or integer percentage) numbers in the simple lotteries, they were figures with up to two decimals in the complex lotteries.

If Program A is adopted, 200 people will be saved. [72 percent]

If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28 percent]

Which of the two programs would you favor?

Problem 2 [*N* = 1551]:

If Program C is adopted 400 people will die. [22 percent]

If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78 percent]

Which of the two programs would you favor?" (p. 453).

In this problem, when options were presented in terms of gains (i.e. lives saved, problem 1) subjects chose more frequently program A, while when the same options where framed as losses (i.e. lives lost, problem 2) subjects chose more frequently program B. This results are explained by Tversky and Kahneman by loss aversion (i.e. individuals are more sensitive to losses than to gains of the same size) which has been formilised in their prospect theory (1979) using a value function.

Framing effects have been shown to be a robust phenomenon in experimental settings (e.g. Maule 1989, Paese 1995, Highhouse and Yuce 1996) although there is also experimental evidence where no framing effects were found.

Levin et al. (1998) distinguish three types of framing, risky choice framing, attribute framing and goal framing. Relevant to this discussion is the risky choice framing of which the Asian disease problem is an example.

"This form of framing, introduced by Tversky and Kahneman (1981), is the form most closely associated with the term 'framing'. In this type of framing, the outcomes of a potential choice involving options differing in level of risk are described in different ways." (p. 150)

Levin et al. maintain that the most common, but by no means the only one, form of manipulation is the one where the options are first presented in a positive frame and then in a negative one. The most common finding is a shift in choices as the one observed in the Asian disease presented (e.g. Maule 1989, Paese 1995, Highhouse and Yuce 1996, Sher and McKenzie 2005).

An example of a decision problem that is not framed in terms of losses and gains is presented by Tversky and Kahneman (1986).

“Problem 7 (N = 88). Consider the following two lotteries, described by the percentage of marbles of different colors in each box and the amount of money you win or lose depending on the color of a randomly drawn marble. Which lottery do you prefer?

Option A

90% white	6% red	1% green	1% blue	2% yellow
$0	win $45	win $30	lose $15	lose $15

Option B

90% white	6% red	1% green	1% blue	2% yellow
$0	win $45	win $45	lose $10	lose $15

Problem 8 (N = 124). Which lottery do you prefer?

Option C

90% white	6% red	1% green	3% yellow
$0	win $45	win $30	lose $15

Option D

90% white	7% red	1% green	2% yellow
$0	win $45	lose $10	lose $15 " (p. S264).

In this decision problems the probabilities of getting prizes ranging from -$15 to $45 are the same both in problem 7 and problem 8, however the outcomes blue and yellow in A are combined in the outcome yellow in option C and the outcomes red and green in option B are combined in the outcome red in option D. This kind of framing manipulation has led subjects to reverse their choice from problem 7 to problem 8. In problem 7 the dominated option A is easily recognized and therefore not chosen. In problem 8 however the majority of subjects chose the dominated option C because less transparent.

A different framing of contingencies, that for a rational agent should not matter, has also been studied by Thaler (1999) and among others Tversky and Kahneman (1981). Here is an example taken from Tversky and Kahneman (1981):

"Problem 8 [N = 1831: Imagine that you have decided to see a play where admission is $10 per ticket. As you enter the theater you discover that you have lost a $10 bill. Would you still pay $10 for a ticket for the play?

Yes [88 percent] No [12 percent]

Problem 9 [N = 2001: Imagine that you have decided to see a play and paid the admission price of $10 per ticket. As you enter the theater you discover that you have lost the ticket. The seat was not marked and the ticket cannot be recovered. Would you pay $10 for another ticket?

Yes [46 percent] No [54 percent]" (p. 457).

These two problems should lead a rational agent to the same choice. However while 88% of the subjects in problem 8 were willing to buy the ticket, only 46% made the same choice in problem 9. This is explained by mental accounts. Subjects have an account for theatre tickets. In problem 8 the ticket account is only charged by 10 dollars while in problem 9 the account is charged by 20 dollars. This, according to Tversky and Kahneman leads to choice reversal.

Now let us go back to our manipulation. From the simple lotteries we obtain, using the method explained in the previous section, the complex lotteries. Our manipulation does not involve a change in frame if we consider the number of outcomes, prizes and probabilities. In fact in these respects the simple and the complex lotteries are substantially different. However when we scramble the order of the outcomes, which is different for the simple and the complex lotteries, we are changing the frame of the two lotteries. It has to be noticed however that the framing does not change when we compare different scrambling orders of the same complex lottery, which is in fact always the same.. As Sonsino et al. (2002) notice "the perceived complexity of a given lottery might depend on editing procedures and be subjected to framing effects" (p. 938). Their manipulation is similar to ours and therefore may lead to framing effects. As discussed before framing effects refer to any sort of switch in choices when the same problem is presented in a different way. However the underlying mechanism that triggers the change, labelled as framing effects, may be different depending on the decision problem individuals are presented with. So in some cases the psychological mechanism may be loss aversion, in some others mental accounting and so on. In our case the framing of the lotteries is such that the complex lotteries should be perceived as more complex. Therefore in this case the framing effects (if we were to observe any) would be triggered by complexity attitudes.

4. Experimental Structure and Implementation

We ran two experiments: a posted offer market experiment with three treatments (Experiment 1) and an individual choice experiment with two treatments and with a posted offer market frame (Experiment 2). A posted offer market setup corresponds to the reality of retail markets where sellers post prices and buyers simply decide whether and how much to buy at the given price. Both experiments involved the same two pairs of products (S1 and C1 *or* S2 and C2), two trial periods using an example product and four phases; each phase had 10 independent trading periods.

In Experiment 1 subjects were randomly assigned to the role either of seller or buyer while in Experiment 2 all subjects were buyers. They were handed instructions, questionnaires and consent forms. After they read the instructions they answered the questionnaire and if they had any doubts they could ask for clarification. When all the participants were ready, after they did the two trial periods, the experiment started. In the trial periods we employed an example product, which was the same across sessions and is available in appendix A with instructions. The reason why we used an example product is two-fold. Firstly, we did not want to disclose any information regarding both lotteries. Secondly, we wanted to avoid any possible anchoring effect to the outcomes occurred in the trial phase that could have affected buyers' decisions. In both experiments we used 'points' as the experimental currency (the conversion rate being 975 points to one pound).

Experiment 1 involved 3 treatments: B (Baseline), IS1 (Informed Seller with one product on sale) and IS2 (Informed Seller with two products on sale simultaneously).

The B Treatment. B is the baseline treatment and involved a posted offer market with 1 seller and 4 buyers. The roles did not change throughout the session. In each period only one product was on sale: in phases 1 and 2 the simple lottery and in phases 3 and 4 the complex product in half of the sessions, or vice versa in the other half.

The seller had to state in each period the price and the quantity at which he or she was willing to sell the product. The price but not the quantity was shown to the buyers. Buyers had to state the amount of the products they wished to buy, if any, at that price. The order in which they bought was random and determined after they had stated the number of units they wish to buy. It might therefore happen that some buyers did not have the chance to buy what they wanted if they were not the first to buy and the seller had run out of stock.

Using standard experimental methodology (e.g., Davis and Holt, 1993), sellers were given a marginal cost function for each unit they sold (see Table 3). Their profits in each period were given by the difference between revenue and cost, and cumulated across the 40 periods to give the final earnings. For experimental simplicity, sellers only produced and so paid costs over units they sold.

Buyers were given an endowment of 390 points every period that they could use to buy units of the product on sale, each corresponding to one of the lotteries displayed in Table 1 and 2, which had an expected value of 60 points each (though they were not told this). The unspent points were accumulated and part of the final earnings calculated at the end of the session. The bought units of the product were accumulated over the periods. At the end of the session there was a single draw of the simple lottery and of the complex lottery determining the final value of all the units of the simple product and of the complex product held by buyers. Buyers were paid these earnings plus those from unspent points.

The IS1 Treatment. IS1 differed from B in that sellers were more informed. The rationale for this is that we wanted to reduce the likelihood that experimental subjects playing the role of sellers would be confused by the complex products in the same way as buyers might. After all, in the real world, while it is plausible to assume that consumers may be confused, companies do know well the products they sell and they may well be aware of the possible strategic implications of this. Sellers were given six extra initial practice periods in the role of buyer, to give them the flavor of what is like to be in buyer's shoes. They were provided a products sheet containing the products in unscrambled order, unlike the way they were presented on the screen. Finally, they were neutrally provided information on possible complexity aversion and complexity exploitation effects as factors working in opposite direction the relevance or irrelevance of which was for them to decide, stressing that it was up to them to decide whether either, both or neither was worth taking into account.[7]

Unit	Cost
1st & 2nd	5
3rd & 4th	5
5th & 6th	10
7th & 8th	10
9th & 10th	47.5
11th & 12th	50
13th & 14th	52.5
15th & 16th	55
17th & 18th	57.5
19th & 20th	60
21st & 22nd	62.5
23rd & 24th	65
25th & 26th	67.5
27th & 28th	70
29th & 30th	72.5
31st & 32nd	75
33rd & 34th	77.5
35th & 36th	80
37th & 38th	82.5
39th & 40th	85

Table 3: Marginal Cost Function for each Unit of the Product Sold by Sellers[8]

The IS2 Treatment. IS2 treatment was the same as the IS1 treatment but for one difference: in each period both products (S1 and C1 *or* S2 and C2) were on sale simultaneously. Production costs were computed over all the overall amounts produced by sellers in any given period, i.e. as a function of the sum of the units of both the simple and of the complex product sold. One of the purposes of this treatment is to understand whether sellers try to make buyers' task more difficult.

[7] See the experimental instructions in appendix A for the exact phrasing. Care is required in cases such as this to avoid distorting the results (Zizzo, 2008). The frame was neutral, only provided as suggestions for subjects to take into account or not as they found best, and symmetrical between the two effects working in opposite directions.

[8] *Note*. Costs are in experimental points.

They can choose to sell only the simple product or the complex one. If they want to confuse exploitable consumers they may decide to sell only the complex product. If on the other hand buyers are risk averse they can decide to sell only the simple.

Experiment 2[10]. The key difference relative to Experiment 1 was that this was an individual choice experiment. Since there was not a seller, prices were randomly generated from a uniform distribution. Half of the subjects faced high prices ranging between 75 and 95, the other half faced low prices ranging from 45 to 65.[11] Subjects knew that prices were randomly generated, but they knew neither the range of the distribution nor that the distribution was uniform. Buyers could buy any quantity they desired at the stated price. Experiment 2 involved two treatments: IC1 (Individual Choice with one product on sale each period) and IC2 (Individual Choice with two products on sale simultaneously).

The IC1 treatment was an individual choice treatment with only one product on sale each period. As in the B and IS1 treatments, subjects faced either the simple product in phases 1 and 2 and the complex product in phases 3 and 4 or vice versa.

The IC2 treatment was an individual choice treatment with both products (S1 and C1 *or* S2 and C2) on sale throughout the session. As such, it was the counterpart of the IS2 treatment.

5. Predictions in the market treatments

In this experiment we implement a monopolistic market and a individual choice task to test for complexity attitudes. In this section we present the predictions for the market treatment. We have already stressed previously that we try to minimize as much as possible the difference in riskiness between the simple and the complex lotteries. We did that by generating complex lotteries from simple ones employing small enough α and β and by additionally scrambling the order of

[10] Instruction of Experiment 2 are provided in appendix A.

[11] The choice of range was decided after observing the average market price in several session we had already run from Experiment 1. The average observed price was 70 and we wanted to discriminate between high and low prices relative to the empirical benchmark.

the outcomes. If our strategy is effective, then the two lotteries should be considered equivalent in terms of riskiness but different just in the complexity dimension. In terms of predictions then what that matters is whether the price charged for the simple and the complex lotteries differ, or, in case the price is the same, whether the quantities sold differ.

Let us focus on the net complexity aversion effect. There would be evidence of a net complexity aversion if the price of the simple product is higher than the price of the complex one. If however the difference in prices is not significant, we might still find evidence of a net complexity aversion were the quantity bought of the complex product higher than that of the simple product. So our first hypothesis is the following:

Hypothesis 1. *There is evidence of a net complexity aversion effect if the price of the complex lottery is lower than the price of the simple one and, in case the price of both lotteries is not significantly different, the quantity traded for the complex lottery is lower than that for the simple one*.

Let us now turn to the complexity exploitation effect. Firms that want to exploit consumers' confusion brought about by complexity of the products on sale, will charge a price higher than the value those products have to the subjects. Complexity aversion and complexity exploitation effects work in opposite directions: consumers may not want to pay a price for a complex product because complexity lowers the utility of the product however, consumers may also be confused by the complexity of the products and pay a higher price for them. At any rate, if the price of the complex product is higher than the price of the simple one, we could deduce that there is evidence of a net complexity exploitation effect. If on the other hand the prices do not differ we can still have evidence of a net exploitation effect if buyers buy more they would because of the complexity of the products.

Hypothesis 2: *there is evidence of a net exploitation effect if the price of the complex lottery is higher than the simple one. In case the price of both lotteries is not significantly different, there is still evidence of a net exploitation effect if the quantity traded for the complex lottery is higher than that for the simple one*.

It would be interesting to see what price a profit maximizing monopolist should charge given different buyers' attitudes towards risk. In this section we only provide predictions for 3 cases (assuming homogeneity of risky attitudes): risk aversion; risk neutrality and risk lovingness[12].

If all buyers are risk averse, a profit maximizing monopolist should never charge a price higher or equal to 60. Risk averse agents by definition prefer the certainty equivalent of a lottery over its expected value. Therefore they are willing to pay at most a price as high as 59, just lower than the expected value of our products, which is 60. The monopolist will maximize profits selling 18 units making profits of 477. Any quantity above 18 would reduce profits, given the cost function, the profits.

If all buyers are risk neutral the price we should observe in our monopolistic market is 60. Agents that are neutral to risk are by definition indifferent between the certainty equivalent of a lottery and its expected value, therefore a profit maximizing seller would not find it optimal to charge a price below 60 because risk neutral buyers would be willing to pay a price equal to the expected value of a risky lottery, which in our case is 60.

If all buyers are risk loving the price that a profit maximizing monopolist should charge is be higher than 60. In fact by definition risk loving agents are willing to pay more than the certainty equivalent for taking the risk.[13].

It is worth noting that when we only consider risky attitudes the price of the simple product should not differ from the price of the complex one. However, when we introduce complexity attitudes, then we are back to our hypothesis 1 and 2.

---

[12] For completeness we present predictions assuming that subject are respectively risk averse, risk neutral and risk seeking however it would be more conform to experimental results to assume that subjects are risk averse (see for example Holt and Laury, 2002).

[13] Appendix B contains a more detailed, albeit informal, discussion on the predictions under different assumptions on risk and complexity attitudes.

6. Experimental Results

The experiments were run at the University of East Anglia between July 2007 and March 2008. 268 subjects (mainly students) were recruited via email, and Table 4 contains details on number of subjects and independent observations for each treatment (there were at least 12 in each case). The average earnings were £ 16.01 for around one hour and a half of work. Table 4 reports descriptive statistics on key variables.

	Treatment	Subjects	Independent observations
Experiment 1	B	80	16
	IS1	60	12
	IS2	60	12
Experiment 2	IC1	32	16 (high prices) + 16 (low prices)
	IC2	36	18 (high prices) + 18 (low prices)

Table 4: Experimental Design and Number of Independent Observations[14]

Response times of buyers and sellers, i.e. the times it takes them to make decisions, are useful as a validation exercise that complex products were indeed perceived as more complex by buyers. This can be determined by looking at the treatments where a single product was sold at a time (i.e., B, IS1 and IC1). Table 5 shows that buyers spent approximately 20% more time in dealing with complex products than in dealing with simple ones. The difference is statistically significant in a Wilcoxon test (p = 0.03).[15] In contrast, there is no statistically significant difference for sellers, and, as shown by Table 5, this is due to the fact that, when sellers were provided with additional information and training (IS1), they spent virtually the same amount of time choosing prices for both products.

RESULT 1. On average buyers, but not sellers (especially when informed), took more time making decisions in relation to complex products than they did in relation to simple products.

[14] *Note*. Experimental treatments are as defined in the main text.

[15] As we would expect, response times generally decline as the experiment progresses, though in the B, IS1 and IC1 treatments they spike up in round 21 when a new product is sold.

Prices as shown in Table 5 are only of course meaningful for Experiment 1, since in Experiment 2 they were randomly chosen by the computer as discussed in section 2. In Experiment 1 mean prices remain however fairly close in the three treatments, in the 65-70 range, with the mean price of the complex product being slightly higher but not statistically significantly so.

RESULT 2. There is no evidence of either a net product complexity aversion effect or net complexity exploitation effect in prices.

As Figure 1 shows, there is a small tendency for prices of the simple product to become lower with time (Spearman $\rho = -0.037$, p = 0.039),[16] but even towards the end of the experiment they are above the expected value of the lotteries (60 experimental points).

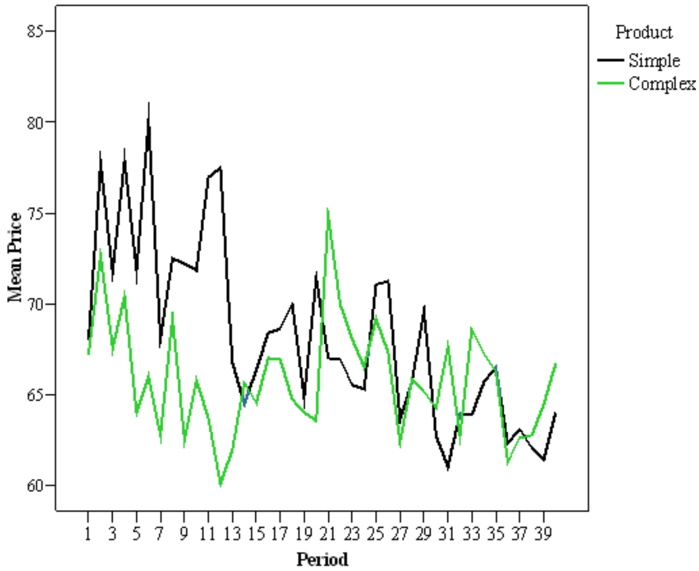


Figure 1: Mean Price Dynamics

If buyers and sellers have an understanding of how the market operates, we would expect a negative relationship between quantity and prices. Figure 2 shows that such a negative relationship exists in relation to both simple products and complex products.

A computation of elasticity coefficients shows that demand for simple products is unit elastic (- 0.99, S.E. = 0.05), whereas the demand elasticity coefficient for complex products is – 1.33 (S.E. = 0.06). As predicted, buyers are

[16] In relation to the complex product, instead Spearman ρ = -0.012 (p = 0.51).

more sensitive to changes in prices in the case of complex products than in the case of simple products.

RESULT 3. Buyers and sellers show a basic understanding of the experimental setup, as revealed by the existence of a negative relationship between quantities and prices. Market demand is more elastic for complex products than it is for simple products.

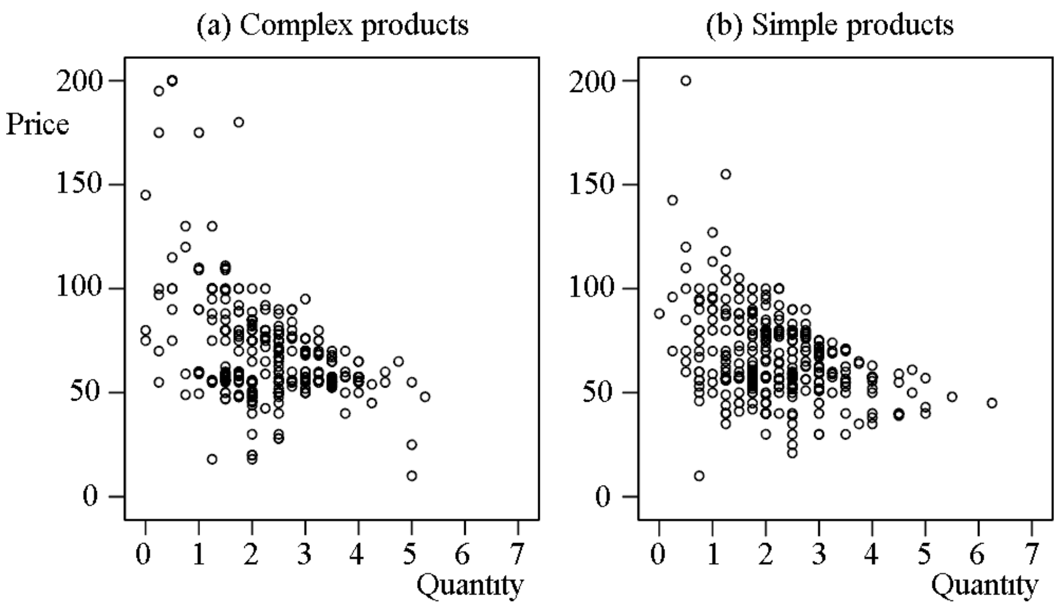


Figure 2: Demand Schedules Scatter plots for Simple and Complex Products[17]

It would be interesting to see what price a profit maximizing monopolist should charge, if she knew the exact distribution of risk and complexity attitudes among buyers, given our estimated parameters. However this would require a considerable amount of theoretical work. At the moment we provide an estimate of the highest price a profit maximizing monopolist should charge should she know the demand function observed in our market treatments, and leave the other question for future research. For this purpose, we estimate the demand function for treatments B, IS1 and IS2, we then use the average quantity bought in these

[17] *Note*: Each dot corresponds to an observed (price, quantity) combination in either of the two experiments.

treatments then we use in our regression to estimate the profit maximising price the monopolist should set.

The estimated demand function is the following:

$$Q = -0.063P + 9.41$$

The average quantity bought in the three market treatments, weighed for the number of observations is 9.36. Substituting this into our estimated demand function we find that the highest price a profit maximising monopolist should set is 79, which is higher than the average price we observe in our markets.

Table 5 shows a small discrepancy between quantities bought and quantities demanded in the case of Experiment 1. This is due to the fact that, while in the individual choice setting of Experiment 2, in Experiment 1 rationing is possible since not enough units may be available at the posted price to cover all the demand. With the help of this table we can also assess whether sellers in the IS2 treatment have some preferences towards any of the two products on sale, as they can choose to sell only one. Since the quantity demanded is greater than the quantity actually bough, the quantity bought is also the quantity supplied and sold. It can be seeing that the quantity sold both for simple and complex products is virtually the same, 1.13 for the simple and 1.12 for the complex. There is not experimental evidence that sellers try to make buyers' task more difficult.

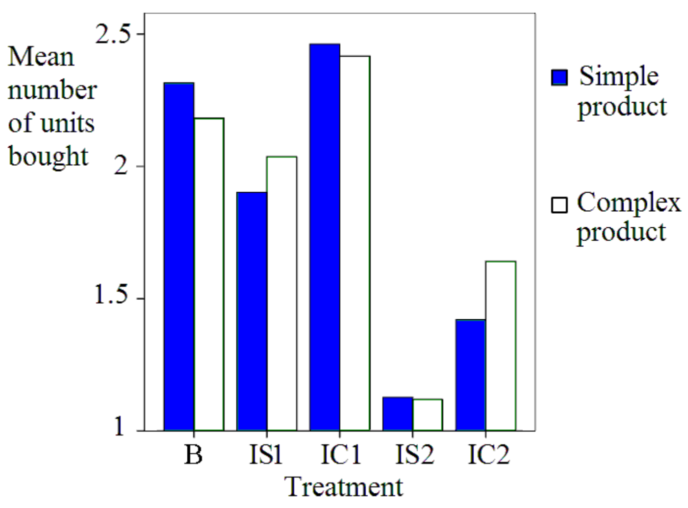


Figure 3: Average Number of Units Bought By Treatment[18]

[18] *Note*. Experimental treatments are as defined in the main text.

Table 5 also shows that the picture with quantities is mixed, with quantities demanded and bought sometimes lower and sometimes higher depending on the treatment. Figure 3 illustrates this in relation to quantities bought. In the B treatment there is a marginally statistically significant effect in the direction of net complexity aversion when one looks at quantities bought (Wilcoxon p = 0.08), but this is an artifact of rationing as quantities demanded are virtually the same.[19]

The only treatment where there is a genuinely statistically significant effect is in the IC2 treatment, and it is in the direction of a net complexity exploitation effect, with quantities demanded of the complex product being around 15% above those of the simple product (Wilcoxon p = 0.05). If we pool IC1 and IC2 treatments, we get suggestive evidence of greater quantity demanded of the complex product in both treatments as a whole (Wilcoxon = 0.08). Similar results are obtained if one looks at expenditures.

One problem in interpreting these results is the univariate nature of the statistical tests. It is possible that a net complexity aversion or complexity exploitation effect can be identified, or identified more clearly, once one controls for additional factors. We employed random effects regression models to do this,[20] controlling for the non independence of observations within each market session.

Table 6 presents the estimates of four regression models. Two have the (average) quantity bought each period, whereas the other three have the (average) quantity demanded each period as the dependent variable. The models also differ in the proxy we use for past prices.[21] The reason why these may be important is because of the shaping effects we mentioned as possible in the introduction: subjects' preferences may be shaped by past prices (e.g., Loomes et al., 2003). Depending on the regression model, we use one of two proxies for shaping: lagged

---

[19] Unsurprisingly, the one effect of rationing we find in Experiment 1 is that subjects that are rationed in one period are more likely to demand more the following period.

[20] Sashegyi et al. (2000) argue that, for this kind of data, where observations over time are taken for different group of subjects, an econometric model must control both for intra-cluster correlation and intra-individual correlation within the same cluster. For our data, panel models are the most appropriate, and specifically more appropriate than spatial models (such as error clustering). See Baltagi (2006) for further discussion.

[21] Since proxies for past prices are used in the regressions, only observations from period 2 are included.

average price (LagAvPrice), which is the *average* of the prices observed in the market for a given product from period 1 to period *t*-1, where *t* is the period of play, or alternatively lagged minimum price (LagMinPrice), which is the minimum price experienced so far, i.e. the *minimum* of the prices observed in the market for a given product from period 1 to period *t*-1.

		Experiment 1			Experiment 2	
Variable	Product	B	IS1	IS2	IC1	IC2
Price	Simple	70.00	70.27	66.15	70.31	70.00
	Complex	66.97	66.15	65.48	70.40	69.94
Quantity Bought	Simple	2.32	1.88	1.13	2.46	1.42
	Complex	2.18	2.09	1.12	2.42	1.64
Quantity Demanded	Simple	2.72	2.32	1.41	2.46	1.42
	Complex	2.70	2.44	1.30	2.42	1.64
Expenditure	Simple	148.78	113.21	71.61	156.49	89.77
	Complex	137.58	120.97	69.41	156.93	103.65
Response Time Buyers	Simple	12.88	10.34	-	11.61	-
	Complex	15.63	12.72	-	14.06	-
Response Time Sellers	Simple	20.75	19.02	-	-	-
	Complex	23.31	19.09	-	-	-

Table 5: Average Values of Key Variables

If past prices matter, then the dynamic pricing strategy employed by firms may matter. Such dynamic strategy, however, differs between Experiment 1 and 2. In Experiment 2, prices are consistently fixed either in a low range or in a high range, whereas prices in Experiment 1 are much more spread all over the place. The average observed variance in the computer generated prices chosen by each firm in Experiment 2 is 34.66 in the low price distribution and 33.04 in the high price distribution.[22] The median is respectively 34.64 and 34.65. Conversely, in the market treatments, the average variance of prices chosen by each firm is 384.22 and the median is 90.14. The variance is significantly higher in the market treatments than that in the individual choice treatments (F test, p = 0.00). As a result, if we

[22] These values are computed by looking at each set of prices chosen for each product by each firm during a session.

assume that preference shaping will occur more in relation to complex products since buyers find it more difficult to ascertain the true value of these products, consumers may be exploitable in Experiment 2 in a way they are not in relation to Experiment 1, since they are faced with a more systematic pricing strategy in Experiment 2. This leads to include not only an IC dummy variable taking the value of 0 in Experiment 1 and 1 in the individual choice Experiment 2, but also to include interaction terms of this variable with the past price variable in each regression model, i.e. LagAvgPrice × IC and LagMinPrice × IC.

Dependent Variable:	Quantity Bought			Quantity Demanded		
	Model 1			Model 2		
Regressors	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
Price	-0.020	0.001	0.000	-0.036	0.001	0.000
IC	2.985	0.519	0.000	1.483	0.568	0.009
PricexIC	-0.104	0.004	0.000	-0.088	0.005	0.000
LagAvPrice	0.003	0.002	0.082	0.007	0.002	0.000
LagAvPricexIC	0.064	0.007	0.000	0.064	0.008	0.000
Complex	-0.041	0.034	0.222	-0.037	0.042	0.380
ComplexxIC	0.185	0.058	0.002	0.197	0.076	0.009
Period	-0.006	0.001	0.000	-0.010	0.002	0.000
Constant	3.124	0.114	0.000	4.239	0.215	0.000
Dependent Variable:	Quantity Bought			Quantity Demanded		
	Model 3			Model 4		
Regressors	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
Price	-0.020	0.001	0.000	-0.035	0.001	0.000
IC	4.441	0.337	0.000	2.882	0.491	0.000
PricexIC	-0.101	0.004	0.000	-0.083	0.005	0.000
LagMinPrice	0.005	0.001	0.000	0.007	0.002	0.000
LagMinPricexIC	0.042	0.005	0.000	0.042	0.007	0.000
Complex	-0.058	0.034	0.089	-0.060	0.042	0.156
ComplexxIC	0.204	0.059	0.001	0.218	0.076	0.004
Period	-0.005	0.001	0.000	-0.009	0.002	0.000
Constant	3.051	0.106	0.000	4.374	0.193	0.000

Table 6: Regression Analysis[23]

As a result, if we assume that preference shaping will occur more in relation to complex products since buyers find it more difficult to ascertain the true value of these products, consumers may be exploitable in Experiment 2 in a way they are

[23] *Notes*: n = 12168; data from period 2 onwards included; random effect regressions used controlling for the non independence of each time series of observations in relation to each session (Experiment 1) or subject (Experiment 2).

not in relation to Experiment 1, since they are faced with a more systematic pricing strategy in Experiment 2. This leads to include not only an IC dummy variable taking the value of 0 in Experiment 1 and 1 in the individual choice Experiment 2, but also to include interaction terms of this variable with the past price variable in each regression model, i.e. LagAvgPrice × IC and LagMinPrice × IC.

Additional variables we have in the regression models include the current price (Price), a Complex dummy equal to 1 for complex products and 0 for simple ones and the period number between 2 and 40 (Period). We also include Price × IC and Complex × IC interaction terms.

The regressions in Table 6 confirm the existence of a significant negative relationship between price and quantity, as already discussed. IC has a significant positive coefficient, implying that more is bought in the individual choice treatments. In particular, in model 1 buyers buy as many as 3 units more in IC than in the others and in model 3 they buy about 4 more units than in the others. The size of the estimates of these coefficients is puzzling. It seems from table 5 that the difference in the units bought between IC and the other treatments is far more less than what estimated in the regressions. A possible explanation for this is the following. Table 5 reports the average quantity bough per treatment. However the average price per treatment is different. It is higher in the IC treatments (on average 70 both for complex and for simple) than it is in experiment 1 (for the complex product the price is on average 66 and for the simple one is about 66 in the treatment IS2 and 70 in the B and IS1 treatments). This means that, if the prices were on average the same both in the market treatments and in the IC treatments, that is 66 in most cases, then subjects would by as many more units as the size of estimate of the IC coefficients than they would in the other treatments. It is worth noting that the estimates of these coefficients vary greatly across models. Let us discuss first the difference in the estimates of the coefficient between model 1 and model 2. In model 1 the dependent variable is the quantity bought while in the other is the quantity demanded. The results show that in model 1 the estimate of the IC coefficient is 2.985 while in model 2 its estimate is 1.483. The fact that this last coefficient is smaller is not surprising. In fact, the quantity demanded in the treatments where rationing is possible (i.e. B, IS1 and IS2) is greater than the quantity bought. However in the IC treatments the quantity demanded is the same as the quantity bought. Therefore, the difference in quantity demanded between the IC treatments and the others must be smaller than the difference in quantity bough between the same treatments. This is reflected in the smaller size of the IC coefficient in model 2 relative to the size of the same coefficient in model 1. The same applies when we compare the size of the coefficients in model 3 and 4. It is however more puzzling the difference in size of the coefficient of the variable IC when we compare respectively model 1 and model 2,model 3 and model 4. Here the difference must lie in the some kind of interaction between the IC variable and the lagged variables, since these are the only ones that change.

The variable Price × IC also has a significant coefficient, implying greater sensitivity to the price observed in the individual choice treatments of Experiment 2. The coefficient on Period is negative and strongly significant: the quantity bought and demanded decreases with time.

While the coefficient on Complex is generally not statistically significant,[24] the one on Complex × IC shows a robust and statistically significant positive coefficient: on average, controlling for the other regression variables including the product price, buyers bought more of the complex product than of the simple one. This result chimes with the earlier result of a net complexity exploitation effect at least in the context of the IC2 treatment.

The coefficients of LagAvgPrice and LagMinPrice are positive and statistically significant, showing evidence of a shaping effect across both Experiments 1 and 2. The interpretation of the coefficient of LagAvgPrice is as follows, the higher the average price observed in the last period the greater the quantity bought. So If the average price is used to as a reference price, then the higher the reference price the higher the quantity bought. Subjects that observe a high average price can be thought of as assigning more value to the product than they would otherwise. Suppose there are two markets, one with a low average price (e.g. 15) and one if a high average price (e.g 40). Subjects in the first market will buy more than subjects in the second one for a given price (e.g. 40). The interpretation of the coefficient of the variable LagMinPrice is similar. The higher the minimum price observed in the market the greater the quantity bought. To

[24] Regression model 3 is the only one where there is evidence of marginal statistical significance (only at the $P < 0.1$ level).

explain this consider the following example. Suppose there are two markets. In one market the minimum price observed is 10 while in the other is 30. It is more likely that subjects buy more in the second market than in the first for a given price. The reason is that subjects' valuation of a good depends on the prices they observe in the past. The smaller the price the less value they assign to the product and therefore the less they buy and vice versa. The coefficients of these variables however are smaller than those on the interaction terms LagAvgPrice × IC and LagMinPrice × IC. Shaping occurs more strongly in the individual choice treatments than in the market treatments. The reason why this is the case can be explained by the fact that in the IC treatments the pricing strategy is more consistent. In these treatments it is possible for subjects to form a clear reference price that affects their purchasing behaviour. In the treatments with real sellers on the other hand the price vary greatly, as noticed before. It is therefore less likely, if at all possible, for subjects to form a reference price that can be used to evaluate any deviation from it. For this reason probably shaping it is more marked in the IC treatments than in the others.

RESULT 4. Preference shaping occurs. Controlling for other factors such as current prices, past experience with prices influences consumers' willingness to purchase products.

RESULT 5. There is no statistically significant evidence of an aggregate net complexity aversion effect, while there is some evidence of an aggregate net complexity exploitation effect in quantities bought and demanded with the computer generated pricing strategies of Experiment 2, especially in relation to the IC2 treatment.

To verify further the extent to which consumer exploitability is a possibility, it might be helpful to consider quantities bought when there are high prices. For comparability between Experiment 1 and Experiment 2, we consider prices between 75 and 95, as high prices in Experiment 2 are randomly generated within this range only.

Figure 4 in the following page shows that quantities bought are virtually the same in the B and IS1 treatments, while they nudge in the direction of consumer exploitation in the remaining Experiment 1 treatment (IS2) and in both Experiment 2 treatments.[25]

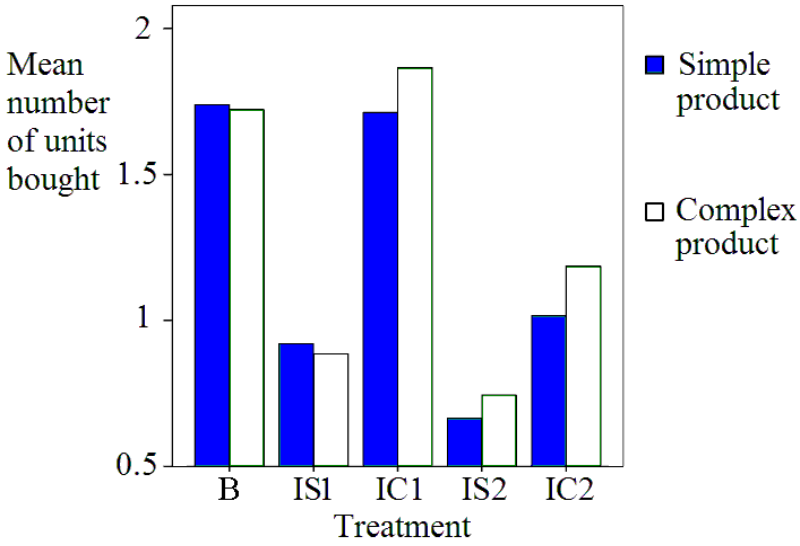


Figure 4: Average Number of Units Bought when Prices are Beteween 75 and 95[26]

When the high pricing strategy is used systematically, as in the individual choice treatments IC1 and IC2, the quantity bought of complex lottery is significantly higher than that for simple one (Wilcoxon p = 0.04). The difference is also significant if we consider the treatments where the products were on sale simultaneously (IS2 and IC2, Wilcoxon p = 0.08). If we consider IC1, IC2 and IS2 altogether, the statistical significance increases (Wilcoxon p = 0.02). These results are confirmed if we look at the difference in the expenditure between complex and simple product.[27]

[25] Due to the negative relationship between prices and quantities (Result 3 above), of course mean quantities in Figure 4 are generally lower than those in Figure 3, as they refer to high prices only.

[26] *Note*. Experimental treatments are as defined in the main text.

[27] The expenditure is significantly higher for the complex lottery than for the simple one both in the individual choice treatments (IC1 and IC2) and in treatments where both products are on sale simultaneously (IS2 and IC2) (Wilcoxon p = 0.03 and 0.08, respectively). If we consider all three treatments the statistical significance lies between the two (Wilcoxon p =*0.04*).

RESULT 6. There is evidence for some potential consumer exploitability: when prices are high, it is possible for firms to exploit consumers into buying more of the complex products than they would otherwise (i.e., were they more certain about their value as in the case of simple products).

7 Discussion and Conclusion

The key motivation of our experiments was to provide a first preliminary study on whether product complexity matters in experimental retail markets. This required us to identify a metric of product complexity that controls for consumer preferences, and we relied on Sonsino et al. to construct a procedure enabling us to do so. Our metric translates product complexity in an inability to understand what the value of the product is, which can be justified in terms of combinations of possible utility outcomes that can be obtained by multiple product features. Our procedure for identifying product complexity was validated by the longer time buyers spent in making decisions for complex products than for simple ones, though undoubtedly future research may wish to consider other ways of varying product complexity building on this work.

While Sonsino et al. claim evidence for complexity aversion, and mainly theoretical research (which is in the spirit of the OFT policy report by Garrod et al., 2008) suggests that complexity exploitation is an issue, we could not detect any evidence of either complexity aversion or exploitation in relation to prices. We did find that, in Experiment 2 where pricing strategies were computer generated and exhibited lower variance than in Experiment 1, there is some evidence of a complexity exploitation effect in quantities: that is, for a given price, more is bought of the complex product than of the simple one. The possibility of consumer exploitability is confirmed by considering consumer behavior when prices were high. Demand for complex products was more elastic than demand for simple products. Consumers may also be exploited not just because of their uncertainty about the value of the products, but more fundamentally because their preferences are uncertain in the first place: as a result, they may anchor their valuation on past experience. This suggests that firms who engage in consistent (lower variance) pricing behavior may be more effective in selling to consumers.

It could be argued that the interpretation of differences in quantities bought for a given price is not really a form of exploitation. We accept the plausibility of this alternative interpretation, although, to counter it, we note that uncertainty about the value of the product is a form of bounded rationality and as such in our view it is normatively appropriate to consider this as a form of exploitation relative to what the consumer would be doing were he or she not confused by the product. In this sense, while further research is clearly needed, there is some *qualified* support for the claim that consumers may be harmed by product complexity, even though prices are not systematically altered.

CHAPTER 3: Products Complexity and Shaping Effects: An Experimental Investigation

1. Introduction

This chapter describes an experiment that is a follow up analysis of the experiment described in the previous chapter. In the previous experiment we did not detect a complexity aversion effect effects as regards to quantities bought and prices. However we found that: a) buyers were exploitable when a consistent high pricing strategy was employed; b) the demand elasticity for more complex products was greater than less complex ones. We also found preliminary evidence that subjects' decisions were affected by past price (i.e. shaping effects) and this may explain results a). We then decided to explore further our results running another experiment that consisted of two parts: a) a binary choice task to study complexity attitudes; b) a posted offer market set up to test for shaping effects. Let us discuss briefly each task.

The binary choice task replicates part of Sonsino et al.'s experiment. The main difference being that we use 9 lotteries, including Sonsino et al.'s ones, with different degrees of complexity, while Sonsino et al. use only 3 lotteries with two degrees of complexity. The reason why we use 9 lotteries, that give rise to 18 binary choices in our treatment, is that we want to check whether complexity aversion is robust to changes in lotteries and complexity levels that may explain why we did not detect complexity aversion in our previous experiment.

The binary choice task combined with the individual posted offer market task allows us to check whether complexity aversion depends on the kind of lottery and/or on the different nature of the task.

In the individual posted offer market task we also explore more systematically the shaping effects that appeared in the regression analysis of our previous experiment. Previous research found that past preferences are shaped by past prices. Shaping effects may have a considerable impact on firms' pricing decisions. If preferences are shaped by past prices in fact it may be more profitable for firms to use pricing strategies that take into account consumers' bounded rationality. If this is the case, firms may first price high, shaping consumers' preferences, and then low exploiting consumers. Our experimental test involves an individual choice task framed as a posted offer market. Subjects observe prices and then decide how much, if anything, they want to buy. There is no interaction whatsoever among subjects and no information about amounts bought and prices observed by any of the subject is made public to the others. This chapter is organized as follows: it is divided into three parts. The first part will deal with the binary choice task. The second part will present the test of shaping effects with a posted offer market set up. The third part will conclude.

PART 1: THE BINARY CHOICE TASK

1. Introduction

Previous research has shown that subjects are complexity averse. In the individual choice experiment Sonsino et al. find that subjects, violating expected utility theory, prefer the simple lottery over the more complex one. Other research shows that indeed products complexity matters (i.e. Rouse, 2008, Bostrom, 2005, Garrod et al., 2008). If complexity aversion is present in real market then this should be an issue to be dealt with for consumer policy.

Sitzia and Zizzo (2009) run a first market experiment using lotteries as products (chapter 2). They use Sonsino et al.'s method to generate more complex lotteries from simple ones. They do not find clear evidence of complexity aversion, although they find that subjects are exploitable when the products are complex. That is, buyers tend to buy greater quantities at higher prices.

Experimental evidence shows contrasting results regarding complexity aversion. Individual choice experiments detect complexity aversion (i.e. Sonsino et al.) while complexity aversion is not observed in experimental markets (i.e. Sitzia and Zizzo 2009). It is therefore important to understand whether complexity aversion is affected by the kind of choice subjects are presented with. In Sonsino et al. subjects are asked to choose between lotteries, in Sitzia and Zizzo (experiment 2) they only have to state whether and how much they are willing to buy at the stated price.

This experiment aims to explore the robustness of complexity aversion to change in the lottery used and to change in the degree of complexity. The lack of evidence supporting complexity aversion in Sitzia and Zizzo (2009), can be ascribed to the kinds of lotteries and complexity levels used and/or to the different individual choice tasks subjects are presented with (i.e. as opposed as Sonsino et al. in Sitzia and Zizzo in the IC treatments subjects are asked whether they are willing to buy at the stated price while in Sonsino et al. subjects choose between lotteries). The same reasoning applies to Sonsino et al. That is, Sonsino et al. only used two lotteries, one simple and one complex. In our previous experiment we use 2 simple lotteries of the same type used by Sonsino et al. and 2 complex lotteries that, however, are much more complex than the ones used by Sonsino et al. Our complex lotteries/products have in fact 27 outcomes while Sonsino et al. only 6. In the binary choice task we then replicate part Sonsino et al.'s experiment using both the lotteries that we use in Sitzia and Zizzo (2009) with 3 different level of complexity (that is lotteries with 3 outcomes, with 6 outcomes and with 27 outcomes) and Sonsino et al.'s lotteries with three different level of complexity as well. This choice has the double advantage of comparability of our results with Sonsino et al.'s and at the same time we are able to check whether complexity aversion, if detected, is robust to changes in lotteries and complexity levels. Similarly if the results in the posted offer market set up (i.e. the second part of the experiment) match with the results in the first part of the experiment, as regards as complexity effects, then we could conclude that complexity aversion does not depend on the different individual task subjects face.

2. Sonsino et al.'s experiment

In this section I will present the first part of the experiment. Since our design is based on part of Sonsino et al.'s experiment, I will first discuss theirs and then turn to ours. Sonsino et al.'s paper presents a thorough analysis on complexity aversion. Their experiment involves several tasks and types of lotteries. Some lotteries are called multi-period lotteries in that the payments involved are deferred. This implies a more complicated decision process for the subjects than in the case of one-period lotteries. Firstly subjects need to discount later payments and then judge the level of complexity as measured by the number of outcomes (although discounting can be consider as an increase in complexity as well). So, multi-period lotteries have two layers of complexity: one is the time; the other one is the number of outcomes. They also use one-period lotteries but with different expected values and complexity levels. Finally they use lotteries with the same expected value but a different number of outcomes, like the ones that we use in our previous experiment. Overall they find that subjects are complexity averse and that complexity increases the noise of the decision process.

The task we are interested in is the simplest that Sonsino et al. run in their paper, that is we use lotteries with the same expected value but a different number of outcomes. We do that mainly because of comparability of the results with our previous experiment since the lotteries in that task were of this kind. When we thought about which lotteries to use in that experiment we decided to use the simplest possible lotteries that Sonsino et al. used. The main reason is that a market experiment is per se more complicated than a binary choice experiment, therefore we wanted to keep the design as simple as possible and isolate complexity of the lotteries as measured by the number of outcomes in the possible simplest environment, avoiding any possible confounding effect, such as discounting by subjects that could have arisen with multi-period lottery.

Let us now focus on the Sonsino et al.'s task that we replicate in this part of our experiment. Sonsino et al.'s task (pp. 950, 951, 952) involves 2 binary choice tasks[1]. The first task choice was to choose between a simple lottery S3 with three outcomes and the certainty equivalent CE, the second task was to choose between the simple lottery S3 and a more complex lottery with 6 outcomes C3, as shown in the table below. The second choice was to choose between S3 and a C3.

The lotteries have the same expected value of 107 experimental points and differ in complexity. The simplest one is CE which is the certainty equivalent. S3 is a relatively simple lottery compared to C3. Complexity is measured in terms of number of outcomes. C3 is derived from S3 using the following formula (see Sonsino et al. p. 950 for a more detailed explanation).

$$C3 = \frac{1}{2}S3 + \frac{1}{2}S3$$

[1] Sonsino et al. used actually 3 binary choices. The third choice however involved multi-period lotteries that we do not use in our experiment for the reasons explained above.

That is, C3 is obtained from two random draws of the simple lottery S3. The method used is the same we use in the previous experiment, although there we had to generate a complex lottery with 27 seven outcomes so the formula changes slightly, in particular, the complex lottery is generated from three random draws of the simple one. Moreover we assign to each draw a different weight depending on the complex lottery we want to generate.

Choice 1

S3

Outcomes	Probabilities	Results
1	30.00%	150
2	40.00%	80
3	30.00%	100

CE

Outcomes	Probabilities	Results
1	100%	107

Choice 2

S3

Outcomes	Probabilities	Results
1	30.00%	150
2	40.00%	80
3	30.00%	100

C3

Outcomes	Probabilities	Results
1	24.00%	115
2	9.00%	150
3	16.00%	80
4	9.00%	100
5	24.00%	90
6	18.00%	125

Table 1: Sonsino et al.' binary choices

According to Sonsino et al., if a subject is both an expected utility maximizer and risk averse she should prefer CE to S3 in choice 1 and C3 to S3 choice 2. They found that 69% of subjects preferred CE in choice 1 and therefore claim that at least 69% of the subjects should prefer C3 in choice 2. However, considering only the subjects that chose CE (risk averse) in choice 1, only 53% of them chose C3 in choice 2 so chose consistently to EUT. The rest of the subjects, that is 47%, switched from the safer (CE) to the riskier (S3) alternative. Sonsino et al. maintain that these results may be ascribed to complexity aversion. Considering only the subjects that chose S3 (acting as risk loving), 52.6% of them chose the safer option C1 in choice 2, violating EUT. That is, 52.6% of the subjects of this sub-sample switched from the riskier to the safer alternative. Sonsino et al. suggest that this might be due to noise in the decision process. The more complex the lotteries the more noisy the decision process is. So, they conclude, some of the switched explained by aversion to complexity can be explained by noise.

3. Binary choice task

3.1 Lotteries/products used

In this experiment (both the binary choice task and the market task) we employ 9 lotteries (table 2), including Sonsino et al. ones.

In our previous experiment we used 4 lotteries. Two simple ones, with 3 outcomes and two complex ones, derived from the simple ones, with 27 outcomes and all with an expected value of 60. In this experiment, all lotteries have an expected value of 107 experimental points (this is done in order to be consistent with Sonsino et al.'s experiment) with a conversion rate of 975[2] to a pound.

Sonsino et al.'s used 3 lotteries, the ones showed in table 1. One simple with 3 outcomes and one complex, derived from the simple one, with 6 outcomes and the certainty equivalent CE. We decided to use 9 lotteries, Sonsino et al.'s ones and ours with three level of complexity: simple lotteries with 3 outcomes, Sonsino et al.'s complexity with 6 outcomes, and Sitzia and Zizzo's complexity with 27 outcomes. We did this both for robustness and comparability reasons.

The lotteries employed are displayed in table 2. CE is the certainty equivalent. S1, S2, S3 are the simple lotteries with 3 outcomes each. C1, C2, C3 are the complex lotteries with 6 outcomes. VC1, VC2, VC3 with 27 outcomes are the very complex lotteries (see the previous chapter for a more detailed explanation of the procedure we used to generate these lotteries from the simple ones).

The lotteries are divided into 3 groups. Each group has a simple, a complex and very complex lottery. The complex and the very complex are derived from the simple one. Group 1 and 2 are the lotteries that we used in our previous experiment, except for C1 and C2 that are new and have been generated to match Sonsino et al.'s complex lottery. Group 3 includes the lotteries used by Sonsino et al. except for VC3 that has been generated from S3 to match our very complex lotteries VC1 and VC2. Including Sonsino et al.'s lottery has the obvious purpose of facilitating

[2] Our lotteries in our previous experiment had an expected value of 60, we had to increase it from 60 to 107 to match Sonsino et al.'s lotteries expected value of 107. The conversion rate and the incentives per lottery did not change with respect to our previous one (this is particularly relevant for the market task). This allows for comparability of the results both with Sonsino et al. experiment and with our previous experiment.

replication. To summarise, there are 3 groups of lotteries, for each group: simple, a complex and a very complex lottery, i.e. three level of complexity.

CE

Outcomes	Probabilities	Results
1	100%	107

S1

Outcomes	Probabilities	Results
1	20.00%	112
2	50.00%	57
3	30.00%	187

S2

Outcome	Probabilities	Results
1	50.00%	198
2	20.00%	50
3	30.00%	113

S3

Outcomes	Probabilities	Results
1	30.00%	150
2	40.00%	80
3	30.00%	100

C1

Outcomes	Probabilities	Results
1	20.00%	84.5
2	9.00%	187
3	12.00%	149.5
4	4.00%	112
5	30.00%	122
6	25.00%	57

C2

Outcomes	Probabilities	Results
1	12.00%	155.5
2	30.00%	124
3	4.00%	113
4	9.00%	198
5	25.00%	50
6	20.00%	81.5

C3

Outcomes	Probabilities	Results
1	24.00%	115
2	9.00%	150
3	16.00%	80
4	9.00%	100
5	24.00%	90
6	18.00%	125

VC1

Outcomes	Probabilities	Results
1	7.50%	1
2	3.00%	110.4
3	12.50%	57
4	2.70%	187
5	7.50%	66.1
6	2.00%	108.15
7	1.80%	184.75
8	3.00%	177.85
9	2.00%	62.5
10	4.50%	183.1
11	1.80%	119.5
12	3.00%	175.65
13	5.00%	58.65
14	5.00%	106.5
15	7.50%	174
16	4.50%	177.9
17	1.80%	181.75
18	1.20%	117.25
19	1.20%	114.25
20	1.20%	179.5
21	0.80%	112
22	4.50%	70
23	2.00%	110.35
24	3.00%	64.75
25	3.00%	115.6
26	5.00%	60.85
27	3.00%	67.75

VC2

Outcomes	Probabilities	Results
1	3.00%	58.85
2	2.70%	198
3	4.50%	187.64
4	7.50%	54.44
5	2.00%	111.11
6	1.20%	118.95
7	3.00%	111.14
8	3.00%	62.25
9	3.00%	185.09
10	12.50%	50
11	1.20%	115.55
12	1.80%	195.45
13	5.00%	54.41
14	1.80%	121.5
15	4.50%	64.8
16	7.50%	60.36
17	7.50%	183.2
18	0.80%	113
19	2.00%	108.59
20	4.50%	193.56
21	2.00%	56.3
22	1.20%	189.5
23	3.00%	117.06
24	1.80%	192.05
25	5.00%	51.89
26	5.00%	106.7
27	3.00%	187.61

VC3

Outcomes	Probabilities	Results
1	2.70%	104
2	2.70%	146.5
3	2.70%	146
4	2.70%	103.5
5	4.80%	139.5
6	2.70%	107.5
7	3.60%	144.4
8	4.80%	84.9
9	2.70%	100
10	4.80%	85.6
11	3.60%	98.4
12	3.60%	101.9
13	6.40%	80
14	3.60%	87
15	3.60%	98.6
16	4.80%	81.6
17	3.60%	141.1
18	3.60%	90.5
19	3.60%	140.9
20	3.60%	145.1
21	3.60%	86.5
22	2.70%	150
23	3.60%	83
24	4.80%	81.4
25	3.60%	102.6
26	2.70%	142.5
27	4.80%	97

Table 2: Lotteries Used in the Binary Choice Task

Including the very complex lotteries allows us firstly, as already pointed out, comparability of the results with our previous experiment. Additionally, we can check whether subjects' behaviour changes with increased complexity. Similarly, using three different groups of lotteries allows us to check whether the results depend on the type of lottery used.

The table below shows a clearer summary of the lotteries employed displayed by group and complexity.

Level of Complexity	Group 1	Group 2	Group 3
Simple	S1	S2	S3
Complex	C1	C2	C3
Very Complex	VC1	VC2	VC3

Table 3: Summary of the Lotteries Used

It is worth noting here that the way we obtain more complex lotteries from the simple ones may be thought of as a framing manipulation . We have already discussed this issue in the previous chapter. However while in that experiment subjects are presented both with the simple lottery and the complex one, in this experiment they only play with one lottery throughout the experiment. So there is no scope for framing effects to show.

3.2 Experimental design.

We ran the experiment at the University of East Anglia during the spring term 2009. It consisted of 2 parts. The first part was a binary choice task with 18 binary choices where subjects had to choose between lotteries of different complexity but with the same expected value of 107 with a conversion rate of 975 points to a pound. The second part was a posted offer market set up with the same lotteries used in the first part that will be discussed in the second part of this chapter.

A sample of 384 subjects took part in the experiment. After their arrival they were asked to sign a consent form, to read the instruction for this part of the experiment and to complete a questionnaire to check whether they had understood how the first part worked, they were given an opportunity to ask questions for clarification. After the experimenters answered all the questions the first part started.

Subjects were asked to choose between two lotteries over 18 rounds. The pair of lotteries was different every time and the sequence of choices was randomized across subjects. The choices were the following:

Group 1				Group 2				Group 3			
Choice	Lottery 1		Lottery 1	Choice	Lottery 1		Lottery 2	Choice	Lottery 1		Lottery 2
1	CE	vs	S1	**7**	CE	vs	S2	**13**	CE	vs	S3
2	CE	vs	C1	**8**	CE	vs	C2	**14**	CE	vs	C3
3	CE	vs	VC1	**9**	CE	vs	VC2	**15**	CE	vs	VC3
4	S1	vs	C1	**10**	S2	vs	C2	**16**	S3	vs	C3
5	S1	vs	VC1	**11**	S2	vs	VC2	**17**	S3	vs	VC3
6	C1	vs	VC1	**12**	C2	vs	VC2	**18**	C3	vs	VC3

Table 4: Choices in the Binary Task

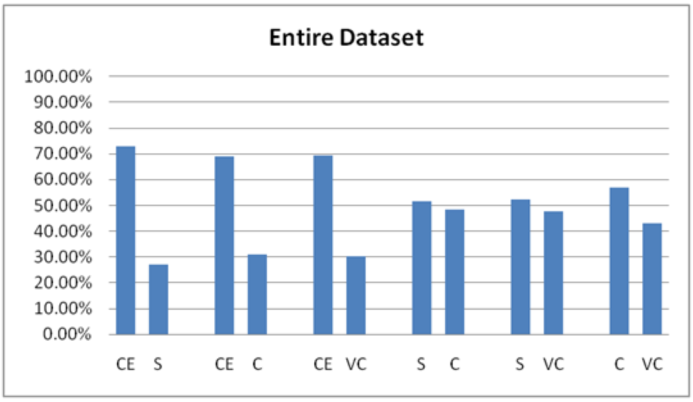
As can be noticed from table 4, we did not mix up pair of lotteries belonging to different group. So for example there were not choices such as S1 vs S3. The lotteries in each choice belong to the same group. This was to enable us to analyse the choices within a group and to compare them with the other two, in particular group 3 (Sonsino et al.'s lotteries). This would allow us to check whether there are systematic differences between groups and therefore if the results are lottery-dependent. For each group we have 6 different choices that enable us for a more complete analysis on complexity as opposed to Sonsino et al.'s where they only had 2 choices. Finally, we can also pool the choices across groups and compare lotteries of different complexity over the entire dataset. This would give us an idea of the subjects' preferences over lotteries of different complexity.

Earnings for the binary choice task. Each lottery chosen for every choice was played at the end of the experiment and all the points accumulated were added up.

4. Results

4.1 Overview

Figure 1 shows the summary of the results for the entire dataset and for type of choice, without considering each group separately.

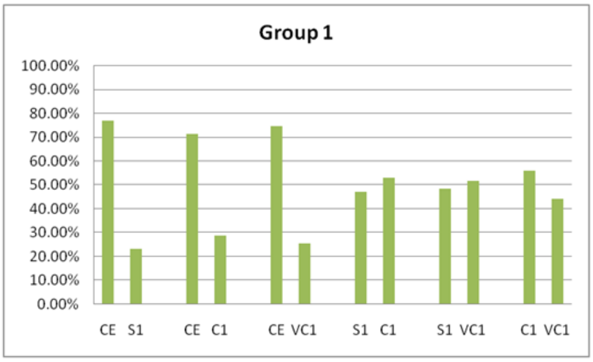


Lottery	Choices
CE	73.00%
S	27.00%
CE	69.10%
C	30.90%
CE	69.40%
VC	30.40%
S	51.60%
C	48.40%
S	52.40%
VC	47.60%
C	56.90%
VC	43.10%

Figure 1: Summary of Binary Choices for the Entire Dataset

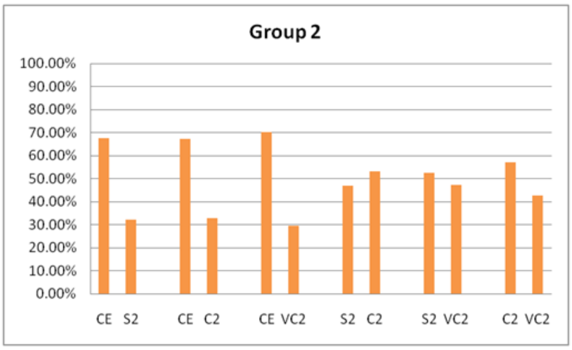
Since the observations are not independent as there are three choices that are the same for each subject, that differ only in the group, we had to create a variable to overcome this problem. We considered the same choice for each group and calculated an average per subject that has then be compared with a dummy variable whose value is always 0.5. So for example, consider the choice CE vs S. Each subject made this choice three times, one for each group, so CE vs S1, CE vs S2 and CE vs S3. We considered the choices made and calculated the average. As can be seen the certainty equivalent is always chosen more often than S, C and VC. The difference is statistically significant in a Wilcoxon test ($p<0.0001$, $p<0.0002$, $p<0.0002$ respectively for choices CE-S, CE-C and CE-VC). The simple lotteries S are not statistically preferred to the complex one C (Wilcoxon ranked signed test, $p<0.284$) and the same holds true for S versus VC (Wilcoxon ranked signed test, $p<0.386$), while C is statistically preferred to VC (Wilcoxon ranked signed test $p=0.001$). The table shows the pattern of choices for the entire dataset.

As we can see (Fig 2, Fig 3, Fig 4) almost the same pattern appears when we consider each single group of lotteries (group 1, group 2 and group 3). We will see in a later section whether the type of lottery is relevant to the results.



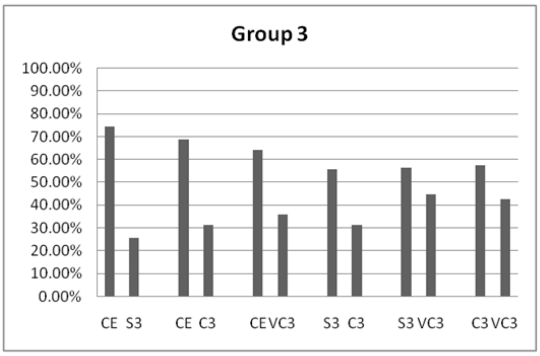
Lottery	Choices
CE	76.85%
S1	23.20%
CE	71.40%
C1	28.60%
CE	74.50%
VC1	25.50%
S1	46.90%
C1	53.10%
S1	48.40%
VC1	51.60%
C1	56%
VC1	44%

Figure 2: Summary of Binary Choice for Group 1



Lottery	Choice
CE	67.70%
S2	32.30%
CE	67.20%
C2	32.80%
CE	70.30%
VC2	29.70%
S2	46.90%
C2	53.10%
S2	52.60%
VC2	47.40%
C2	57.30%
VC2	42.70%

Figure 3: Summary of Binary Choices for Group 2



Lottery	Choice
CE	74.50%
S3	25.50%
CE	68.80%
C3	31.30%
CE	64.10%
VC3	35.90%
S3	55.70%
C3	31.30%
S3	56.30%
VC3	44.80%
C3	57.30%
VC3	42.70%

Figure 4: Summary of Binary Choices for Group 3

4.2 Differences between lotteries

As explained in the introduction we added several lotteries that were not part of Sonsino et al.'s experiment. After all, the results that Sonsino et al. got may be dependent on the lotteries used and obviously also our previous market experiment results may be as well. To investigate this issue we created a variable called Difference. Consider table 5 that explains how the variable has been created:

Group 3		Group 1		Difference
S vs C (group 3)	**Value**	**S vs C (Group 1)**	**Value**	
S	1	S	1	**0**
S	1	C	0	**1**
C	0	S	1	**-1**
C	0	C	0	**0**

Table 5:The Variable Difference

Group 3 is the group with Sonsino et al.'s lotteries, while group 1 is the group with our lotteries. Consider now for both groups the choice S vs C. For each group there are 2 possible options. If the subject chooses S, we assign a value of 1 to that choice. If the subject chooses C, we assign a value 0. Then we consider choices in group 1 and 3 simultaneously. In this case there are 4 possibilities: SS, SC, CS and CC. We take the difference of the values and get the variable Difference. So, if the choice is SS we simply subtract 1 to 1, and so on for the other cases. We do that for every binary choice under the null hypothesis that the lottery type does not matter (e.g. whether it comes from group 1, 2 or 3). The average of the variable Difference considering all 4 cases is 0. However, since for each group in table 5 only one choice is possible, the variable difference can assume different values, ranging from -1 to 1. In this specific example, the question we ask is: Do subjects tend to choose more frequently lotteries in one group than in the others? To check this we created a variable (Dummy0) with all entries equal to 0. If subjects tend to choose the same lotteries with the same frequency in both groups, then the variable Difference and the Dummyu0 should not be statistically significantly different and therefore we can conclude that there are not differences between the lotteries in the groups and then the choice and results do not depend on the lotteries used. We performed a sign test using these two variables.

We overall found that there is not statistical difference between choices of different groups except for few cases. A sing test shows that the choices S vs C in group 1 and the same choices in group 3 are statistically significant (($p<0.004$), this means that subjects tend to choose more frequently the simple lottery in group 3 than the same lottery in group 1. We also found that the variable Difference for the choices S vs VC in group 1 and group 3 is different statistically from the Dummy0, this means that in that choice subjects tend to choose more frequently the simple lottery in group 3 than the same lottery in group 1 ($p<0.011$). Finally, for the binary choice C vs VC for group 2 and 3, subjects tend to choose more frequently C in group 3 than the same lottery in group 1 ($p<0.0001$). It seems then that, except for these few cases, there is not a clear pattern in our results, so we can conclude that we do not have enough evidence to claim that the type of lottery (e.g. group) affects the choices made by the subjects.

Result 1. We do not detect any clear pattern in choices that allows us to claim that the type of lotteries matters for subjects' choices. Our results suggest that complexity aversion is robust to changes in lotteries.

4.3 Complexity aversion or noisy decision process?

As said before, Sonsino et al. compared the choices CE-S vs S-C choice. We will do the same for each group to see whether we get similar results. Sonsino et al. found that 69% of the subjects chose the certainty equivalent CE in the first choice. 53% of these subjects were risk averse and consistently to the expected utility theory they chose C in the second choice, which was the safer lottery. The rest of the subjects chose the simpler and riskier lottery S in the second choice. Sonsino et al.'s maintained that this can be explained by complexity aversion. The subjects that did not chose CE in the first choice (31%), that is the riskier lottery S, switched to the more complex and safer lottery C (52.6%). Sonsino et al. maintain that this can be explained by noise in the decision process: when more complex lotteries are involved noise in the decision process increase. It has to be noticed that also the choices explained by complexity aversion involve an increase in complexity from the first choice to the second one. We will discuss this in more detail later.

We will replicate Sonsino et al.'s analysis but with a dataset that allows us for more comparisons. In fact we can also analyse another case, that is, CE-S vs S-VC choice.

For clarity purposes, we will use the following notation: Case 1 = choice 1 - CE vs S; and choice 2 - S vs C; Case 2 – choice 1 - CE vs S and choice 2 VC vs S.

Let us first consider each group separately.

Group 1 - CASE 1: CE vs S and S vs C

We are interested in the switches from safe to risky and from risky to safe (that is from CE in choice 1 to S in choice 2 and from S in choice 1 to C in choice 2). In the first case, we can argue that choices can be explained by complexity aversion in the second case by noise (as Sonsino does).

Choice 1 - Subjects that chose CE in the first choice. There are 50.8% of the subjects that chose CE in the first binary choice and switched to risky (S) in the second choice. This, according to Sonsino et al. can be explained by complexity aversion. The rest of the subjects, consistently to expected utility theory, chose CE in the first pair and C in the second one.

Choice 2 - Subjects that chose S in the first choice. There are 60.6% of subjects that chose the risky lottery (S) in the first choice and switched to the safe one in the second choice (C), this can be explained by noise. The rest of subjects chose the risk lottery in the first choice and the risky lottery in the second one, behaving according to expected utility theory.

I used a χ^2 test, to check whether switches are more frequent in the first case than in the second one. The difference is statistically significant ($p<0.066$).

Group 1 - CASE 2 – CE vs S and S vsVC

Choice 1 - Subjects that chose CE in the first choice. In this case 49.5% of subjects switched from the safer lottery (CE) to the riskier and simpler (S). This can be explained by complexity aversion. The rest of the subjects chose CE in the first choice and, consistently to EUT, VC in the second.

Choice 2 - Subjects that chose S in the first choice. 55.1% of the subjects chose the riskier lottery (S) in the first choice but the safer one (VC) in the second choice. This can be explained by noise in the decision process. The others behave consistently to EUT

I performed as before a χ^2 test to check whether the difference was significant and it is not (*p<0.264*)

Group 2 - CASE 1 – CE vs S and S vs C

1) Subjects that chose CE in the first choice. 51.2% of these subjects chose the risky option in the second task (S) this can be explained by complexity aversion. The rest stuck to the safer (C).

2) Subjects that chose S in the first choice. 45.20% of subjects that chose the risky option in the first choice switched to the safer one C, this may be due to noise. The other choices werer consistent to EUT.

A χ^2 test shows that there is not statistically significant difference between both switches safe-risky and risky-safe (*p<0.286*).

Group 2 - CASE 2 – CE vs S and S-VC

*1) Subjects that chose CE in the first choice***.** In this case 54.6% chose CE switched to riskier lottery VC in the second choice. The rest chose VC, consistently to EUT.

2) Subjects that chose S in the first choice. 51.6% of these subjects chose the riskier lottery (S) in the first choice but the safer one (VC) in the second choice. The other chose consistently to EUT.

A χ^2 does not show statistical significant difference between the two kinds of switches, (*p<0.151*).

Group 3 - CASE 1 – CE vs S and S vs C

1) Subjects that chose CE in the first choice. 55.6% of these subjects chose the risky option in the second choice (S). The rest chose the safer one (C).

2) Subjects that chose S in the first choice. 56.6% of these subjects switched to the safer one (C). The rest behaved consistently to EUT.

The difference is not significant in a χ^2 test (*p<0.512*).

Group 3 - CASE 2 – CE vs S and S vs VC

*1) Subjects that chose CE in the first choice***.** In this case 43.9% switched to the simpler and riskier lottery in the second choice. The rest chose VC, consistently to EUT.

2) Subjects that chose S in the first choice. 44.8% of these subjects switched to the safer lottery (VC) in the second choice. The others chose consistently to EUT.

A χ^2 test shows not statistical significant difference between the two kinds of switches, ($p<0.440$)

Case			Group 1				P-value
			Choice	%	Choice	%	
1	Choice 1 Choice 2	CE vs S C vs S	CE (safe) S (risky)	50.80%	S (risky) C (safe)	60.60%	0.066
2	Choice 1 Choice 2	CE vs S VC vs S	CE (safe) S (risky)	49.50%	S (risky) VC (safe)	55.10%	0.264
Case			**Group 2**				**P-value**
			Choice	%	Choice	%	
1	Choice 1 Choice 2	CE vs S C vs S	CE (safe) S (risky)	51.20%	S (risky) C (safe)	45.20%	0.286
2	Choice 1 Choice 2	CE vs S VC vs S	CE (safe) S (risky)	54.60%	S (risky) VC (safe)	51.60%	0.151
Case			**Group 3**				**P-value**
			Choice	%	Choice	%	
1	Choice 1 Choice 2	CE vs S C vs S	CE (safe) S (risky)	55.60%	S (risky) C (safe)	43.90%	0.512
2	Choice 1 Choice 2	CE vs S VC vs S	CE (safe) S (risky)	56.60%	S (risky) VC (safe)	44.80%	0.440

Table 6: Results from χ^2 Tests on Switches from Risky to Safe (Noise) and Safe to Risky (Complexity Aversion)

The table above summarises the results of the last three pages. Except for one case (group 1, case 1) where the statistical significance is only at 10% level, the differences in switches between safe to risky, that can be explained by complexity aversion, and the switches from risky to safe, that can be explained be noise, are not statistically significant. Therefore we cannot claim that these switches that may be explained by complexity aversion are actually ascribable to that, because it could be due to noise as well, but not vice-versa. Noise is symmetrical in both switches, complexity aversion is not. Both pair-wise choices in fact involve an increase in complexity, and there is not statistical difference between the ones explained by complexity aversion and the ones explained by noise. That is, there seems to be no additional effect, according to the χ^2 tests results, due to complexity aversion in the switches that can be explained by it. We therefore conclude that the switches that are explained by complexity aversion can be explained by a noisy decision process that increases with complexity. This interpretation has the advantage of explaining both switches by noise, that is, is parsimonious, contrary rather than by complexity aversion and noise. This leads us to our second result:

Result 2*:* Complexity aversion can explain switches from the safer lottery (CE) to the riskier but more complex one (C or VC). These switches can also be explained by noise. However, switches from the riskier but simpler lottery (S) to the safer but more complex one (C or VC) can only be explained by noise in the decision process. We claim however that both switches may be explained by a noisy decision process.

Since noise is relevant for the interpretation of the data, the next section will explore in more detail the noise in the decision process and whether this increases as complexity increases.

4.4 Analysis of the noise in the decision process

Sonsino et al. suggest that as the lotteries become more complex, the more the decision process is noisy[3]. We then decided to perform a test for noise based on transitivity and intransitivity of choices made by the same subject[4]. We considered

[3] There is a vast literature that proposes stochastic econometric models that take into account noise in the decision process. Among others, Loomes (2005) suggests that noise, or stochastic component of behaviour, as he calls it, is relevant especially for binary choices between lotteries. He also maintains that not taking into account this stochastic component may lead to a different interpretation of the data.

[4] Sopher and Gigliotti (1993) run an experiment aiming to test the rational choice hypothesis and the random error hypothesis. They found evidence that intransitive choices are explained by random error.

two groups of three lotteries each and checked whether the choices were transitive or intransitive. The lotteries were CE, S and C in the first case and S, C and VC in the second case. The first group of lotteries is simpler than the second one. One example of intransitive choice is the following:

$$S > C; S > CE;\ CE > S$$

If a subject prefer S to C and S to CE then she must prefer S to CE. For each of the three groups of lotteries we considered all the possible combinations such as the one showed before. We then created a variable Transitive-Simple and Transitive-Complex with a value for each subject. If the triple of choices were transitive then the value of the variable was 1, otherwise zero. Since each subject made the same triple of choices (like the one showed above) three times and observations need to be independent, we divided the value we obtained for each subject by three getting an average for each subject, which represents the fraction of transitive choices per subject. If a subject's choices are always transitive then the variable is 1, if not the value is 0, any value between 0 and 1 represents choices that in some cases are transitive in some others are not. We then performed a Wilcoxon Signed Ranks Test on Transitive-Simple and Transitive-Complex. The results show that subjects tend to make more intransitive choices when more complex lotteries are involved (*p=0.010*). This suggests that the decision process was actually noisier with more complex lotteries. If we combine these results with the ones that we get with the χ^2 test results in the previous section, then we do not have enough evidence to claim that subjects' choices were affected by complexity aversion. In fact it may well be that the switches from safe to risky and risky to safe, as explained before, were due to noise rather than to complexity aversion.

Result 3: As the complexity of the lottery increases, the noise in subjects' choices, as measured by intransitive choices, increases as well. Therefore, complexity aversion can be explained as well by a noisy decision process.

5. Brief note on the incentive system

In this experiment and in the previous one we use lotteries that have the same expected value. The reason why we decided to do this is, given our design, to maximise experimental control on the variables we were interested in studying, i.e. complexity. If more than one dimension varies (complexity and expected value) then it would not be possible to understand which one is responsible if some pattern in the data is found. Nonetheless, it can be argued that, particularly in the binary choice task, since subjects make 18 choices between lotteries that have the same expected value (107 experimental points) then it does not matter what they really chose since on average they get the same amount either way. In this sense it can be said that we lose control over the incentives used. However, while it is true that in principle subjects may choose randomly between lotteries, and this would explain why our data is consistent with the noise hypothesis, it has to be noticed that we also detect some patterns in the data that are not consistent with random choices made by subjects. This is clearly seen in figure 2, 3 and 4. Subjects always choose more frequently (about 70% of the times) the certainty equivalent. The same critique can be directed to the experiment discussed in chapter 2 and the second part of the experiment discussed later in in this chapter. However, also in the previous experiment we detect some patterns in the data, such as a greater demand price elasticity for complex products or decreasing price over time when human sellers are employed. Similarly, in the second part of this experiment we detect clear evidence of shaping effects. Therefore, insofar as we are able to detect some pattern in the data we can be confident enough that we are not losing control over the incentives.

We have nevertheless to admit that the noise we found in the data may be due a loss of control over the incentives. Our experiment replicates part of Sonsiono’s experiment where they also use lotteries with the same expected value. Their entire experiment however provides evidence that noise increases with complexity, also when choices do not have the same expected value. This suggests that our results, that are consistent with theirs, may not be driven by a loss of control over the incentives. Having said that, extending this research using the same structure of this experiment, but with lotteries with a different expected value, is a possible avenue for future research.

6. Summary

Our results replicate the same pattern Sonsino et al. found in their experiment. We had 9 lotteries instead of three and this makes our results robust to changes in lotteries.

Subjects tend to prefer the certainty equivalent over other lotteries, however when they have to choose between other lotteries they switch either to the riskier (or simpler) or to the safer (complex) lotteries. The first case can be explained by complexity aversion while the second by noise. We then investigated whether noise increased for more complex lotteries and we found that actually it does. This gives support to our claim that actually the switches from safe to risky, that can be explained by complexity aversion, can in fact be explained by the increased noise in the decision process due to an increased complexity of the lotteries involved. Similarly, switches from risky to safe can be explained by noise.

PART 2: THE INVIDUAL CHOICE TASK AND SHAPING EFFECTS[5]

This section is devoted to the second part of the experiment, which is a test for shaping effects. This part is organised as follows, there will be an introduction that review the literature on shaping effects, then the second section will describe the design, the following will be devoted to the results and finally there will be a section concluding the chapter.

1. Introduction

Dan Ariely (2008) has noted how the Apple IPhone was originally priced at $ 600, and, when it moved down to $ 200, everyone thought it was a bargain to buy as a result. This paper devises an experiment trying to test this intuition: namely, that it may be profitable for companies to choose to price a new product high and then reduce the price, rather than provide a low introductory price and raise the price later. There are reasons why the latter has often been considered a good strategy by economists: in the presence of switching costs by at least a fraction of consumers, a low introductory price may be used to 'lock in' consumers and the

[5] This section is almost completely based on Sitzia and Zizzo (2010).

price may be raised afterwards (e.g., Cabral, 2010; Klemperer, 1995); it may be used to signal low cost (Bagwell, 1987); it may be used to facilitate buyer experimentation when there is uncertainty about the product's quality (Schlee, 2001). We do not wish to deny that there are settings where a 'low price – high price' ('low high') strategy may be profitable. This paper, however, presents an experiment trying to test the opposite intuition: namely, that it may be profitable for firms first to price high and then low. Reasons why this might be the case, which have been identified in the industrial economics literature, include intertemporal price discrimination with durable goods (e.g., Conlisk et al., 1984), special parametrizations of switching costs models (Klemperer, 1995), and – more cognate this paper – models of an introductory high price as a signal of quality under the assumption of heterogeneous (in part informed, in part uninformed) agents (Bagwell and Riordan, 1991). Our experiment has no durable goods, no (at least monetary) switching costs, and does not have the game theory structure that would justify Bagwell and Riordan equilibria.

We aim to show that a 'high price – low price' ('high low') strategy may nevertheless be profitable for companies because of *shaping effects*: if consumers have unclear preferences, their willingness to buy may be affected by anchors provided either artificially or through the operation of auction mechanisms (Ariely et al., 2003, 2006; Loomes et al., 2003; Brooks et al., 2008). Having observed high prices implies a high reference price and a belief that a good deal is made when the price is decreased; conversely, having observed low prices implies a low reference price and a belief that a bad deal is made when the price is increased.[6] This psychological mechanism is consistent with adaptation theory in marketing (Morris and Morris, 1990) and bears a close analogy with the success of the so called 'black hat/white hat' strategy in negotiation experiments: one can get better bargaining outcomes by starting tough and then softening up in the negotiation process than starting soft and then hardening up (Hilty and Carnevale, 1993). In spirit, it fits with the Bagwell and Riordan model insofar as a high price can indeed be here

[6] Isoni et al. (2008) contains a recent formalization of the idea of 'bad deal aversion,' which he employs to explain the willingness to pay – willingness to accept disparity observed in contingent valuation studies. Note that, although we talk here of 'high low' and 'low high', one could more precisely identify these strategies as price high and then lower vs. price low and then higher. We use the 'high low' and 'low high' terminology here simply as less cumbersome.

interpreted as a signal of high quality, even though the conditions of the Bagwell and Riordan model do not hold. Our behavioral mechanism may provide a powerful reason why 'high low' price strategies are observed, not only in the context of Ariely's (2008) mobile phones example but also, for example, in relation to appliances (Consumer Reports, 2008), video consoles (e.g., Fitzgerald, 1992), or color television sets (Krishnan et al., 1999). It reflects the admonition by marketers that price discounts may undermine the perceived economic value of a good (Lucke and Hogan, 2007).[7]

This experiment is the first that tries to systematically test the profitability of a shaping effect related 'high low' strategy in an experimental retail market. It is most closely related to the experimental work described in the previous chapter where we found evidence from regression analysis that demand seemed to depend on past prices, and, namely, that demand at time t was higher if the price at time t – 1 was higher. The purpose of that experiment was not however to test shaping effects in general, nor to test the profitability of a strategy trying to exploit such shaping effects.

As in the previous experiment, we used lotteries of different degrees of complexity as products that consumers could buy. The choice of lotteries as products[8] was not only for comparability to the previous experiment, but also and more significantly to ensure the novelty of the product for all subjects, to ensure that it was a product that subjects could buy over a number of rounds without quickly and heterogeneously getting tired of it, and so ultimately to maximize experimental control. Experiments on reference dependent preferences have shown that similar behavioral features to those found with lotteries are found with real commodities (compare, e.g., Bateman et al., 1997, with Kahneman and Tversky,

[7] There is a technical literature in marketing science that looks at optimal pricing strategies based on decreasing prices (e.g., Bass, 1980; Krishnan et al., 1999). This research, however, takes the empirical sales curve in time as a given rather than attempting to explain it as a function of price-dependent consumer preferences (or the kind of informational issues highlighted in Bagwell and Riordan, 1991).

[8] In the previous part of the experiment we refer to lotteries rather than products, since the purpose was to check for complexity aversion, we thought that that was a more appropriate terminology. In this part of experiment we refer to lotteries as products, since this part is related to firms and consumers, we think that it is more logical to refer to them as products.

1979), and this, together with the danger of loss of control from other product choices (e.g., due to satiation), implies the usefulness of our choice. Another reason for having lottery products is the simple way we can control for level of complexity and therefore potential product value uncertainty using a lottery paradigm: we indexed complexity using the same procedure as already stressed in the first part of this chapter, which in turns relies on Sonsino et al. (2002); research still needs to be conducted on how to index complexity with real commodities. Of course, lottery tickets are a real commodity by themselves and one that is in high demand in the real world.[9]

Our key finding is that shaping effects do matter, and that a high low strategy would indeed be profitable for firms under different assumptions about cost and volume of demand. This is true no matter the type of product employed. The rest of this paper is structured as follows. Section 2 presents the experimental design, section 3 presents the results and section 4 concludes.

2. A close-up: Anchoring

In this section I am going to further the discussion on anchoring and its relationship with shaping, since the two phenomena are strictly related.

Tversky and Kahneman (1974) ran an experiment asking subjects, after they span a wheel of fortune, to estimate several quantities. They showed that these estimates were closely correlated to the random number selected by the wheel. They call this phenomenon anchoring and define it in the following way "People make estimates by starting from an initial value that is adjusted to yield the final answer" (p.1128). The initial value is the so-called 'anchor' and the extent to which individuals' evaluation is affected by the anchor depends on how precise the information they have on the quantity they are asked to estimate is. So, if an individual is asked how many countries there are in Africa after having spun the wheel of fortune, she/he will not be affected by the anchor if she/he knows that number. However the anchor needn't be a plausible figure, as shown by Tversky

[9] For example, in the 2008/09 financial year the U.K. National Lottery sold £ 5.15 billions in lottery tickets, up from £ 4.96 billions in 2007/08 (source: Camelot Stakeholder Report, 2009, p. 2).

and Kahneman experiment. Similarly, the anchor does not need to be explicitly given, but it could also be "the result of some incomplete computation" (p. 1128). What however characterises the process of anchoring is the fact that adjustments of the estimate starting from the anchor are usually insufficient, to the point that "different starting points yield different estimates, which are biased towards the initial values" (p. 1128).

Anchoring has been shown to be a robust phenomenon in laboratory settings. Leboeuf and Shafir (2009) provide experimental evidence that subjects judgments of time and distances are affected by anchors. More relevant to this chapter discussion is however whether anchors also affect preferences or willingness to pay and willingness to accept. Several studies have shown that this is actually the case. Ariely et al (2003) ran several experiments where subjects were asked how much they were willing to accept to be exposed to various unpleasant experiences, such as listening to some horrible noise. They found that the valuation was relatively low when the anchor provided was low and higher when the anchor provided was high. They also found that irrelevant cues, such as the last three digits of their social security number affected subjects' willingness to accept. Similar results are found by Green et al (1998) in referendum contingent evaluation whereby willingness to pay for public goods can be elicited.

The fact that anchors affect individuals judgment about the value of items (or products in general) has indeed led many economists to develop models that assume that buyers' decisions are not only affected by the actual price but also by the perceived price. The psychological mechanism assumed is usually based on the adaptation-level theory developed by Helson (1964). Emery (1974) observes that one of the implications of this theory when applied to a pricing context, is that the standard price serves as an anchor that is used to judge the other prices. That is, when buyers decide whether to buy a product, she/he observes the actual price of the product then compare this to the perceived price, that is either the standard price (i.e. the prevalent price in market) or the last price observed or the average of the past market prices. Many models using this insight have been developed (Thaler

1985, Winer 1986, Putler 1992, Isoni 2010). In these models it is normally assumed that buyers decisions are made by comparing a reference price (i.e. “an adaptation level for the product’s price that is based primarily on past price levels for the product”; Putler 1992 p. 289) to the actual price. If the actual price is higher than the reference price then the consumer perceives a loss, if its lower she/he perceives a gain. Consumers are assumed to be loss averse (Kahneman and Tversky 1978). That is the utility produced by a gain of £20 is less than the utility in absolute value produced by a loss of the same amount.

Our experiment is a posted offer market with a computerised seller. In the first phase subjects face a consistent pricing strategy that could be low, medium or high. The range of each pricing strategy is relatively narrow (20 experimental points) so it is likely that subjects form a perceived price that, as the literature on anchoring suggests, is the average price observed in that phase. When in the second phase the price suddenly changes, except in T3 where the price is always drawn from the medium distribution, subjects that compare the new prices to the perceived ones are likely to change their purchasing behaviour. If in the first phase prices were lower, an increase in the price level in the second phase it is likely to be perceived by subjects as a loss that leads them to buy less units that they would have bought otherwise. If on the other hand prices were higher in phase 1, a decrease in the price level would be perceived as a gain leading them to buy more units. If subjects are loss averse, the variation in quantity bought should be greater in the low-medium treatments than in the high-medium treatment. So if anchoring effects are at work we should observe a greater quantity bought in treatments where the strategy high-medium was implemented than in treatments where the strategy implemented was low-medium. In the treatments where the strategy is always the same (i.e. medium-medium) we should not observed a significant difference since there would not be either a perceived loss or a perceived gain.

3. A close up: Shaping

It is interesting to discuss what exactly it is meant by shaping and how shaping relates to anchoring.

Shaping effects refers to the hypothesis that individuals do not have well defined preferences, as assumed by standard economics textbooks. Several experiments (e.g. Knetch et al 2001, Loomes et al. 2003, Isoni et al. (2010) have shown that bids tend to converge to the market price even if the market price is not associated to individuals' values. This is consistent with Ariely et al. (2003) findings that show that subjects willingness to accept is anchored to irrelevant cues (e.g. three last digits of the social security number). Among others Loomes et al. (2003) test the hypothesis that preferences are altered or shaped by the market rather than discovered, as suggested by Plott (1996). They define the shaping hypothesis as follows:

> "..., in repeated auctions in which prices have no information content, there is a tendency for agents to adjust their bids towards the price observed in the previous market period....The intuition behind the hypothesis is that, prior to her involvement in a specific market, an agent may not have well-articulated preferences waiting to be 'discovered'. Instead, values may only be partially formulated and/or imprecise, so that when confronted by an elicitation mechanism, responses are generated using heuristic in which market prices act as cues" (pp. C155, C156)

As we can see from this passage, the market affects subjects preferences through a dynamic process where the market price acts as an anchor or reference price. So we can think of shaping as a phenomenon that is the result of the more general psychological mechanism of anchoring, which is used as a heuristic in the subjective assessment of values (i.e. WTP and WTA). It is worth noting that the anchor in the shaping hypothesis is assumed to convey no information at all. However, this is not a strict requirement for anchoring effects to show. As Monroe and Dodds (1985) argue the willingness to buy (or as they put it preference for a product) is directly proportional to the perceived value of a product which in turn depends on the price, which is used as a reference price (anchor). Reference price perceptions are likely to be affected when the seller's pricing behaviour is consistent and the offering prices (as in the case of sales) are considered plausible by the consumers (Lichtenstein and Bearden 1989). So what that matters is that the price is judged by the buyers as plausible, and this is more likely to happen when prices do not vary too much. In fact were that the case, buyers would either disregard them so that their internal reference price will not be affected or, in the case they do not have one, they will not be able to form it.

To sum up, anchoring and shaping of preferences are two closely related phenomena. Anchoring is a psychological mechanism that affects individual's judgments of any sorts, while shaping relates to preferences. In this context the market price serves as an anchor, or reference price, that subjects use to adjust their bids or make purchasing decisions.

Shaping effects are normally measured in experiments in terms of WTP and WTA convergence towards the market price. This implies that the variance of bids decreases over time (Loomes et al. 2003). The market institutions that have been implemented are variations of the Vickrey auctions. However in a posted market institution with a computerised seller, as is the case in this experiment, the measurement of shaping effects cannot rely on the convergence of bids towards the market price, since the price is given. They only variable that can be used is the quantity bought by subjects. If shaping effects are at work then we should observe a greater quantity bought in the second phase of the treatments that implement a high-medium pricing strategy than in the treatments that implement a low-medium pricing strategy. As can be noticed the predictions are the same as in the case of anchoring, but this should not be surprising, being anchoring the underlying psychological mechanism of shaping.

One may wonder to what extent shaping is a social phenomenon, considering that bids are normally influenced by the market price which is a concise index of what other agents in the market do. As noticed before, the main difference between anchoring and shaping is the fact that shaping relates to preferences while anchoring relates to any sort of judgment. Anchoring is not a social phenomenon, Tversky and Kanheman (1974) shows that subjects' estimates relative to the number of African countries is affected by random numbers that vary among subjects. Similarly Ariely et al (2003) find that subjects with higher social security numbers had a WTA higher than that of subjects with lower social security numbers. There is therefore no reason to expect that shaping only shows in experimental markets where subjects know what other subjects are doing both in terms of quantity bought or in terms of products evaluation. Nor is there any reason to expect shaping to be present in certain markets but not in others, being anchoring a heuristic not peculiar to a specific context. However, it has to be stressed that, in line with the results we obtain in the previous experiment and with what said above, it is likely that shaping effects will be triggered when prices are not too volatile. if prices are too erratic subjects may not be able to form an internal reference price (anchor) and therefore to perceive price differences that are a key element (at least in the literature reviewed here) in buying decisions. Variance of prices can therefore be considered a determinant factor of shaping effects, and for this reason we have decided to implement in this experiment consistent pricing strategies. The intuition[10] behind is that when subjects are presented with a consistently low pricing strategy in the first phase, they will form a reference price that influences the perceived value of the lottery, that will therefore be relatively low. When in the second phase however, the pricing strategy increases, subjects will buy less because the new price is too high when compared to the low value they have assigned to the product. After a while however, when they adapt to the new price range, the anchor will change (so preferences will again being shaped) and so the perceived value of the product, as a result of this the quantity bought will increase again. The opposite will happen when the high pricing strategy is followed by the medium one (i.e. in the second phase the price decreases).

4. General Outline

This task is an individual task framed as a posted offer market[11] with only one product involved. The reason why we frame this as a posted offer market is because we want to reproduce, as much as possible, real retail markets where sellers post

[10] This intuition is in line with Winer's paper (1986). Winer develops and tests a reference price model where households purchasing behaviour is assumed to depend on the discrepancy between the expected price, which in turn depends on the past price and on a trend, and the actual price.

[11] The word market has also being retained to make this task clearly distinguishable from the previous task and to indicate as well that the set up is almost completely similar to a posted offer market with the difference that there is a computerised seller and no rationing is possible (buyers do not interact in any way between each other).

prices and buyers decide how much they want to buy. For reasons explained before however we cannot have a real seller because we want to have control over the pricing strategies used and this can be easily achieved by using a computerised seller. Before beginning the task, subjects completed a questionnaire to check their understanding; they could then ask questions of clarification. After the experimenters answered any question of clarification the posted offer market task started. It involved a trial period with a example product[12] (table 7) and two phases, phase 1 and phase 2 of 10 independent trading periods each.

Outcomes	Probabilities	Results
1	35.00%	50
2	15.00%	15
3	18.00%	80
4	22.00%	139
5	10.00%	10

Table 7: Example Product

We use the same lotteries used in the first part of the experiment, S1, S2, S3, C1, C2, C3, VC1, VC2 and VC3 (see table 2 in part 1) except for the lottery used in the trial periods (table 7) , however each subject played only with one of these lotteries.

It is not in any way relevant for the scope of this experiment to believe that products of different levels of complexity differ only in terms of complexity. The key message is simply that, depending on the experimental treatment, we used at least three of these products, or all nine of them, to verify the robustness of our results, as detailed below.

5. Experimental design

Each trading period subjects were endowed with 650[13] experimental points. Their task was each period to decide whether and how much they wanted to buy of

[12] The reason why we used an example product in the trial periods have already been explained in the previous chapter. We did not want to disclose information about the products we were going to use in the experiment.

[13] In our previous experiment the endowment was 390 experimental points, however, since these lotteries have an expected value of 107 while the ones used in the other experiment had an expected value of 60, we had to change the endowment accordingly, to keep the incentive system consistent across our experiments.

a product on sale at a randomly chosen price. The product did not change throughout the experiment and was one of the products presented in Table 2.

The key variable of the second task is the price. The main purpose of this task is to test for shaping effects and verify their robustness across a range of products and to check whether we can detect any complexity effects. In each period a price is chosen randomly from a uniform price distribution within a specific range. We label the price distributions as Very Low (ranging between 57 and 77), Low (87 – 107), Medium (117 – 137), High (147 – 167) and Very High (177-197). In all experimental treatments, phase 2 (10 periods) is run with prices drawn from the Medium distribution; in phase 1 however the treatments differ in the price distributions used. We began the experiment by running treatments 1, 2 and 3. Each of these treatments was run with 108 subjects per product sold, and all 9 products were used with different subjects, so a total of 324 subjects participated in these treatments. Treatment 2 is our control treatment: prices are drawn from the Medium distribution throughout the experiment, including phase 1. Treatment 1 draws phase 1 prices from the Low distribution; Treatment 3 from the High distribution. Treatment 1 implements a 'low-high' strategy where prices begin low (phase 1) and they then increase (phase 2); conversely, treatment 3 implements a 'high-low' strategy where prices begin high and then decrease.

Price Distribution	Range
VeryLow	57-77
Low	87-107
Medium	117-137
High	147-167
VeryHigh	177-197

	Periods			Subjects
Treatment	1-5	6-10	11-20	
0	VL	L	M	30
1	L	L	M	108
2	M	M	M	108
3	H	H	M	108
4	VH	H	M	30

Table8: Market Task: Summary of Treatments

As we found our results (described in section 3) to be insensitive to the level of product complexity, and we had an opportunity to run a small number of additional sessions, we ran an additional two treatments, 0 and 4, with three products (namely S1, S2 and S3), to test further the effectiveness of different dynamic strategies on sales and profits. An additional 60 subjects participated to these treatments. Treatment 0 draws prices from the Very Low distribution in periods 1-5 and from the Low distribution in periods 6-10; Treatment 4 draws prices from the Very High distribution in periods 1-5 and from the High distribution in periods 6-10. Overall, the average price in phase 1 increases as we move from treatment 0 to 1, 2, 3 and 4, with treatment 0 also implementing a low and rising price strategy and treatment 4 a high and decreasing price strategy. A summary of the experimental treatments is provided in Table 8.

5.1 Earnings

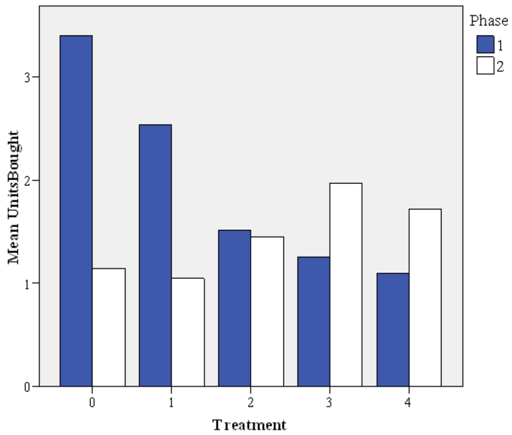
Earnings for the posted offer market task

The earnings are the points accumulated in the experiment. The product on sale in the posted offer market tasks is played out at the end of the experiment to determine its value, and then multiplied by the number of units bought throughout the second part of the experiment. The unspent endowment accumulated throughout the 20 periods is also part of the earnings.

6 Experimental Results

6.1 Sales and Expenditure

Figure 5 reports the quantity bought in phase 1 and phase 2 for each treatment.



Units Bought

Treatment	Overall	Phase 1	Phase 2
0	2.27	3.40	1.14
1	1.79	2.54	1.05
2	1.48	1.51	1.45
3	1.61	1.25	1.97
4	1.41	1.09	1.72

Figure 5: Average Quantity Bought Broken Down by Phase

Phase 1 average prices increase as we move from treatment 0 to 4 and, unsurprisingly, this is borne out in a declining quantity bought. In the absence of shaping effects, it is not obvious however that we should observe significantly differences in the quantity bought in phase 2, since the pricing strategy is the same across treatments in this phases and so prices are on average the same.[14] Conversely, the shaping effects prediction is that in treatments 3 and 4 we should observe greater sales than in the other treatments, and especially treatments 0 and 1. Figure 5 shows that sales are higher in treatments 3 and 4 than in the other treatments, with treatment 3 phase 2 sales being around 80% greater than those in treatment 0 and 1. Treatment 3 sales are significantly greater than those in treatments 0, 1 and 2 (Mann Whitney $p < 0.01$); the shaping effect prediction of greater sales in treatment 4 than in treatment 0 and 1 is also supported (Mann Whitney $p < 0.05$), though not for treatment 2. We also perform a Kruskal-Wallis test whose result shows that the differences in units bought in phase 2 across treatments are overall statistically significant ($p < 0.001$). Therefore, it appears that a different pricing strategy in phase 1 has an effect in phase 2 sales which is consistent with shaping effects, as more sales occur with a 'high and reducing' pricing strategy than with a 'low and increasing' pricing strategy or the control treatment, particularly in the case of treatment 3[15].

Treatment 3 seems the more effective pricing strategy. When we move from Strategy 0 to 1 (price sequence respectively Very Low – Low – Medium and Low - Medium) and from strategy 3 to 4 (price sequence respectively High– Medium – Very High- High- Medium) the number of units bought does not change significantly. Gradual adjustment of prices from extremely high or low levels seems to produce not particularly significant differences.

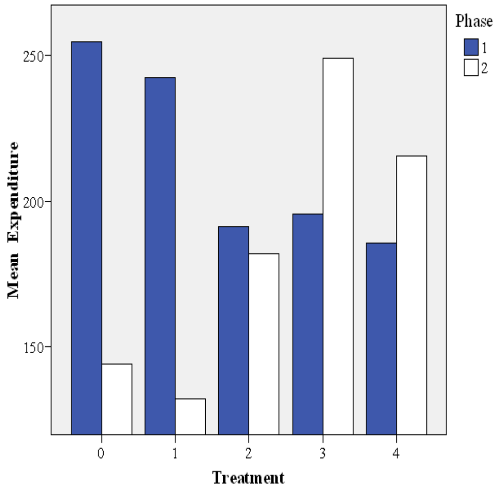
Result 4: The number of units bought is significantly greater in treatment 3 and 4 (high-low strategies) than in treatments 0 and 1 (low-high strategies). This is consistent with shaping effects.

[14] There may be income and portfolio effects leading to different outcomes in phase 2. These are discussed and controlled for in the regression analysis of section 2.3. We also checked, and were able to confirm, that the prices chosen randomly by the computer were indeed on average the same across treatments.

[15] A more detailed discussion of the results, linked to the psychological mechanism of anchoring and loss aversion if provided in the appendix D.

.

Higher sales do not necessarily mean higher expenditure and thus higher revenue for firms, of course. For example, high sales with a low price may imply lower revenue for firms than low sales with a high price. Figure 6 shows the average expenditure broken down by phase.



Expenditure

Treatment	Overall	Phase 1	Phase 2
0	199.45	254.71	144.19
1	187.33	242.41	132.24
2	186.60	191.23	181.98
3	222.29	195.47	249.11
4	200.54	185.74	215.34

Figure 6: Average Expenditure across Treatments Broken Down by Phase

In phase 1 the expenditure decreases as we move from treatment 0 and 1 to the others, reflecting the price elasticity of demand for given average price levels across treatments. In phase 2 the expenditure is lower in treatment 0 and 1 than in 2, 3 and 4. The shaping effect predicts that expenditure is greater in treatment 3, and this is supported in relation to treatments 0 ($p < 0.001$), 1 ($p < 0.001$) and 2 ($p = 0.01$); and for 4 in relation to treatment 0 ($p < 0.06$) and 1 ($p < 0.05$). Figure 6 confirms that, even though prices per treatment are on average the same, phase 2 expenditure is greater by around 50% or more in treatments 3 and 4 than in treatments 1 and 2, and, as before, the treatment 3 pricing strategy seems the most effective. A Kruskal-Wallis test confirms that the differences among treatments in phase 2 are globally statistically significant ($p < 0.001$).

Result 5: Consistently with shaping effects, expenditure in phase 2 is significantly greater in treatments 3 and 4 that implement a high-low pricing strategy than in treatments 0 and 1 that implement a low-high pricing strategy.

Product	Treatment 1			Treatment 2			Treatment 3		
	Phase 1	Phase 2	Overall	Phase 1	Phase 2	Overall	Phase 1	Phase 2	Overall
S1	2.80	1.17	1.98	1.83	1.70	1.76	1.03	2.08	1.55
S2	2.45	0.97	1.71	1.52	1.44	1.48	0.97	1.85	1.41
S3	1.73	0.77	1.25	1.26	1.43	1.34	1.16	1.72	1.44
C1	2.43	1.08	1.76	1.92	1.98	1.95	1.10	1.84	1.47
C2	3.08	1.27	2.17	1.28	1.04	1.16	2.08	2.68	2.38
C3	3.48	1.01	2.25	1.22	1.03	1.13	0.91	1.61	1.26
VC1	1.58	0.60	1.09	1.19	0.91	1.05	1.35	2.03	1.69
VC2	2.30	1.49	1.90	1.78	1.94	1.86	1.28	1.66	1.47
VC3	2.99	1.08	2.04	1.63	1.59	1.61	1.41	2.29	1.85
Type 1	2.27	0.95	1.61	1.64	1.53	1.59	1.16	1.99	1.57
Type 2	2.61	1.24	1.93	1.53	1.48	1.50	1.44	2.06	1.75
Type 3	2.73	0.95	1.84	1.37	1.35	1.36	1.16	1.87	1.52
Simple	2.33	0.97	1.65	1.53	1.52	1.53	1.05	1.88	1.47
Complex	3.00	1.12	2.06	1.47	1.35	1.41	1.36	2.04	1.70
V. Complex	2.29	1.06	1.68	1.54	1.48	1.51	1.35	1.99	1.67
Overall	2.54	1.05	1.79	1.51	1.45	1.48	1.25	1.97	1.61
	Treatment 0			Treatment 4					
Product	Phase 1	Phase 2	Overall	Phase 1	Phase 2	Overall			
S1	3.85	1.05	2.45	0.72	0.86	0.79			
S2	3.07	1.32	2.20	1.13	2.12	1.63			
S3	3.28	1.06	2.17	1.43	2.17	1.80			
Overall	3.40	1.14	2.27	1.09	1.72	1.41			

Table 9: Average Sales by Treatment, Phase and Product

Product	Treatment 1			Treatment 2			Treatment 3		
	Phase 1	Phase 2	Overall	Phase 1	Phase 2	Overall	Phase 1	Phase 2	Overall
S1	267.65	147.66	207.65	232.68	212.35	222.52	160.54	261.78	211.16
S2	236.27	121.61	178.94	190.65	182.10	186.38	149.64	233.74	191.69
S3	163.26	96.23	129.75	159.58	178.17	168.88	180.58	215.98	198.28
C1	231.86	137.05	184.45	242.19	248.63	245.41	172.84	232.78	202.81
C2	293.13	158.35	225.74	160.11	129.70	144.90	322.40	338.15	330.28
C3	332.39	128.68	230.53	154.71	129.60	142.15	322.40	203.74	172.35
VC1	150.87	75.93	113.40	148.97	112.74	130.85	210.91	258.54	234.73
VC2	219.02	188.33	203.67	225.41	243.14	234.28	200.42	207.55	203.98
VC3	287.30	136.37	211.83	206.76	201.36	204.06	220.94	289.69	255.32
Type 1	216.79	120.21	168.50	207.95	191.24	199.59	181.43	251.03	216.23
Type 2	249.47	156.09	202.78	192.06	184.98	188.52	224.15	259.81	241.98
Type 3	260.98	120.43	190.70	173.68	169.71	171.70	180.83	236.47	208.65
Simple	222.39	121.83	172.11	194.31	190.87	192.59	163.59	237.16	200.38
Complex	285.79	141.36	213.58	185.67	169.31	177.49	212.06	258.23	235.14
V. Complex	219.06	133.54	176.30	193.71	185.75	189.73	210.76	251.93	231.34
Overall	242.41	132.24	187.33	191.23	181.98	186.60	195.47	249.11	222.29
	Treatment 0			Treatment 4					
Product	Phase 1	Phase 2	Overall	Phase 1	Phase 2	Overall			
S1	288.69	132.26	210.48	123.75	106.62	115.19			
S2	231.47	167.99	199.73	188.02	265.20	226.61			
S3	243.98	132.33	188.16	245.46	274.19	259.83			
Overall	254.71	144.19	199.45	185.74	215.34	200.54			

Table 10: Average Expenditure by Treatment, Phase and Products

Tables 9 and 10 broadly show the same pattern of results for different products and complexity levels. This is confirmed in statistical tests checking, e.g., for whether product complexity has any impact on sales and productivity. We found no evidence that it did. However this is discussed later in more detail in the regression analysis section.

Result 6: Product complexity does not affect sales and expenditure.

6.2 Regression analysis

While the analysis of the previous section is useful, it does not control for a number of other factors such as period by period dynamics in prices and sales, learning, and the current asset portfolio held by each consumer. We do so through the regression analysis considered in this sub-section. We used random effect regressions controlling for subject level effects.[16]

We then estimated two models that have the same regressors but a different dependent variable: sales and expenditure. We do that both for phase 1 and phase 2, although phase 2 is the most relevant for testing shaping effects.

6.2.1 Description of the models

In model 1 the dependent variable is the quantity bought. We expect the number of units bought to depend negatively on the price. The lagged price shows whether the number of units bought depends on the past price. In the previous experiment we found that sales were *positively* dependent on past prices, and interpreted this as period by period evidence for shaping effects. Products1 is a dummy whose value is 1 when the products are S1, C1 and VC1 and zero otherwise. Products2 is 1 when the products are S2, C2 and VC2 and zero otherwise. Complexity is 0 when the product is simple (S1, S2, S3), 1 when the product is complex (C1, C2, C3) and 2 when the product is very complex (VC1, VC2, VC3).

[16] For our data, which comes from repeated observations by the same subject, panel models are more appropriate than spatial models (such as error clustering). See Baltagi (2005) for a discussion. However we also run cluster regressions, and the results did not change.

The lagged income is the number of experimental points that each subject has earned so far in the experiment. The income is computed as the endowment saved plus the number of units bought times the expected value of each unit (that is, 107). It is not actual subjects' income, since the return of the product they have bought will be known at the end of the experiment. With this variable we control for possible income and risk aversion effects; as it is normally assumed that wealthier subjects are more risk loving, these effects would work in the direction of a positive coefficient (richer subjects should purchase more units of the product). Lagged Total Sales is the overall number of units bought up to the previous period. It shows whether subjects take into account the number of units they have bought in the previous periods to decide how much they want to buy of the current one. This variable helps to control for portfolio effects. In particular, it can be argued that a rational investor has an optimal portfolio to achieve over the 20 periods (in terms of safe and risky assets) but does not know how prices are determined. She initially assumes that prices are based on average as per the prices observed so far (as she has no better information). Given those prices, she aims to obtain a given number of units initially and at any given point of the experiment, and decides how many units to buy in each period based on her price expectation. When the prices increase in phase 2 of treatments 0 and 1, she realises that now prices are much higher and, given her optimal portfolio of assets, the new prices and the amount of units she already holds, she will become the more reluctant to buy further units the more units she has already bought. The reverse argument applies if subjects are in treatments 3 and 4. If there are meaningful portfolio effects, we should therefore expect a negative coefficient on Lagged Total Sales, particularly in phase 2 as there is considerable variance in amounts of units bought from phase 1 across treatments (as shown by Figure 5).

The MarketPeriod variable is the period of play, and shows how the number of units bought changes over time. T0, T1, T3 and T4 are 4 dummy variables that are equal to 1 when the treatment is respectively 0, 1, 3 and 4. Treatment 2 is the reference control treatment. When controlling for income and portfolio effects, and the other variables, the sign and statistical significance (if any) of these dummy variables in phase 2 regressions help us test for shaping effects. The variable Complex*Price aims to test whether product complexity affects the elasticity of demand. Other interaction terms relate the period of play and the treatment dummy (e.g., MarketPeriod*T0). With these variables we are able to see whether behavior over time changes across treatments. Finally there are the interaction terms that relate the centered prices and treatments. These variables allow us to see whether subjects' sensitivity to change in prices varies across treatments. The centered price variable is computed in each period as a deviation from the average price observed in the previous periods.[17]

Model 2 differs from model 1 in that expenditure is the dependent variable instead of sales.

Phase 1 – Regressions

Although phase 1 is not strictly relevant to test our shaping effects hypothesis we also run the two models described earlier for this phase, as it is relevant to check for complexity effects.

As expected, the quantity bought depends negatively on the price, and this is reflected on a negative strongly significant coefficient, and positively on the lagged price, being the coefficient positive and strongly significant. This is consistent with period by period shaping effects.

Products1 and Products2 are not significant in both models, that is, providing evidence that the type of lottery does not affect either sales or expenditure, confirming the results we found in the binary choice task. The coefficient of the variable Complexity is not statistically significant, meaning that complexity does not affect subjects' choices. These results are interesting when coupled with the binary choice task results. In the binary task, where we replicate part of Sonsino et al.'s experiment, we are not able to explain the results just by complexity aversion effects, in fact we argue that, more consistently, they could be explained by noise. The results of the regressions make us more prone to think that noise can actually be more relevant than otherwise thought.

Complexity*Price is not statistically significant showing that demand elasticity is not affected by the complexity of the lottery. This is not consistent with the results we obtained in the previous experiment.

[17] Centering prior to computing interaction terms is useful to eliminate undesirable multicollinearity effects (e.g., Marquardt, 1980).

	Model 1			Model 2		
	Dep. Var.: Sales			Dep. Var.: Expenditure		
Regressors	*Coefficient*	*SE*	*p>z*	*Coefficient*	*SE*	*p>z*
Price	-0.040	0.008	0.000	-3.569	0.892	0.000
Lagged Price	0.014	0.003	0.000	1.375	0.391	0.000
Products1	-0.023	0.062	0.704	-3.502	6.978	0.616
Products2	0.029	0.062	0.637	4.387	6.986	0.530
Complexity	-0.169	0.174	0.329	-24.654	19.620	0.209
Complexity*Price	0.001	0.001	0.284	0.218	0.152	0.150
Lagged Income	-0.0006	0.000	0.000	-0.161	0.011	0.000
Lagged Total Sales	0.120	0.003	0.000	13.456	0.360	0.000
MarketPeriod	0.182	0.066	0.006	80.922	7.390	0.000
T0	-0.397	0.577	0.492	-5.812	64.558	0.928
T1	-0.620	0.253	0.014	-64.284	28.341	0.023
T3	0.141	0.241	0.559	40.653	26.976	0.132
T4	0.359	0.581	0.537	93.586	64.913	0.149
MarketPeriod*T0	-0.198	0.059	0.001	-10.387	6.626	0.117
MarketPeriod*T1	-0.082	0.024	0.001	-3.978	2.727	0.145
MarketPeriod*T3	0.012	0.024	0.623	-1.628	2.632	0.536
MarketPeriod*T4	0.018	0.058	0.755	-5.182	6.493	0.425
centeredPrice*T0	-0.064	0.011	0.000	-1.043	1.262	0.408
centeredPrice*T1	-0.068	0.011	0.000	-4.348	1.189	0.000
centeredPrice*T3	0.017	0.011	0.101	1.149	1.178	0.329
centeredPrice*T4	0.014	0.011	0.218	0.638	1.262	0.613
Constant	4.488	1.129	0.000	379.138	126.386	0.003

Table 11: Phase 1 Random Effects Regressions results (n=3840)

Both the coefficients of the variables controlling for Lagged Income and Lagged Total sales are strongly significant. Lagged Income coefficient is negative and very small, showing that subjects tend to buy fewer products the more income they have. The coefficient of this variable Lagged Total Sales is positive and strongly significant. This works in the opposite direction as our portfolio effect hypothesis. However, as already stressed, these two variables are more relevant in Phase 2 than in 1. However, a possible explanation for the positive coefficient of the Lagged Total sales is that in some way captures subjects' attitude towards buying. Subjects that buy more at the beginning keep buying more and subjects that buy less tend to do that for the rest of the phase. MarketPeriod is positive, that is, subjects overall tend to buy more as the game goes on. The coefficients of our 4 treatment variables, T0, T1, T3 and T4 are all not statistically significant except for T1. The coefficient is negative, meaning that subjects tend to buy less than in T2, our reference treatment. The fact that the 3 dummies T0, T3 and T4 are not statistically significant show that a different pricing strategy used in those treatments with respect to T2 does not affect subjects' behavior. These dummies are however more relevant in phase 2 where the pricing strategy is the same. So any significant difference can be explained by shaping effects. The interaction terms MarketPeriod*T0, MarketPeriod*T1 are negative and statistically significant, showing that subjects buy more at the beginning of the game and then reduce the amount in later periods at least in treatments T0 and T1. However these results are inconsistent with the sign of the coefficient of the variable MarketPeriod. Probably this is due to a balancing effect when all the treatments are considered, as in this case.

Finally, the coefficients of centeredPrice*T0, centeredPrice*T1 are negative and strongly statistically significant. This shows evidence that subjects are more sensitive in changes in prices when the price is low than when is high. This result is somewhat intuitive. When the price is low, for example 10, an increase of 1 is proportionally greater than when the price changes from 100 to 101, this makes subjects more sensitive to price changes when it is low..

Model 2 differs from model 1 in that the dependent variable is expenditure. As can be noticed from the table the results, except for the size of the coefficients, are the same as for model 1. The main difference is that MarketPeriod*T0, MarketPeriod*T1, MarketPeriod*T3, MarketPeriodT4, centeredPricesT0, centeredPriceT3 and centeredPriceT4 that are not statistically significant, with the only exception of centeredPriceT1 where the coefficient is negative and statistically significant showing that subjects are more sensitive to changes in the price with respect to T2.

Result 7: The regression analysis shows evidence that complexity and type of product in phase 1 have no relevant effect on sales and expenditure.

Phase 2 regressions

The results of phase 2 are more relevant to what we want to test, i.e. shaping effects (table 12).

Consistently with the law of demand, sales are negatively affected by current prices. This is also reflected in lower expenditure. The lagged price has a positive effect on sales: the higher the price in the previous period, the greater the sales and expenditure in the present period. As discussed earlier, this can be interpreted as period by period evidence for shaping effects and replicates our previous experiment results. In the regression analysis reported in the last chapter we used the variables LagAvPrice, LagAvPricexIC, LagMinPrice and LagMinPricexIC. The size of these coeffients is in line with the size of the variable Lagged Price in this regression. In particular the size coefficients of LagAvPrice and LagMinPrice ranges from 0.003 to 0.007. The size of coefficients of the variables LagAvPricexIC and LagMinPricexIC, which ranges from 0.042 to 0.064, is however relatively greater than the size of the coefficient of the variable Lagged Price used in this regression. The reason for this difference can be explained by referring to the kind of variables used. In this regression we use the price observed in the previous period while in the other regression we use an index which summarizes a different kind of information, that is the average past price or minimum past prices. We can expect this index to reflect more the shaping effects (as explained in sections 2 and 3) than the last price observed. This then explains why the coefficient of LagAvPrice and LagMinPrice have a similar size to the coefficient of the lagged price used in this regression. The reason is that we measure shaping in the treatments IC where a consistent pricing strategy is used, while the other treatments were not designed to trigger them, as already discussed in chapter 2. The shaping effects are therefore present but weaker. However we also use more informative indexes and this makes the coefficients of that regression undistinguishable in size from the coefficient of this regression.

Neither the types of products used nor their different complexity level, at least as we have measured it, have a significant effect on sales or expenditure. This result confirms the results we found in the binary choice task. Complexity*Price is not statistically significant, so demand elasticity for more complex products is not significantly different than the one for simpler ones. This result is consistent with what we find in phase 1 but does not confirm what we found in our previous experiment. In that experiment demand elasticity was greater for more complex products. There are wealth effects, but they work in the opposite direction than predicted, with richer subjects buying and spending less. The coefficients on Lagged Total Sales are statistically significant but positive, which is the opposite sign to the one that we would expect to find if portfolio effects were important.

When controlling for all these effects, as noted earlier, the T0, T1, T3 and T4 are especially useful to identify shaping effects. If there were no significant shaping effects, there should not be any differences between the treatments being the pricing strategy the same.

	Model 1			Model 2		
	Dep. Var.: Sales			Dep. Var.: Expenditure		
Regressors	*Coefficient*	*SE*	*p>z*	*Coefficient*	*SE*	*p>z*
Price	-0.074	0.009	0.000	-8.009	1.153	0.000
Lagged Price	0.007	0.002	0.000	0.913	0.231	0.000
Products1	0.001	0.081	0.987	0.163	10.288	0.987
Products2	0.076	0.082	0.350	9.746	10.302	0.344
Complexity	0.147	0.478	0.758	11.600	60.076	0.847
Complexity*Price	-0.001	0.004	0.747	-0.097	0.471	0.837
Lagged Income	-0.0008	0.000	0.000	-0.095	0.010	0.000
Lagged Total Sales	0.033	0.002	0.000	4.141	0.262	0.000
MarketPeriod	0.432	0.054	0.000	54.164	6.845	0.000
T0	-1.279	0.489	0.009	-149.548	61.413	0.015
T1	-1.640	0.278	0.000	-197.474	34.890	0.000
T3	1.374	0.275	0.000	185.496	34.560	0.000
T4	0.772	0.474	0.103	111.920	59.504	0.060
MarketPeriod*T0	0.025	0.022	0.269	2.789	2.794	0.318
MarketPeriod*T1	0.042	0.015	0.005	5.101	1.863	0.006
MarketPeriod*T3	-0.057	0.015	0.000	-7.415	1.869	0.000
MarketPeriod*T4	-0.012	0.022	0.600	-1.871	2.799	0.504
centeredPrice*T0	0.048	0.012	0.000	5.760	1.535	0.000
centeredPrice*T1	0.053	0.007	0.000	6.285	0.939	0.000
centeredPrice*T3	0.019	0.007	0.010	2.879	0.936	0.002
centeredPrice*T4	0.027	0.012	0.023	3.758	1.512	0.013
Constant	9.392	1.208	0.000	1009.595	151.687	0.000

Table 12: Phase 2 Random Effects Regressions Results (n=3456)

The coefficients on T0, T1, T3 and T4 are all significant and in the directions predicted by shaping effects. Our benchmark treatment is the control treatment 2 where the pricing strategy does not change across phases. In T0 and T1 prices are lower in phase 1 and then increase in phase 2. The estimate of those coefficients is negative and strongly significant. In these treatments subjects buy and spend considerably less in phase 2 than in the same phase in treatment 2. Conversely, in treatments 3 and 4, where prices are higher in phase 1 than in phase 2, the coefficients are positive, larger and more statistically significant in the case of treatment 3 than in treatment 4, thus confirming the picture from section 3.1 of greater buying and spending with a 'high low' strategy[18]. These results are in contrast with the imprinting account given by Ariely et al 2003. In their experiment

[18] See appendix D for a more detailed discussion of these results and how they relate to perceived loss and gain.

subjects were exposed to three different anchors, low medium and high, high medium and low. They found that the WTA of subjects exposed to the decreasing treatment was lower than that of subjects exposed to the decreasing treatment. This suggests that the first anchor has a stronger effect than the last one. This result is consistent with the shaping effects that we find overall in this experiment, but is in contrast with the results we find for treatments 0 and 4. Since in those treatments the first anchor was respectively lower and higher than the one employed in treatments 1 and 3, we should expect stronger shaping effects. It has to be noticed that, whether or not this is the reason why we do not find evidence of imprinting account, Ariely et al. only use one anchor for one period, while we use anchors for more than that. So, it may be that the effect of the anchor provided in the first half of the phase is weakened by the anchor provided in the second half of that phase. This would suggest, if the interpretation is correct, that imprinting is not a long lasting phenomenon and it also depends on the consistency of the pricing strategy used (that is the more the prices are erratic the more difficult it is for individuals to form an anchor, as discussed in sections 2 and 3).

Sales and expenditure generally increase with time. The reverse signs on the interaction terms of the treatment dummies with MarketPeriod – statistically significant in the case of MarketPeriod*T1 and MarketPeriod*T3 – suggest that the shaping effect tends to partially decrease in size with time as subjects adjust to the new reference price. This being said, as late as in the last period of the experiment (period 20), in treatment 3 subjects still buy on average 0.8 units more than in the baseline treatment 2, and some 2 units more than in treatment 1. A final result is that of a decreasing price sensitivity of demand in treatments other than the baseline treatment 2. This is consistent with shaping effects: due to the large average price change relative to the reference price, subjects pay less attention to the exact value of the price than they otherwise would.

Result 8: The regression analysis supports the existence of treatment level and period by period shaping effects in sales and expenditures even controlling for a number of factors, such as price, product type, and income and portfolio effects.

Result 9*:* As for phase 1, we do not find evidence that complexity and type of products affect subjects' decisions.

6.3 Profitability Analysis

So far we have seen that the differences in the pricing strategies bring about shaping effects leading to significant differences in sales and expenditure in phase 2. The question then becomes whether the increased sales and expenditure of the 'high low' strategy of treatments 3 and 4 is successful in making such a strategy profitable for firms. Two considerations are relevant for determining this. First, we need to pool phase 1 and phase 2 as the firm may rationally decide to make a loss in phase 1 if this is more than compensated by gains in phase 2: the key is the overall profitability of a given strategy rather than the profitability of part of it. Second, profits are equal to revenue minus costs. In our setup, revenue is the consumer expenditure, but we need to make assumptions about costs in order to be able to provide an answer on profits. Furthermore, the size of the market will also matter as it may affect the marginal cost of producing more units. We assume a simple and general cost function which allows us to model both different market sizes and different returns to scale:

$$(1) \qquad Y = x^{\alpha}$$

Y is the cost and x is the number of units produced, which in our case corresponds to the number of units sold. According to the value that we choose for α, we can estimate the cost for decreasing ($\alpha < 1$), constant ($\alpha = 1$) and increasing ($\alpha > 1$) returns to scale. The values that we used are $\alpha = 0.5$, 1, 1.5 and 2, and so cover all three cases. To estimate the costs, and assuming a market with a single consumer, we calculated the average number of units bought in each period for every treatment and then substituted this value into the cost function. From these estimated costs we then derived the profits for each period and treatment. In order to get cost and revenue estimates for different market sizes, we also multiplied the average number of units in each period by factors of 10, 100, 1,000, 10,000, 100,000 and 1,000,000. Table 14 in the following page describes the results of our profitability analysis: it shows the average estimated profits per treatment and cost function. The interesting values are the positive ones, as firms obviously would not choose to produce if they are making a loss.

We find that, for all our α and market size combinations yielding positive profits, profit values are greater for treatments 3 and 4 than for the other treatments. For every profitable α and market size combination, the predication that profits are greater with a 'high low' strategy is always supported (at $p < 0.05$ or better) for the treatment 3 strategy;[19] the same prediction receives some, though less, support for treatment 4, especially if α = 1.5 or 2. Table 15 in page 78 contains the relevant Mann Whitney P values.

		Treatment				
α	**Market Size**	0	1	2	3	4
0.5	1	203.17	187.47	188.29	223.40	201.30
	10	203.40	187.67	188.48	223.59	201.49
	100	203.48	187.73	188.53	223.65	201.54
	1000	203.50	187.75	188.55	223.67	201.56
	10000	203.51	187.76	188.56	223.68	201.57
	100000	203.51	187.76	188.56	223.68	201.57
	1000000	203.52	187.76	188.56	223.68	201.57
0.75	1	199.16	188.03	187.22	222.26	200.29
	10	199.94	188.69	187.81	222.88	200.85
	100	200.38	189.07	188.14	223.23	201.16
	1000	200.62	189.28	188.32	223.43	201.34
	10000	200.76	189.40	188.43	223.54	201.44
	100000	200.84	189.47	188.49	223.60	201.50
	1000000	200.88	189.50	188.52	223.64	201.53
1	1	201.19	186.00	187.08	222.07	200.16
	10	201.19	186.00	187.08	222.07	200.16
	100	201.19	186.00	187.08	222.07	200.16
	1000	201.19	186.00	187.08	222.07	200.16
	10000	201.19	186.00	187.08	222.07	200.16
	100000	201.19	186.00	187.08	222.07	200.16
	1000000	201.19	186.00	187.08	222.07	200.16
1.25	1	197.97	187.42	186.93	221.85	200.02
	10	195.66	185.76	185.65	220.42	198.82
	100	191.55	182.80	183.39	217.88	196.68
	1000	184.25	177.55	179.36	213.37	192.88
	10000	171.26	168.21	172.19	205.35	186.11
	100000	148.16	151.61	159.46	191.08	174.08
	1000000	107.09	122.07	136.80	165.70	152.68
1.5	1	188.03	176.96	180.49	214.52	194.12
	10	154.55	153.61	163.04	194.71	178.02
	100	48.68	79.76	107.84	132.07	127.09
	1000	-286.12	-153.77	-66.69	-66.02	-33.95
	10000	-1344.84	-892.25	-618.61	-692.44	-543.21
	100000	-4692.82	-3227.55	-2363.94	-2673.34	-2153.64
	1000000	-15280.05	-10612.40	-7883.16	-8937.50	-7246.26
2	1	203.42	187.69	188.43	223.60	201.52
	10	193.22	181.13	184.17	218.48	197.62
	100	100.54	121.40	144.61	171.65	162.09
	1000	-826.27	-475.87	-250.94	-296.65	-193.24
	10000	-10094.31	-6448.52	-4206.50	-4979.62	-3746.48
	100000	-102774.79	-66175.06	-43762.07	-51809.31	-39278.93
	1000000	-1029579.53	-663440.47	-439317.78	-520106.22	-394603.43

Table 13: Estimated Average Profits per Treatment, Market Size and α Values[20]

[19] The only exception is for α = 0.5 in the comparison of treatment 3 with treatment 1, where P = 0.055.

[20] Notes: α is the return to scale coefficient in equation 1: the market size is expressed as the number of consumers buying form the firm.

				Market Size			
				Treatment 0 - Treatment 2			
α	**1**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**
0.5	0.21	0.21	0.21	0.21	0.21	0.21	0.21
0.75	0.14	0.14	0.15	0.15	0.15	0.15	0.15
1	0.19	0.19	0.19	0.19	0.19	0.19	0.19
1.25	0.13	0.13	0.12	0.10	0.08	0.05	0.00
1.5	0.11	0.02	0.00				
2	0.20	0.11	0.00				
				Treatment 0 - Treatment 3			
α	**1**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**
0.5	0.03	0.03	0.03	0.03	0.03	0.03	0.03
0.75	0.02	0.02	0.02	0.02	0.02	0.02	0.02
1	0.02	0.02	0.02	0.02	0.02	0.02	0.02
1.25	0.01	0.01	0.01	0.01	0.00	0.00	0.00
1.5	0.02	0.00	0.00				
2	0.03	0.02	0.00				
				Treatment 0 - Treatment 4			
α	**1**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**
0.5	0.15	0.15	0.15	0.15	0.15	0.15	0.15
0.75	0.10	0.10	0.10	0.10	0.10	0.10	0.10
1	0.14	0.14	0.14	0.14	0.14	0.14	0.14
1.25	0.10	0.10	0.09	0.07	0.04	0.01	0.00
1.5	0.07	0.02	0.00				
2	0.15	0.09	0.00				
				Treatment 1 - Treatment 2			
α	**1**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**
0.5	0.45	0.45	0.45	0.45	0.45	0.45	0.45
0.75	0.45	0.45	0.45	0.45	0.45	0.45	0.45
1	0.45	0.45	0.45	0.45	0.45	0.45	0.45
1.25	0.45	0.45	0.45	0.45	0.45	0.45	0.19
1.5	0.42	0.37	0.00				
2	0.44	0.42	0.09				
				Treatment 1 - Treatment 3			
α	**1**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**
0.5	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.75	0.07	0.08	0.08	0.08	0.08	0.08	0.08
1	0.05	0.05	0.05	0.05	0.05	0.05	0.05
1.25	0.07	0.07	0.06	0.05	0.03	0.00	0.00
1.5	0.05	0.02	0.00				
2	0.05	0.05	0.01				
				Treatment 1 - Treatment 4			
α	**1**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**
0.5	0.31	0.31	0.31	0.31	0.31	0.31	0.31
0.75	0.39	0.39	0.39	0.39	0.39	0.39	0.39
1	0.31	0.31	0.31	0.31	0.31	0.31	0.31
1.25	0.38	0.37	0.37	0.34	0.29	0.12	0.00
1.5	0.27	0.12	0.00				
2	0.31	0.27	0.02				
				Treatment 2 - Treatment 3			
α	**1**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**
0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.00	0.04				
2	0.00	0.00	0.03				

Table 14: Mann-Whitney P Values of Predictions of Greater Profits with a 'High-Low' Pricing Strategy.[21]

---

[21] Notes: the table contains the P values of the relevant pair-wise Mann-Whitney tests. α is the return to scale coefficient in equation 1.

Results 10*:* A pricing strategy of pricing high and then decreasing price, especially as implemented in treatment 3, is more profitable than a constant pricing strategy or a 'low high' strategy.

7. Summary

The market task shows that starting with a high price and then decreasing it can be profitable for firms. Reasons, consistent with rational choice, have been identified in the literature for why a 'high low' strategy may be profitable, in the form, most notably, of price discrimination with durable goods (Conlisk et al., 1984) and of a game theoretical setup quality signalling (Bagwell and Riordan, 1991). Our experimental results suggest that, independently of such reasons, although in the spirit of quality signalling, a strong reason why a 'high low' pricing strategy may be a good one lies in shaping effects: subjects may not have clear preferences about the value of a product and rely on past prices to provide an indication of what such value is. Firms, in turn, may then exploit consumers' bounded rationality to gain more profits than they would otherwise.

PART 3: CONCLUSIONS

This experiment has been designed to test for complexity aversion effects and the robustness of the phenomenon to a change in the individual choice task and for shaping effects in an individual task that uses a posted offer market set up.

In the first task we replicate Sonsino et al.'s results, although we give another explanation to complexity aversion that we also detect. We maintain in fact that noise can both explain pair-wise choices that can be explained by complexity aversion and pair-wise choice that cannot be explained by that. Our analysis shows that in both pair-wise choices, complexity increases and therefore noise can accommodate the data without referring to complexity aversion. Our explanation is therefore more parsimonious. The regression analysis in the second task confirms the results obtained in the experiments discussed in the second chapter. We can therefore conclude that subjects' attitudes towards complexity are not affected by the different nature of the individual task nor by the type of lottery used. It is possible to extend this research in a number of ways. One obvious possibility would be that of using multi-period lottery that Sonsino et al. use in their experiment. Our binary choice deals with lotteries, however it is possible to extend it using tariffs, like the ones that one sees in the real world, (e.g., electricity tariffs). Consider electricity tariffs for example, one could present subjects with two different tariffs of different complexity. So for example, a tariff could be linear (i.e. only one tier) and the other one could be a non linear one (i.e. one tier and a standing charge or two tiers). Subjects then have to choose between the two tariffs, the simple and the complex. Induced values can be set up in such a way that the complex tariff is better than the other one (at the optimal level of consumption) but this is not a strict requirement. Such an experiment may have some relevance for the UK electricity market where it has been shown that a proportion of consumers, when they decide to change tariff or electricity supplier, switch to a worse tariff (Price and Waddams, 2007). Using lotteries in our experiments was a choice dictated by several reasons. The most relevant ones were for comparability purposes with the main work we are benchmarking our work against and for achieving maximum experimental control. It has to be said though that lotteries are highly abstract objects, the advantage therefore of using products (e.g. tariffs) taken from the real world would make it easier to export the results to the relevant markets without, with the proper manipulations, loss of experimental control.

The second task had the main purpose of checking for shaping effects. We find strong evidence that subjects' decisions are affected by past prices. In phase 2 of the low-high price treatments subjects tend in fact to buy fewer products than in the high-low price treatments. Since in phase 2 the pricing strategy does not change we should not observe a different buyers' behaviour as regards to quantities bought. We therefore explain this by shaping effects. Our analysis on the expenditure shows similar results. Expenditure is, from the point of view of a firm, the revenue that it gets from sales. Therefore, if shaping effects are at work in the real world, firms should be aware that they can make more profit just by following a high-low price strategy. We show this in our profitability analysis, although based on a theoretical cost function. A number of extensions to this research are possible, for example trying to get a better understanding of what price dynamics yields the highest profits or verifying how the profit-maximising price dynamics are affected by the presence of multiple products on sale simultaneously.

CHAPTER 4: Models and Experiments – A Methodological Analysis

1. Introduction

Before 1945 economics was commonly thought to be a non-experimental science. The study of economic phenomena was primarily based on models. Since the late 40s however, starting with Chamberlin's experimental markets (1948) and Allais's tests (1953) of Von Neumann and Morgenstern expected utility theory (1944), the boundaries of economics were pushed further. Nowadays, experimental economics, although not yet a mainstream discipline, is widely known and used as an alternative to or combined with other methods of economic enquiry. Experiments have been used to study a wide range of issues. It seems however that most experiments rely on theoretical models to investigate reality. I will investigate whether experimental tests based on those models are a good approach and then I will explore different approaches that do not use models to investigate reality or use models only partially. I will base my analysis on formal methodological theories and concrete case studies: two influential models in economics, Varian (1980) and Bikhchandani et al. (1992), and two influential experiments that purport to test those models, Morgan et al. (2006) and Anderson and Holt (1997).

In order to test a model we need to know what to test, that is, what is its connection with the real world? Although this may seem a simple question to answer, it is not. Theoretical models are often criticized in economics because they employ abstract and unrealistic assumptions and their connection with the real world is vague and not well defined. This has given rise to a heated methodological debate. Therefore, in the first part of the chapter I will review different methodological accounts of models trying to understand how models connect to the real world and then ask whether these accounts are a good description of what Varian and Bikhchandani do. In the second part of the chapter I will ask whether the methodological accounts reviewed in the first part provide different views on how to test experiments. I will then turn to the experimental tests of the models as examples of model-implementing experiments and consider whether this is a good investigative strategy. I will then discuss two different experimental approaches –

illustrated by the experiments of Chamberlin (1948) and Schelling (1960) – that substitute partially or wholly for models. I will then draw some conclusions.

I will show that, despite deep philosophical differences between rival methodological interpretations of models, all these accounts provide a good description of what theoretical models do and provide essentially the same recommendations on how to test models. Specifically, in order for an experimental test to be able to tell us something about the real world (target domain), the experimental lab should resemble the real world in some relevant respects. We will see however that the experiments I will discuss in the next sections implement the models almost completely, and although they claim to test those models as explanation of specific target domains, they turn out to be just tests of certain generic component of the model – specifically hypotheses about individual rationality. I will ask whether it is better to test such generic components in an environment that resembles the model or in isolation. In some cases, I will argue, the model is too complicated for a lab implementation of the model to be a clean test of that component. With respect to those experimental approaches which are either only partially based on models or not based on model at all, I will argue that experiments of these kinds can be seen as models themselves and are informative of the real world by virtue of the similarities between the lab and the target domain.

PART 1: MODELS IN ECONOMICS AND THEIR METHODOLOGICAL STATUS

1. Models in economics

A sub-class of economic models, i.e. theoretical economic models (from now on I will refer to those as models, however there are also experiments that can be considered as models as well but different from the theoretical ones, it will be clear from the context which ones I am referring to) have the common feature of employing highly unrealistic assumptions, some of them often disconfirmed by empirical evidence, that are used to obtain vague empirical claims. In particular, their connection with the real world seems so loose that they have been the object of numerous critiques that raise doubts about their ability to increase our knowledge about economic phenomena. Several methodological theories have been proposed with the purpose of understanding and explaining what an economic model is, how it connects to the real world and how it helps in understanding real phenomena.

Although, from a methodological standpoint, the dispute is far from being settled, it is possible to distinguish three main approaches, which will be examined in the next sub-sections. The first one is the instrumentalist approach that sees models and assumptions as false. Models interpreted in this way connect to the real world only through predictions. If predictions obtained from such a model are confirmed by empirical evidence, the model is a good model even if its assumptions are false. The opposite perspective is the realist approach. According to realists, models isolate real mechanisms; they are therefore representations of reality. If the assertions made by the model turn out to be confirmed, the model is "true" of the real world or represents/describes it correctly (approximately). The third approach is the "fictionalist" one. This approach interprets models as worlds created by the modeller, for this reason they are neither true nor false. They relate to the real world through a relationship of similarity. It is the hypothesis of similarity/empirical claim then that can be true or false depending on whether this similarity or the empirical claim really exists or not. If this hypothesised similarity/empirical claim does exist this does not change the status of the model as a counterfactual world.

Different methodological theories may lead to different recommendations about how to test a model. We will see however that despite the deep philosophical differences between these theories, their recommendations share many common features. In particular they require the lab environment to be similar to the target domain, to which the models are supposed to be applied, in order for the tests to be able to teach us something new about the phenomena under investigation.

1.1 The instrumentalist approach

At least in its conventional reading, Friedman's well-known essay on the methodology of positive economics (1953) is a good description of the instrumentalist view of economic models.

The starting point for Friedman's essay is the open debate that sees economics as often criticised for employing highly unrealistic assumptions. Friedman stresses that this should not be considered as a problem, since many of the most significant models in science employ unrealistic assumptions. What the economist should be interested in is in testing hypotheses derived from those assumptions.

> "The two stages of constructing hypotheses and testing their validity are related in two different respects. In the first place, the particular facts that enter at each stage are partly an accident of the collection of data and the knowledge of the particular investigator. The facts that serve as a test of the implications of a hypothesis might equally have been among the raw material used to construct it, and conversely. In the second place, the process never begins from scratch; the so-called 'initial stage' itself always involves comparisons of the implications of an earlier set of hypotheses with observation; the contradiction of new hypotheses or revision of old ones." (p. 13)

In Friedman's view the hypotheses are the part of a theory that must be confronted with reality. It is misleading to judge a theory by the truth or falsity of its assumptions. It is the truth or falsity of the hypotheses derived from the theory that should be judged.

> "Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate description representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense)." (p. 14)

Therefore, according to Friedman, assumptions are not really what we should be testing, for they are rarely true, in fact most of the times they are not meant to be good representations of reality. What we should consider in the appraisal of a theory is its predictive power rather than the realism of its assumptions.

> "The difficulty in the social sciences of getting new evidence for this class of phenomena and of judging its conformity with the implications of the hypothesis makes it tempting to suppose that other, more readily available, evidence is equally relevant to the validity of the hypothesis –

to suppose that hypotheses have not only 'implications' but also 'assumptions' and that the conformity of these 'assumptions' to 'reality' is a test of the validity of the hypothesis *different from or additional to* the test by implications. This widely held view is fundamentally wrong and productive of much mischief." (p. 14)

Testing an assumption and its conformity to reality is not in any way, according to Friedman, a test of a theory. In order to test a theory we should only test whether its hypotheses, or implications of these hypotheses, conform to reality. Assumptions should not be interpreted to the letter, but as "as if" statements that cannot be tested, for even if we did, the results would not be in any case evidence for or against the theory. To make this point clear, Friedman gives several examples. I shall report one of the most famous ones.

"Consider the density of leaves around a tree. I suggest the hypothesis that the leaves are positioned as if each leaf deliberately sought to maximize the amount of sunlight it receives, given the position of its neighbors, as if it knew the physical laws determining the amount of sunlight that would be received in various positions and could move rapidly or instantaneously from any position to any other desired and unoccupied position. … Is the hypothesis rendered unacceptable because, so far as we know, leaves do not 'deliberate' or consciously 'seek', have not been to school and learned the relevant laws of science or the mathematics required to calculate the 'optimum' position, and cannot move from position to position? Clearly, none of these contradictions of the hypothesis is vitally relevant; the phenomena involved are not within the 'class of phenomena the hypothesis is designed to explain'; the hypothesis does not assert that leaves do these things but only that their density is the same *as if* they did. Despite the apparent falsity of the 'assumptions' of the hypothesis, it has great plausibility because of the conformity of its implications with observation." (p. 19, 20)

Clearly, again, Friedman claims that any evidence against any of the assumptions of a model is not to be considered as evidence against the theory. The reason is simple; the assumptions are not part of the hypothesis of the theory. The theory does not claim that leaves move from position to position or that they deliberately seek to maximise the sunlight they receive, the only claim made is about the density of leaves in a tree. The assumption is false in the domain of application and therefore we should not test it literally.

A second way in which we can interpret an "as if" assumptions is that it is false but is predictively accurate. I will explain with the help of an example that Friedman gives in his essay:

"Consider the problem of predicting the shots made by an expert billiard player. It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas that would give the optimum directions of travel, could estimate accurately by eye the angles, etc, describing the location of the balls, could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas. Our confidence in this hypothesis is not based on the belief that the billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be *expert* billiard players." (p. 21)

In this example the assumption is false, (i.e. the billiard player behaves "as if" he knew), but predictively accurate in the sense that it predicts the behaviour of the billiard player. The difference from the assumption about the leaves lies in the fact that leaves do not even behave as assumed (i.e. they do not move). Friedman suggests that the confidence we have in the hypothesis about billiard player (that is that he behaves as if) derives from the fact that if they were not capable of reaching the same result (that of making good shots), they would not be expert billiard players. Let us consider the assumption about the leaves. Following the same line of reasoning, we should say that:

“Our confidence in this *hypothesis about the density of the leaves* is not based on the belief that the *leaves* can or do go through the process described *(calculate the optimum position that maximises sunlight exposure);* it derives rather from the belief that, unless in some way or other they *(leaves)* were capable of moving to the sunlight-maximising position *they (the leaves) would not have evolved by natural selection”*

However this is false, in fact the evolution process is at the tree level not at the leave level.To put it differently, in the case of the leaves we have two layers of “as if’s”, one regards the behaviour of the leaf, the other one regards their capability of making calculations. In the billiard player example we only have one layer, the capability of making calculations. In this sense, in the billiard player example the assumption is about the billiard player and the hypothesis is about the billiard player, however in the leaves example the maximisation assumption is about the leaves while the behaviour is about the tree. So, the reasoning could be as follows:

“Our confidence in this hypothesis *(i.e.* t*hat the leaves are positioned the way they are)* is not based on the belief that the *leaves* can or do go through the process described; it derives rather from the belief that, unless in some way *trees* were not capable of reaching essentially the same result *they (trees) would not have evolved*.

However in this case no confidence is gained on the hypothesis about the leaves’ behaviour, but only about tree’s behaviour. So, in the leaves example, leaves are the maximising agents, but the predictions are about the tree not the leaves. In the billiard player example, the billiard player is the maximasing agent and predictions are about her/him. So while the leaves assumption is false both about the process the leaves go through and about their behaviour, the billiard player assumption is “true” about the behaviour. This suggests that, although both assumptions are instrumentalist, their degree of “falseness” is different.

If we were to test the billiard player assumption then we should not ask him to solve highly abstract physics problems, in fact if taken to the letter the assumption is false (it is inaccurate as a description of the billiard player's thinking)..

Here is another example of such an 'as if' assumption that relates to economics:

"It is only a short step from these examples to the economic hypothesis that under a wide range of circumstances individual firms behave *as if* they were seeking rationally to maximize their expected returns (generally if misleadingly called 'profits') and had full knowledge of the data needed to succeed in this attempt; *as if*, that is, they knew the relevant cost and demand functions, calculated marginal cost and marginal revenue from all actions open to them, and pushed each line of action to the point at which the relevant marginal cost and marginal revenue were equal." (p. 21)

As in the case of the billiard player, firms do not actually make all the calculations to solve a profit maximisation problem, in fact this is just a way economists express their model.

"The billiard player, if asked how he decides where to hit the ball, may say that he 'just figures it out' but then also rubs a rabbit's foot just to make sure; and the businessman may well say that he prices at average cost, with of course some minor deviations when the market makes it necessary. The one statement is about as helpful as the other, and neither is a relevant test of the associated hypothesis." (p.22)

"Confidence in the maximization-of-returns hypothesis is justified by evidence of a very different character. This evidence is in part similar to that adduced on behalf of the billiard-player hypothesis – unless the behavior of the businessmen in some way or other approximated behaviour consistent with the maximization or returns, it seems unlikely that they would remain in business for long." (p. 22)

The third case is when the "as if" assumption works because there are other factors not included in the theory that are however present in the domain of application and make the assumption work. In this case, again, a literal test of the assumption would not be appropriate because we would be omitting mechanisms/factors that are not described in the theory, and therefore not included in the test, but are relevant in order for the assumption to work.

According to the instrumentalist view of models (if we interpret Friedman's position as instrumentalist), in order to test a theory, we should test its hypothesis in the domain of application of the theory (the phenomena that the hypothesis is designed to explain) but not its assumptions.

The instrumentalist account of models has problems that are well summarised by Hausman (1992, p.166), who interprets that account as using the implicit reasoning schema in (1), (2) and (3):

"1) A good hypothesis provides valid and meaningful predictions concerning the class of phenomena it is intended to explain (premise).

2) The only test of whether a hypothesis is a good hypothesis is whether it provides valid and meaningful predictions concerning the class of phenomena it is intended to explain (invalid from 1).

3) Any other facts about a hypothesis including whether its assumptions are realistic, are irrelevant to its scientific assessment (trivially from 2).

If (1) the *criterion* of a good theory is narrow predictive success, then one is tempted to say that surely (2) is narrow predictive success, and Friedman's claim that the realism of assumptions is irrelevant follows trivially. This is an enticing argument.

But it is fallacious. (2) is not true, and it does not follow from (1). To see why consider the following analogous argument.

1' A good used car drives reliably (over-simplified premise).

2' The only test of whether a used car is a good used car is whether it drives reliably (invalid from 1').

3’ Anything one discovers by opening the hood and checking the separate components of a used car is irrelevant to its assessment (trivially from 2’).”

If the theory is just a set of “as if” assumptions that do not tell us anything about the world, then if the theory does not work we will not be able to understand the reasons why it does not, and consequently what we need to do to improve it. Like checking the engine of a car, a check of any single component of a theory would not be of any help. However, Hausman maintains we are able to learn something by analysing the assumptions.

Without going into depth into this philosophical debate, maybe what Friedman is doing, other than giving clear instructions on how economists should test models, is describing what economists do and providing a way to use unrealistic models to test predictions.

In the section “the use of the assumptions as an indirect test of a theory” (p. 26) Friedman asserts that we can use assumptions to indirectly test a hypothesis if the assumptions can be considered as implications of the hypothesis. However, it is not quite clear when the assumptions can be thought of as implications of the theory.

“More generally, a hypothesis or theory consists of an assertion that certain forces are, and by implication other are not, important for a particular class of phenomena and a specification of a the manner of action of the forces it asserts to be important. We can regard the hypothesis as consisting of two parts: the first, a conceptual world or abstract model simpler than the ‘real world’ and containing only the forces that the hypothesis asserts to be important; second, a set of rules defining the class of phenomena for which the ‘model’ can be taken to be an adequate representation of the ‘real world’ and specifying the correspondence between variables or entities in the model and observable phenomena.” (p. 24)

Here theory and hypothesis are used as synonyms suggesting that a hypothesis (or theory) is an assertion consisting of two parts. This definition of hypothesis seems to imply that the assumptions cannot be considered as implications of a hypothesis unless the hypothesis clearly says so. However later on Friedman seems to suggest that assumptions can be used to get indirect evidence on the hypothesis if we test the implication of a hypothesis outside the domain of application.

"In presenting any hypothesis, it generally seems obvious which of the series of statements used to expound it refer to the assumptions and which to implications; yet this distinction is not easy to define rigorously. It is not, I believe, a characteristic of the hypothesis as such but rather of the use to which the hypothesis is to be put. If this is so, the ease of classifying statements must reflect unambiguousness in the purpose the hypothesis is designed to serve. The possibility of interchanging theorems and axioms in an abstract model implies the possibility of interchanging 'implications' and 'assumptions' in a substantive hypothesis corresponding to the abstract model, which is not to say that any implication can be interchanged with any assumption but only that there may be more than one set of statements that imply the rest" (p. 26, 27)

It is difficult, Friedman maintains, to determine what in a hypothesis refers to the assumptions and what to the implications. This makes it possible to interchange assumptions and implications.

"To state the point more generally, what are called the assumptions of a hypothesis can be used to get some indirect evidence on the acceptability of the hypothesis in so far as the assumptions can themselves be regarded as implications of the hypothesis, and hence their conformity with reality as a failure of some implications to be contradicted….The reason this evidence is indirect is that the assumptions or associated implications generally refer to a class of phenomena different from the class which the hypothesis is designed to explain;…..The weight attached to this indirect evidence depends on how closely related we judge the two classes of phenomena to be." (p. 28)

Assumptions can be used to get some indirect evidence on the hypothesis, if the assumptions are implication of the hypothesis themselves. The evidence, Friedman goes on, is indirect because it is not used to predict the class of phenomena for which it has been designed. The strength of this evidence depends on how related the phenomena in the two different domains are judged to be.

Another way in which we can get indirect evidence on the hypothesis relates to the use of the same assumptions in different theories and therefore, according to Friedman terminology, hypotheses.

> "Another way in which the 'assumptions' of a hypothesis can facilitate its indirect testing is by bringing out its kinship with the other hypotheses and thereby making the evidence on their validity relevant to the validity of the hypothesis in question. For example, a hypothesis is formulated for a particular class of behaviour. This hypothesis can, as usual, be stated without specifying any 'assumptions'. But suppose it can be shown that it is equivalent to a set of assumptions including the assumption that man seeks his own interest. The hypothesis then gains indirect plausibility from the success for the other classes of phenomena of hypotheses that can also be said to make this assumption;" (pp.28-29)

If an assumption has been used successfully to predict different classes of phenomena, like for example assuming self-interested behaviour, then it may give, if it is used to build another theory, reasonable confidence that it may work to predict successfully another class of phenomena.

Friedman's position has usually been interpreted as instrumentalist, as in Hausman's discussion. However, Mäki (2009) suggests a reading of Friedman as partly realist. Mäki maintains that this interpretation is suggested by some passages of Friedman's text. For example, the passage in which Friedman suggests that 'the model can be taken to be as an adequate representation of the real world' (p. 24) can be interpreted as realist. Before however we discuss that, it is better to briefly summarise the realist interpretation of models.

1.2 The realist approach

The realist sees models as representations of the real world. Van Frassen (1980) defines realism in the following way:

> "Science aims to give us, in its theories, a literally true story of what the world is like;"

The realist perspective sees theories as aiming to provide a description of how things really are. In this sense, theories are either true or false, depending on whether the description of the process at hand is correct or not. Theories though do not describe every feature of reality, rather they are isolations whose purpose is that of isolating (theoretically) a mechanism from the influence of other factors that are not of interest to the modeller or thought not to be relevant to explain the phenomenon under investigation. Mäki (1992) suggests that these isolations are achieved using idealising assumptions and omitting other relevant factors that might have an effect on the mechanism the modeller wants to focus on. Idealising assumptions are false in the sense that they omit some relevant features and exaggerate others. Cartwright's account of economic models (2009) can also be considered realist. She suggests that economic models are "blueprints for Galilean experiments". Their purpose is to set up "nomological machines" that show how a "capacity" works without any disturbing factors, that is, in isolation. Let us see what a capacity and nomological machine are according to Cartwright. A capacity is or can be thought of as a property of a factor. This property does not always display according to Cartwright, but there must be an ideal (for the capacity) situation for it to display

> "…capacities are not to be identified with any particular manifestation….in Ryle's account: 'They signify abilities, tendencies, propensities to do, not things of one unique kind, but things of lots of different kinds'" (1999, p. 64)

No element in the model should have any effects that prevent the natural working of a capacity. Capacities display when we set up a nomological machine, which is what, according to Cartwright:

"…a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws". (p. 50)

Models according to Cartwright are then set up to discover how a capacity, and usually only one, works without any other disturbing factors. In this view models are then isolating tools. Economic models in particular are seen by Cartwright as blueprints for "Galilean experiments". In a Galilean idealisation only one cause is analyzed in order to see how this cause affects the process under study. Galileo was able to find out how the earth's pull affects the motion of a body by eliminating all the other causes that might have an effect on the motion of bodies. So, experiments of this kind are able to find out tendencies to use Mill's word, or capacities to use Cartwright's word.

The purpose of economic models, according to Cartwright, is the same as that Galilean experiments. Usually these models only focus on one cause discovering the contribution of that cause to the phenomenon. What Cartwright stresses is that if the general principles that we use in such an idealisation are the correct principles, that is they are true of the target system, then the results of the models will hold true of it. If a model is just like a Galilean experiment should be, then it should be able to tell us what happens in a real world-like situation and more than that. It will also be able to tell us what will happen in an experiment and therefore help us discover the capacity at work.

Models in economics aim at establishing tendencies or capacities of the factor under study. According to Cartwright, economics models have a problem. They describe regularities that hold ceteris paribus (the same is true in physics) however "ceteris paribus conditions have a special role to play in economics laws... They describe the structure of a machine that makes the laws true-" (p. 148 – The dappled world). However, this ceteris paribus conditions describe a very special set of circumstances rather than what usually happens in reality. The shielding assumptions that make the model work are extremely unrealistic. So, rather than having a nomological machine that generates law displaying the working of capacities we have "socio-economic laws" that "are created by socio-economic machines" (p. 149).

We have seen that one of the features of the realist interpretation of models is that they are isolating tools. Models isolate mechanisms omitting factors that are not relevant. Those factors are excluded because either we want to see how that mechanism works in isolation or because those omissions are thought not to be relevant. In this case the "as if" assumptions used do not deny the existence of such factors but only state that they are negligible in the domain of application. Mäki maintains that:

> "The point is this: one should not criticize what appears as a false assumption without understanding what assertion is intended when using it….Such paraphrasing include turning an assumption into an assertion of the negligibility of a factor or of the applicability of a model to a domain of phenomena". (2009, p. 102)

Therefore to the realist, false assumptions are perfectly compatible with realism:

> "Indeed, it has been my contention that theories with false assumptions may be true, and that realism (as a theory of theories) is perfectly comfortable with unrealistic assumptions. The truth-value of a theory cannot be read off the truth-value of its assumptions. The key to understanding the gap between the two is to ask the question: what is the theory about, what claim does it make about the world, if any? A theory may be true about the functioning of some important casual factor while making false assumptions about the existence of functioning of other factors." (p. 101)

This long quote underlines the fact that, according to realist philosophers, unrealistic assumptions are perfectly compatible with the realistic interpretation of models. In fact idealized assumptions serve the purpose of isolating the mechanism described in the model. Therefore, when testing a model we should not be concerned with its assumptions but with the assertions of the theory. As a consequence of this, if rationality is not an assertion we should not test it. In the realist view many assumptions serve the purpose of neutralising the effect of other factors on the isolated mechanism. So for example the assumption of vacuum in Galileo's law implies that other factors such as air pressure have negligible (for the purpose at hand) effects on the phenomenon he is interested in, therefore in this case the assumption should be true, or approximately true for the purpose at hand. Let us consider then the assumption about rationality. This can be interpreted as saying that all the other forces that influence people's behaviour have a negligible effect on the phenomenon under investigation.

Mäki goes on by discussing the use of "as if" assumptions by Friedman. He stresses that usually "as if" assumptions are seen as a distinctive mark of an instrumentalist view. However "this is a mistake" (p.104):

"The as-if formulation is a flexible tool that can be used for expressing a number of ideas about a theory and its relationship to the world. F53 [i.e. Friedman's essay] itself appears to be of two minds as to the import of 'as if'. It is one thing to say that

(a) phenomena behave *as if* certain ideal conditions were met, namely conditions under which only the theoretically isolated real forces are active;

and it is quite another thing to say that

(b) phenomena behave *as if* those forces were real."

The first statement is consistent with realism. The forces that shape the phenomena are real but the shielding conditions are defined in the model. If those shielding conditions were true of the real world, then we would observe the model's conclusions. Mäki presents further evidence from F53 suggesting that not all Friedman's claims are instrumentalist. Mäki stresses that in p. 40 of his essay Friedman asserts:

"…the phenomenon occurs as if nothing but those powerful forces were in operation. The reality of none of these forces is denied. If one wants to identify a fiction here, it is the isolation that is a fiction, not the isolated forces." (p. 105)

Statement (b) on the other hand is instrumentalist. It denies that all the forces included in the model are necessarily real. Friedman's example about the leaves belongs to this category.

So according to the realist view, models isolate mechanisms using abstracting and idealising assumptions. Assumptions are abstractions because they exclude the existence of factors that although present in the real world they are not taken into account because assumed not to have a substantial influence on the mechanism. Assumptions are idealisation in that they represent extreme cases of some instances. For example, one can assume that economic agents are self-interested; this does not imply that they are always self-interested in the real world but that they are, approximately, in the domain where the model applies. Therefore assumptions that are abstractions and idealisation are false.

A key assumption in Friedman essay (and more importantly in economics) regards the profit maximization, however Mäki does not clearly say if the assumption is compatible with realism. What he says is that assumptions can be paraphrased to understand what assertions they make about the real word. It is the assertions then that are claimed to be true (approximately) in the target domain.

"The assumption in question is profit maximazation, understood literally as dealing with goals and computations taking place in businessmen's extended minds: by going through a series of marginalist calculations, they consciously infer what the maximasing abehaviour is, then take action accordingly. The paraphrase of this assumption dispenses with the decision procedures behind overt behaviour, and makes claims about that behaviour only. The assertion now made is about behaviour that ends up with maximum profits – rather than about profit maximizing motives and calculations. F53 indicates optimism that so paraphrased, the assumption has a fair chance of being (approximately) true.” (p. 104)

An assumption can also make a claim that certain factors are not relevant for the phenomenon under study. Consider for example Galileo’s law about falling bodies in a vacuum. If one were to test such an assumption reproducing perfectly Galileo’s assumptions then one would just be reproducing the model. However, in order to make sure that Galileo’s law works also in the target domain, then we have to remove that assumption, that is, to include some features that are assumed to be negligible, and see whether it still holds. In this case we would be able to say for example that it does not work for feathers but it works approximately for heavier bodies. The same applies to economic theories, if a model makes an assumption that certain factors are negligible we need to verify this in the target domain, or a domain that reproduces some of its relevant features to check whether that assumption has actually no relevant effect on the mechanism under study.

1.3 The fictionalist approach

The third and final methodological interpretation of models that will be discussed is the one that sees some of the assumptions as principles that are employed in a wide variety of models because of their reliability. The fictionalist approach sees models as counterfactual worlds that are connected to the real world through a relationship of similarity. Their assumptions therefore are neither true nor false. Two accounts will be discussed, those of Giere and Sugden.

Giere’s (1988) account of models is embedded in a systematic view of science that he sees as a cognitive social process of production of knowledge. I will however only discuss the features that are relevant for the purposes of this chapter, that is, Giere’s account of models in the natural sciences, hypotheses of similarity and his notion of theories.

According to Giere models are “sets of objects” or “idealized systems”. A clarification is needed here. The realist also refers to idealized systems, but these idealizations are idealizations of reality, in fact models in the realist view represent reality. Giere interprets idealized systems as abstract entities that have no connection with the real world at all. They do not make any empirical claims therefore we cannot ask if they are true or false in an epistemological way. Considering that models are abstract entities that "have no reality beyond that given them by the physicists" (p. 78), the only way in which such models can be and are connected to the real world is through a theoretical hypothesis. According to Giere a theoretical hypothesis is "a statement asserting some sort of relationship between a model and a designated real system" (p. 80). What kind of relationship is the hypothesis claiming? Giere maintains that the relationship is that of similarity.

> "Hypotheses, then claim a similarity between models [abstract entities] and real systems (or class of real systems). [targets systems]" (p. 81).

Claims of similarity make no sense if we don't specify respects and degrees: "anything is similar to anything else in some respects and to some degrees" (p. 81)

Here is an example of such a claim:

> "The position and velocities of the earth and moon in the earth-moon system are very close to those of a two-particle Newtonian model with an inverse square central force. Here the respects are 'position' and 'velocity', while the degree is claimed to be 'very close'". (p. 81)

A theoretical hypothesis is either true or false if the relationship holds or does not in the specified degrees and respects. The specification of the respects is usually explicit in the hypothesis however the degrees are not usually referred to. In this case the acceptable level of approximation (degrees) to which a hypothesis is considered to be true (or false) is determined by the scientific community. In cases in which a theoretical hypothesis tacitly specifies the degrees and respects, its truth or falsity depends on prior social agreement of the community.

According to Giere therefore assumptions should not be tested because they are not representation of any real world mechanism and do not make claims about the real world.

Giere goes on suggesting that a scientific theory consists of a population or family of models and hypotheses that link the models to the real world. Classical mechanics provides, according to Giere, a clear example of what he means by a scientific theory. In standard textbooks one finds a cluster of models that are obtained by combining Newton's second law of motion with specific force functions. In this context, the laws of motion can be thought of as the link between the models. So these laws 'would be embedded in the models' (p. 85). The model as an abstract entity then is connected to real world through a hypothesis of similarity. A family of models so constructed and the hypotheses of similarity are then defined by Giere as a theory.

At a close inspection, the scientific enterprise in classical mechanics, which is Giere's focus, shares many features with the scientific enterprise in neoclassical microeconomics. On this view the role of rationality and utility (or profit) maximisation within neoclassical economics can be thought of as similar to that of principles such as the second law of motion within classical mechanics. These principles are used in a wide variety of models and, consistently with what Friedman says, give confidence to the theorist that a model is able to tell something about the real world because the principles have been used successfully previously. I would also add that the principle, which defines the models as being part of the family, gives confidence in the model just because of its belonging to that family of successful models.

To sum up, Giere sees models as abstract entities that make no empirical claims. Models are connected to the real world through a similarity relationship. The structure of the models however does not allow defining the two important ingredients of this similarity, respects and degrees. It is the hypothesis of similarity that defines the respects, while the degrees, most of the time, are decided by the scientific community. In this interpretation then it does not make sense to ask whether a model is true or false. What we should ask is whether the hypothesis is true or false.

While Giere's account of models is tailored on physics and more specifically on classical mechanics, Sugden (2000) develops a similar account to explain the role of models in economics.

Sugden interprets models as counterfactual, constructed worlds. Nonetheless, the aim of those models is to tell us something about the real world. But if models are just parallel worlds, how do they connect to the real world?

"Somehow, a transition has to be made from a *particular hypothesis*, which has been shown to be true in the model world, to a *general hypothesis*, which we can expect to be true in the real world too". (p.19)

This transition, according to Sugden, is made by inductive inference. Sugden asserts that many economic models, especially those that use unrealistic assumptions and make vague empirical claims, seem to suggest that: 1) in the model world a certain phenomenon R is caused by F, and since F and R are at work in the real world, "*there is reason to believe"* that R is caused by F. Or alternatively, 2) in the model world, a certain phenomenon R is caused by F, and since F is at work in the real world, "*there is reason to believe*" that R is at work as well. Or, finally, 3) in the model world, R is caused by F, and since R is at work in the real world, " *there is reason to believe*" that F is at work as well. (pp. 19-20)

All these three possible interpretations of models need inductive inference to move from the model world to the real world. But how, Sugden asks, can induction be justified? How does it come to be the case that *"there is reason to believe"* that those models have some bearing on to the real world?

In order for the transition from the model world to the real world to be justified we have to be convinced that there are some similarities between the regularity described in the model and the regularity observed in the real world.

"We gain confidence in such inductive inferences, I suggest, by being able to see the relevant models as instances of some category, some of whose instances actually exist in the real world. Thus, we see Schelling's checkerboard cities *as possible cities,* alongside real cities like New York and Philadelphia. We see Akerlof's used-car market as a possible market, alongside real markets. ...We recognize the significance of the similarity between model cities and real cities, or between model markets and real markets, by accepting that the model *could be real* – that it describes a state of affairs that is *credible*…" (p.25)

Credibility of the model acquires great importance on Sugden's view, but what exactly does it mean for a model to be credible? Sugden discusses the following characteristics of models:

1) "A theoretical model has to be logically consistent.
2) The assumptions must "fit" naturally together.
3) The assumptions must cohere with what is known about real processes in the real world". (p. 26)
4) A model should be robust.

The first condition requires the modeller not to have made any mistake in deriving the conclusions. The second condition requires that the assumptions are not chosen arbitrarily, *ad hoc*, but should give a consistent, and therefore credible, picture of "how the world *could* be." (p.24). The third condition requires that the assumptions employed in the model are consistent with what is already known. The fourth condition is about the robustness of the model. If the modeller is able to get the same results by changing minor assumptions then this gives confidence that the mechanism reproduced is really at work in the real world.

We have seen in this section two accounts of models that are very similar to each other. According to Giere, the model is a world on its own connected to the real world through a similarity relationship. We are confident in such a similarity if the model belongs to a family of models whose similarity with the real world has already been shown. The membership to a family depends on whether the model employs the same principles as other models that belong to the same family.

Sugden interprets models in economics as counterfactual worlds invented by the economists' minds and connected to the world via inductive inference that may be more justifiable if the model depicts a credible world. If this credible world satisfies the four principles aforementioned then we gain some confidence that the model may actually be describing a mechanism similar to a real one.

These two views see models as either abstract systems (Giere) or counterfactual worlds (Sugden). For both testing assumptions is a pointless exercise. What we should test instead is whether the similarity hypotheses or empirical claims made on behalf of the models hold true in the real world.

1.4 Summary

We have discussed in the previous subsections three different theories that explain what economic models do.

The instrumentalist view sees models as machines that use false assumptions that cannot be tested and that have no bearing on the world. These machines are used to produce predictions that can be tested. Additionally Friedman proposes also some methods that can help us testing those theoretical models. We can get indirect evidence on assumptions if we test the model in a domain that is different from the target domain. The more similar the two phenomena observed are, the stronger the confidence of the scientist in the results is. Similarly, we can have confidence that an assumption works if it has been used successfully in previous models (for example the assumption that economic agents are rational).

The realist sees models as representations of the real world. Assumptions however are not true, in fact they idealise and abstract from the real world. It would not make any sense to test those assumptions, because the modeller makes no claim that they fully represent reality. In order to test those models we therefore have to understand what the assertions of the assumptions are and then test them in a domain that is similar to the one in which the theory is supposed to be applied.

The fictionalist approach interprets models as idealised (not real) systems or counterfactual worlds. We should not test assumptions but the similarity claimed between the model or parts of it and the domain of application, or domains enough similar to it, is true or false. So, what we test is the hypothesis of similarity (Giere) or the empirical claim (Sugden) that a theoretical model makes.

2. Two model: Varian's model of sales and Bikhchandani et al's model of fads.

In this section I will present two examples that fit the sub-class of models that are the object of this deep philosophical dispute. The two models are: Varian's model of sales and Bikhchandani et al.'s model of fads.

2.1 “A model of sales” by Hal R. Varian (1980)

Varian’s model of sales is an interesting case of study. As well as representing a good example of a standard modelling strategy in economics, it is one of the rare cases where the author has also been interested in methodological issues concerning the role of models in economics. Varian has written with Gibbard the well-known paper Economic Models (1978) addressed to an audience of experts in the field and a less formal paper (1997) addressed to an audience of young economists on how to build an economic model. Therefore, the analysis of “A model of sales” can be quite insightful because of this inside perspective provided by Varian himself.

The model Varian proposes belongs to a class of models that deal with the price dispersion commonly observed in retail markets. The main feature of this model is that it explains price dispersion over time rather than over space or product characteristics. The target domain is vaguely defined and so is its connection with the model. The target domain can be identified as firms’ pricing decisions; specifically, temporary reductions (‘sales’) by retailers.

Here there is a brief description of the model. There a large number of consumers that have a reservation price r and want to buy at most one unit of a good. They come into two types: informed or uninformed. There are n stores that maximise profits. Every period they choose a pricing strategy f(p) from which they randomly select a price. Informed consumers buy at the store that charges the lowest price while uninformed consumers shop randomly. If a store charges the lowest price among all stores it gets all the informed consumers and its share of uninformed ones. If a store does not charge the lowest price gets only its share of uninformed consumers. The model’s main results are that there is no pure-strategy equilibrium for firms but a symmetric mixed strategy Nash equilibrium (MSNE) in which: each firm randomises its price; as the proportion of informed consumers (λ) increases, the expected price paid by both types decreases; as the number of firms (n) increases, the price paid by informed consumers decreases while the expected price paid by uninformed ones increases.

In order to understand what Varian is actually telling us about the real world, it is useful to have a close look at his paper. The introductory section begins as follows:

> “Economists have belatedly come to recognize that the “law of one price” is no law at all. Most retail markets are instead characterized by a rather large degree of price dispersion. The challenge to economic theory is to describe how such price dispersion can persist in markets where at least some consumers behave in a rational manner.” (p.651)

From this quote it is clear that Varian is going to develop a model that deals with the real phenomenon of price dispersion observed in most retail markets. This phenomenon is of recognised importance because standard economic theory, that predicts that the same good sells for the same price in every market, cannot account for it. It is not clear yet whether his model describes or explains the real mechanism (if there is one) of price of dispersion or simply builds a mechanism that reproduces this phenomenon. It seems clear though, from the way the quote is phrased, that the challenge he takes on is to build a model that incorporates as few changes as possible to the standard approach.

Varian goes on noticing that other models have dealt with the same phenomenon. One prominent example he discusses is the paper “Bargains and rip offs: A model of monopolistically competitive price dispersion” by Salop and Stiglitz (1977). In Salop and Stiglitz’s model there are informed and uninformed consumers. The informed consumers know the entire distribution of prices, while the uninformed consumers know nothing about the prices. Some stores sell at the lowest possible price, the rest of the stores sell at a higher price. The informed consumers buy at stores charging the lowest price, the uninformed consumers shop randomly. Varian stresses that Salop and Stiglitz’s model has a feature that makes the mechanism they propose quite implausible:

> “In the Salop and Stiglitz model – as in all the models of spatial price dispersion – some stores are supposed to persistently sell their product at a lower price than other stores. If consumers can learn from experience, this persistence of price dispersion seems rather implausible.” (p.651)

Part of the reason why Varian proposes a model that deals with temporal price dispersion is that spatial price dispersion models neglect completely the possibility that consumers can learn from experience. Here Varian is talking about real consumers, suggesting that a model, in order to be plausible, should at least have some credible features that resemble the real world for it to respond to the challenge reported in the first quote. So this last quote should be read in this way: it is not plausible to have consumers in a model that do not learn from experience because in the real world people surely do so.

Varian goes on discussing the merits of temporal price dispersion. This kind of price dispersion is observed in a theoretical market (i.e. his model) where stores intentionally vary prices over time. Consumers cannot learn from experience because of this intentionality and therefore the market exhibits persistent price dispersion. As Varian points out:

"One does not have to look far to find the real world analog of such behaviour. It is common to observe retail markets where stores deliberately change their prices over time – that is, where stores have sales. A casual glance at the daily news paper indicates that such behaviour is very common. A high percentage of advertising seems to be directed at informing people of limited duration sales of food, clothing, and appliances." (p.651)

This passage is interesting because it reveals the model's target domain. Varian provides casual evidence that a model like his, that adopts temporal price dispersion, is not out of line with what is commonly observed. However the evidence he presents is not a formal test of his model, but it is what he calls in "Economic models" (1978) casual empiricism. The meaning of this passage is made clearer in the following quote:

"In this paper, I explicitly address the question of sales equilibria. … firms engage in sales behaviour in an attempt to price discriminate between informed and uninformed costumers. *This is only one aspect of real world sales behaviour*". (italics added) - (p.652)

It is clear that Varian is proposing a model with the purpose of describing, maybe explaining or simply predicting, a real world phenomenon. He is going to provide a rationale of price dispersion by means of sales. It is not clear however, if the mechanism he proposes describes a real mechanism or only a plausible one which he does not assert to be real but thinks could at least be possible. The fact that he says that he will address the question of sales equilibria could be a sign that the third interpretation is the most likely one.

In the concluding section Varian refers one more time to the phenomenon his model deals with:

> "The form of the resulting pricing strategy … does not seem out of line with commonly observed retailing behaviour. … Although this casual empiricism can hardly be conclusive, it suggests that the features of the model described here *may have some relevance in explaining real world behaviour*." (italics added) – (p.658)

As the quote shows, Varian is rather cautious and vague about the relevance of his model to the phenomenon of sales. Varian's model generates a pricing strategy that is similar to what stores do in the real world. The evidence provided is, as he says, casual empiricism and it is not proposed as a test of his model. However, this casual empiricism provides some support to his model and this leaves some hope that it may explain some aspects of the phenomenon under study.

The features of this model fit quite well with the features of those that have been often criticised in economics. The model is rather abstract and the assumptions are quite unrealistic. It is a self-contained world where some of the attributes are chosen because it is normal practice in economics to include them in their models (e.g. consumers maximise utility and firms maximise profits). The target domain is loosely defined. No testable hypotheses are provided and no empirical claims about the relevance of the model results are made. So, even if the model is meant to be saying something about the real world, the connection with it seems quite loose and vague.

2.2 "A model of fads, fashion, custom and cultural change as informational cascades" by S. Bikhchandani, D.Hirshleifer and I. Welch (BHW)

BHW propose a model that shows how a sequence of rational individuals imitate the choices of those ahead of them, without considering their own contrasting private information, starting what they call an informational cascade.

Here is a brief description of the model. There is a sequence of individuals who decide whether to adopt or reject some behaviour. The sequence is exogenously given and common knowledge. The cost of adopting is ½ and the gain is V. V is the same for all individuals. V can be either 1 or 0 with equal probability of 0.5. Each individual, before deciding, observes both the action of all the individuals ahead of him in the sequence and a private signal about the value of the action. The signal, identically and independently distributed, can be either H or L. H is observed with probability $p>1/2$ if the true value is 1 and $p<1/2$ if the true value is 0. The expected value of the gain V, conditional on the signal, is the posterior probability that the true value is one. Suppose that the first individual in the sequence observes H (the reasoning is the same in case L is observed). In this case, he/she adopts. Then, the second individual adopts if the signal is H and is indifferent if the signal is L, in this case he adopts with probability of ½. If he/she adopts, the third individual adopts independently of the signal. If the second individual rejects, the third individual adopts if H is observed or is indifferent if L is observed. With this decision rule, BHW derive the probability ex-ante, after two individuals have decided, of an up cascade (that is a correct cascade), a down cascade (a wrong cascade) and no cascade. They find that, the more precise is the signal, that is, the closer to 1 is p, the more likely an up cascade will occur and the sooner the cascade forms. The closer p is to ½ the more uninformative is the signal and therefore the later the cascade forms and the higher the probability of ending up in the wrong cascade. The later in the sequence the individual is, the more likely a cascade has already started. Once a cascade has started, signals become uninformative, therefore an individual behaves rationally, that is maximises his/her expected value, by imitating others' actions and ignoring his/her private signal. BHW generalise the results of this simple model by assuming that C (the cost of an action) and V (the value of adoption) are not constant. An individual before deciding observes a signal's value $x_i= x_1, x_2,\ldots,x_r$ with $x_1< x_2<\ldots< x_r$, and the actions of all the individuals ahead of him/her in the sequence. They show that a cascade will start sooner or later and, as for the simple model, when it has started it will last forever, independently of whether it is right (up) or wrong (down). In fact, when a cascade has started, individuals' decisions do not convey any information about the signal's realisations. BHW also show that a wrong cascade may easily break up, that is it is *fragile*, if the set of private signals differs from individual to individual and if public information is released.

As for Varian's model, it is necessary to have a close look at the structure of BHW's paper in order to understand what strategy they follow.

In the introductory section of their paper BHW notice that human behaviour is characterised by localised conformity that is spread throughout the world. Previous theories have tried to account for this conformity in several ways. However, BHW stress that all these theories fail to explain why conformity is fragile and subject to sudden changes brought about by apparently small shocks. They provide real life examples to support their claim that fragility is a feature of conformity. So, for example, in the 1950s couples who were unmarried were not well-thought-of while in the 1980s they were hardly noticed. Students in the late 1960s protested vividly and were more concerned about political issues than in the 1980s. Their theory, as they explicitly state, is concerned with the fragility and idiosyncrasy of human behaviour:

> "This paper offers an explanation not only why people conform but also of why convergence of behaviour can be idiosyncratic and fragile. In our model, individuals rapidly converge on one action on the basis of some but very little information." (pp. 993,994)

This quote defines the target domain of the model, which is not only conformity but also the fragility and idiosyncrasy of the phenomenon. The authors of this paper promise to give us an explanation for the wide-spread phenomenon of conformity which we can see everywhere around us.

As for Varian's model, rationality is a key element:

"Although the outcome may or may not be socially desirable, a reasoning process that takes into account the decisions of others is entirely rational even if individuals place no value on conformity for its own sake". (p. 995)

This quote is interesting because, as Varian does, BHW stress the importance of rationality in their model as if rationality makes their model more credible and plausible.

The body of the paper is devoted to the description of the model. BHW first propose a specific model (the model I described earlier) to introduce the reader to a basic mechanism of informational cascade. They then generalise the model in different ways.

After they have finished exploring all the theoretical implications of their model, BHW devote a large section to examples, casual empiricism, taken from different areas such as politics, zoology, medical practices and so on. These examples give informal support to the main results of their model and are selected according to several criteria that BHW explicitly list. Some criteria concern the model assumptions: actions are sequential, individuals do not decide according to verbal communication but to observation of others' actions, sanctions and externalities are absent and individuals decide according to others' actions and to their private information. Some other criteria are chosen according to some of the model implications: the phenomenon is local, fragile and some individuals ignore their private information. Here is an example, concerning decisions made by physicians.

"Taylor (1979) and Robin (1984) discuss numerous surgical fads and epidemics of treatment-caused illness ('iatroepidemics'). Some operations that have come and gone in popularity are tonsillectomy, elective hysterectomy, internal mammary ligation, and ileal bypass. They argue that the initial adoption of these practices was frequently based on weak positive information. Robin also points out that most doctors are not well informed about the cutting edge of research; this suggests that when in doubt, they may imitate….We now compare the adoption of one surgical procedure, tonsillectomy, with the predictions of the cascades model. As Robin says (1984, p.75) ‘For many decades tonsillectomy was performed in millions of children….in most cases, the operation was unnecessary’….The adoption of tonsillectomy was not associated with any definitive public information….A critical English panel…claimed that tonsillectomy was being ‘performed as a routine prophylactic ritual for no particular reason and with no particular results’. The rate of tonsillectomy has declined in recent years….There has also been significant idiosyncratic geographical variation in the frequency of tonsillectomy…” (pp. 1011, 1012)

BHW offer these examples as further evidence supporting their model. BHW show that in their model a cascade forms when people start imitating others’ actions ignoring their own poor private information. In the example reported, the fact that tonsillectomy was adopted on the base of weak information, was not associated by any definitive release of public information, and in most cases was unnecessary, may be evidence of a wrong cascade. In addition to that the fact that in recent years tonsillectomy is hardly used may be evidence that cascades are fragile. And the fact that it was used more in certain regions that in others may be evidence that cascades are idiosyncratic.

In the concluding section BHW point out the ability of informational cascades in explaining conformity and sudden change in behaviour. An important result of their model is that cascades are fragile, as pointed out by the previous example.

To summarise, BHW build a mechanism where rational individuals who maximise the expected value of adopting some decision, end up imitating others’ behaviour and ignoring their own information. The main purpose of the model is to explain something that previous models could not account for, using the standard assumption of rationality. They do this by using rather abstract and unrealistic assumptions that they do not assert to be true for some portion of the real world. However, they want to show us that the results they get in this way are similar to the ones we observe in the target domain (e.g. medical practice). The connection they draw between two worlds, the real and the counterfactual, that they want us to believe are alike, is not based on a one-to-one relationship between their elements. They only consider a few assumptions that should hold in the examples, but we are not given any formal proof that they actually hold. They also are quite precise about the predictions they consider but the examples they offer are interpreted as confirming these predictions only informally.

3. Which methodological approach best describes what is an economic model is?

The two models described so far have many common features. In the first place both of them have a mathematical structure that employs unrealistic assumptions that enable the authors to produce some results. The authors employ some of the assumptions because this is standard practice in economics: for example, Varian employs MSNE and Bikhchandani et al. employ Bayes' rule. The authors however are silent about whether these assumptions are true of the real world.

The authors of each model say very little about whether the mechanism they describe is a mechanism that is at work in the real world. This is perhaps because they employ assumptions that are quite conventional in economics and therefore do not feel the need to say anything more about their mechanisms. This is in some sense justifiable from a methodological perspective. Friedman suggests that we can get indirect evidence on an assumption if it has worked well in the past when employed in many other models, like for example assuming that agents are self-interested. If this is the case then we are more confident in the hypothesis and its implications and the authors are somewhat relieved from the task of explaining anything further about the mechanisms they propose. The same applies to the other two methodological approaches. The realist view relies on principles that if true make the model true. Regarding the fictionalist view, Giere seems to suggest that if a model belongs to a family of models it can in some sense inherit similarity from other members of that family, and families of models are identified by their use of common principles such as Newton's second law. Sugden talks about robustness. If a principle has been used in the past then it makes the model more robust and therefore makes it more credible.

The authors of both paper use stories that are not about the real world in order for the reader to understand what their models are about.

Finally, the target domain, that is the domain to which the model should be applied, is loosely defined. This makes those models vaguely related to the real world. Although the authors provide casual evidence in support of their results, the empirical claims derived from them are as vague as the target domain. It seems clear that the authors timidly claim that in some sense and somewhere those empirical claims refer to the real world.

What can be said about these models and the philosophical accounts that describe what they do?

The instrumentalist approach considers assumptions to be false in the target domain. The realist has a completely opposite perspective, i.e. assumptions are true albeit they are idealized and abstract instances of the real world. The fictionalist considers assumptions as neither true nor false, but as properties of a model world that is fictional. Which of these approaches best describe what Varian and Bikhchandani et al. do? Well, it is not easy to answer and this is so because the authors say very little about the assumptions. The same can be said about the mechanisms described in those models. The authors do not give the reader any clue on how we should interpret them. False, real or fiction? So again, we do not have enough information to say which philosophical approach describes best how and what these models tell us about the real world.

How then should we answer the question about which is the "correct" philosophical theory for these models? I think there is not a unique answer to this question. Each of the three methodological approaches can give a good account of what an economic model is. The reason is that the authors say so little about the relationship between their models and the real world that there is scope for the philosopher to fill this interpretative gap in different ways.

Maybe after all, for practicing economists it is not so important to understand which of these accounts describes best what a model is. Economists, after all, seem to communicate between themselves through their models quite well without the need to understand which philosopher is right and which is wrong. Most of the economists have learnt economics at school and when they decide to become researchers they usually learn economics in the field. More experienced researchers transfer their knowledge to the new ones. This applies to the entire community of economist. This shared knowledge among economists, which is the result of many years of economists' work, makes the modeller perfectly aware of what she is communicating with her model and also that the other economists are able to understand what she wants to communicate. Economists therefore do not usually need to be, and in fact they are not concerned with methodological issues Methodological analysis though becomes important when some practitioners start using less standard tools and have the need to convince the scientific community about the validity of the new approach, or to settle disputes on different approaches in use.

PART 2: EXPERIMENTS AND MODELS IN ECONOMICS

1. Experiments which implement models

So far I have analysed a subclass of economic models that employs highly unrealistic assumptions that are used to obtain vague empirical claims. The characteristic features of these models can be summarised as follows:

1) They describe a self-contained world, created by the modeller but resembling some aspect of the real world (the 'target domain');
2) The mechanisms they describe induce some 'result' in the model world;
3) No definite hypotheses are offered about the relationship between the model and the real world;
4) The modeller refers to features of the target domain (casual empiricism) that resemble the 'results' and suggests that in some (unspecified) way the model explains these.

Several methodological theories have been proposed with the purpose of understanding and explaining what an economic model is, how it connects to the real world and how it helps understand real phenomena. Although from a methodological standpoint the dispute is far from being settled, my concern here is in fact with a methodological strategy commonly used in experimental economics: implementing and testing those models in the lab.

I will describe this strategy using two examples: Morgan et al.'s test of Varian's model and Anderson and Holt's test of Bikhchandani et al.'s model. The reasons why I have chosen these examples are that they use an approach which is common to many economic experiments; they exhibit good practice as judged by the profession and the authors are clear about what they are doing. I will start by examining whether the methodological interpretations of models have some suggestions so as to test them Then I will present the experimental test of each model. I will then discuss that test and ask whether it is informative and if so, of what it is informative about (the target domain, the model or something else).

2. Methodological theories and consequences for experimental testing of modelsIntroduction

We have seen in the previous part that the philosophical accounts that were discussed can be a good description of what models in economics do.

In this section we are going to consider the suggestions that those accounts give for testing models in the lab.

All three accounts agree about testing in two respects:

(I) Tests of models as making claims about target domain need to include relevant features of the target domain.

(II) All three philosophical accounts refer to some generic component that many models employ. So for example, for the realist that component would be a principle like Newton's second law. These generic components can be tested in a relatively abstract way; for this, the lab has to be similar to the generic target domain to which models using the generic components are applied.

Let us discuss these two common features in order.

(I) All three philosophical accounts suggest that we should test the empirical claims the model makes about the target.

The instrumentalist position refers definitely to the predictions that a model makes. Since assumptions are false, they are not claims about the target domain. The predictions are the part of the model that has to be confronted to reality. How then should we test such a model in the lab? Well, in the section where Friedman discusses indirect test of assumptions, he maintains that we can get indirect evidence in a domain of application that is not the one to which the model should be applied. The strength of the evidence depends on the similarity between the domain of application and the surrogate target domain. This means that we can test the predictions of a model in the lab, even though the lab is not the intended domain of the model. However we should be aware that the lab is only useful if the implementation of the model includes some relevant similarity to the real world, that is to target domain, which is the ultimate goal of the experiment.

The realist account sees models as approximately true of the real world. The assumptions are not literally true, they are isolations and abstractions of reality, but the mechanism the model reproduces is real. This is not to say that the mechanism is real in the sense of exactly reproducing something in the world. The mechanism is isolated and shielded from the influences of other factors that might or might not have an impact on the phenomenon. What we should test according to the realist is therefore the assertions that the model makes about the target domain. So for example, if the model makes an assertion that a certain factor should not have a relevant effect on the predictions of the model, then we have to test that that factor is actually irrelevant. How should we do that in the lab? Well, we should include in the lab the factor that is known to be at work in the real world but not considered in the model. So, also as for the realist view, the lab should implement some relevant features about the target domain.

The fictionalist view sees models as idealised systems (in Giere's sense, not in the realist sense) or counterfactual worlds that do not represent anything in the real world. Giere and Sugden maintain that a model claims a sort of similarity to the target domain. For Giere it is the hypothesis of similarity that should be tested; for Sugden it is the empirical claims that the modeller is making about the target domain. So, also for the fictionalist the lab should have some relevant (what is implied in the hypothesis of similarity or what is claimed by the modeller) similarities to the real world.

We have seen that, although there are profound differences between these three philosophical accounts, they all suggest that in order to test a theory we should implement some feature of the real world, that is relevant to the model, in order for the experimenter to be sufficiently confident that the results of the test may have some bearing outside the lab.

(II) The second aspect that all three philosophical accounts have in common is that they refer to some general principle or component that can be tested in an abstract environment.

Let us start with the instrumentalist view. In the section where Friedman examines how to get indirect evidence on assumptions, he maintains that one way is if a model employs an assumption that has already been used in other models that made successfully predictions. One example is the self-interest or rationality assumption used in economics. The rationality assumption is used in many models for a wide class of phenomena. In this case, if there are doubts related to that generic component, it seems that we could reasonably test the prediction of the assumption in a lab environment that is similar to the general class of target domains of models in which the assumption is used. So if all models in economics assume rationality, the target domain could in principle be any economic phenomenon, including the base domain of the model. Bardsley et al. (2010) argue as well that a theory can be tested in its "base domain", defined in the following way:

> "The close correlates of the formal objects of a theory and other model entities indicate the *base domain* for it, i.e. the set of possible real phenomena to which application of the theory seems reasonably unambiguous, if made. … Typically, the base domain constitutes only a small subset of the set of phenomena to which the theory could conceivably be applied. For example, expected utility theory attributes to agents preferences over prospects. These formal objects are probability distributions over some well-defined set of possible outcomes. The theory is actually applied to choices between many items, such as investment portfolios or careers, which are not identical to prospects because, for example, some outcomes or probabilities are unknown. However, monetary gambles implemented by randomly drawing balls from a bingo cage (where the number of balls and the prize associated with each ball are known) are close correlates of prospects….so choices among them can be reasonably taken in the base domain of expected utility theory" (p.57,58)

Bardsley et al. go on asserting that testing a theory in its base domain is useful because it provides unambiguous predictions and tests. They maintain that:

> "Any laboratory environment E in the base domain of a theory should be presumed to be in the T-domain [target domain]…." (p.66)

The realist view, in particular Cartwright's version of realism, maintains that, if a true principle (true approximately) is used in a model, then one can reasonably expect the model to work. So for example, if the Galilean gravitational law is used in a model, one should expect the predictions that the model derives from that law be successful if no other disturbing factor is present. Therefore, testing it in the base domain, where there are no disturbing factors it is a legitimate test of the law. Then, after one has checked that it works in the base domain, the test should start including factors that are assumed to be negligible.

The fictionalist view includes a similar idea. Giere suggests that a family of models is defined according to similarity between the models. In the case of models in mechanics, the second law of Newton is the common principle that defines a family of models. Sugden talks about robustness of models. Robustness in this case should be interpreted in two different ways: a) Robustness within the model, that is, a model is robust if changing some irrelevant assumptions does not affect the results; b) Robustness related to a family of models, similar to Friedman's view. In this case, using some principle that is common to many existing models, for example rationality or mixed strategy Nash equilibrium, can give a new model more credibility in Sugden's sense. Where should we test the generic component? If the component is completely generic, that is, if it is used in every economic model, then we could test it in any economic environment, even one that is extremely abstract. In principle we should test the model that employs the principle in an environment that is similar to the general class of target domains to which the model is applied, but since all models are linked by a similarity hypothesis, the base domain belongs to this class.

3. Experimental tests that implement models – two examples

In this section I will present two case studies of experiments that have the common feature of implementing all the assumptions of the models they purport to test except one, that is, they use real subjects instead of theoretical ones.

3.1 An experimental study of price dispersion – Morgan, Orzen, Sefton

This work is an experimental test of a model derived from Varian's model of sales.

The strategy Morgan et al. use in their paper is similar to the one used in many models. They provide some casual evidence, see quote below, showing that those models that deal with the same phenomenon are relevant to the real world, and therefore the experiment too.

In the introduction Morgan et al. briefly discuss some "clearinghouse models" that predict price dispersion. These models include Salop and Stiglitz (1977), obviously Varian (1980) and Rosenthal (1980).

"...the Internet, in the form of price comparison sites, offers a concrete example of such a clearinghouse". (p. 135)

This quote shows that Morgan et al. are treating price dispersion as a real-world phenomenon. So they seem to imply that tests of clearinghouse models have real world relevance. They therefore are going to test one of these models: a modified version of Varian's model of sales. They manipulate the model in order to get predictions. The solution concept they use to derive those predictions is the mixed strategy Nash equilibrium (MSNE), used commonly in many economic models. The predictions they derive are the following:

1) An increase in the number of the uninformed consumers, leads to an increase in the price to both types of consumers;
2) An increase in the number of firms leads to an increase in prices to uninformed consumers and a decrease in prices to more informed. These predictions are derived by analysing the differences in the mixed strategy equilibria due to the change in the market structure.

It is not clear from these statements whether Morgan et al. interpret these predictions as referring to the model or to the target domain. However, as I interpret it, since the experiment is presented as a test of those predictions, and since the model's predictions refer to the real world, it seems that those predictions must refer to the target domain too. The results of the experiment are then to be interpreted as confirming or disconfirming those predictions. And in fact the predictions are partly confirmed.

> "In our experiment we observe substantial and persistent price dispersion. We find that, as predicted, an increase in the fraction of informed consumers leads to more competitive pricing for all consumers. We also find, as predicted, that when more firms enter the market, prices to informed consumers become more competitive while prices to captive consumers become less competitive. Thus, our experiment provides strong support for the model's comparative static predictions about how changes in market structure affect pricing." (p. 134)

This last quote shows that the aim of the experiment is to test Varian's comparative statics. Their results in fact, as they stress, provide strong support for them. This is confirmed in the conclusions where they state:

> "Overall, we find strong support for the ability of clearinghouse models to predict the comparative static effects of changes in market structures" (p.153)

It is not clear from these quotes whether they interpret their results as having relevance to the target domain. The following quote will shed some light on these doubts.

> "More broadly, our results have some general implication for understanding the effects of price competition. The word 'competition' is ubiquitous in discussion of economic and current affairs, but its precise meaning is sometimes unclear. Often a competitive market is taken to be one with many firms; indeed various concentrations indices are used as measures of competitiveness. Our results show that increased competition in this sense does not necessarily result in lower prices. An alternative view of competition is based on consumers' ability to substitute away from high-priced firms. Our results suggest that increased competition in this sense does lead to lower prices." (p. 154)

As I interpret this quote, it seems that Morgan et al. are in some way implying that since the comparative static implications of the model are confirmed in the lab they will also hold in the target domain.

Morgan et al. in the second part of the paper, where they justify and explain the purpose of the experiment, refer to some doubts they have about the ability of subjects to follow MSNE. These are the doubts:

1. "First, in a mixed strategy equilibrium, there is no positive reason for a rational player to conform to the equilibrium strategy since she will receive the same expected payoff from any pure strategy within the support of the equilibrium distribution.
2. Second, the equilibrium price distribution is difficult to compute, and so it seems unlikely that *subjects* will reason their way to an equilibrium.
3. Third, for the parameters we employ in our experiment, the equilibrium is unstable under the class of positive definite adjustment dynamics, and so it is unclear whether *subjects* could reach the equilibrium through some learning process." (p.139)

It is worth noting that the doubts that Morgan et al. have about the MSNE refer to the subjects in the lab and not to retail firms, which is the target domain of Varian's model. However, having found that the model's predictions work in the lab they are more confident that the model works in the target domain and this despite the fact that their doubts refer to subjects' behaviour and in particular to MSNE predictions, which is a generic component of many economic models.

To sum up, Morgan et al. propose a modified version of Varian's model. They derive some predictions regarding some changes in the market structures that would be testable in the lab. They test those predictions, show that they are confirmed by the results, and then draw some conclusions about their relevance for the target domain.

The third section of the paper is devoted to the description of the experimental design that consists of an almost complete implementation of the model. The main difference is that in the lab real subjects are substituted for the theoretical ones. Here's a brief description of the design. The experiment involved a two-seller treatment and a four-seller treatment. Sellers in both treatments faced six computerized buyers. The experiment consisted of three phases of ten periods each. In each period sellers had to post prices and the quantities they were willing to sell at those prices. In phase 1 and 3, three buyers were informed of all prices posted and bought 12 units each from the sellers that posted the lowest prices, the remaining three were not informed and bought from each seller 6 units, in the treatment with two sellers, and 3 in the treatment with four sellers.

Having presenting the experimental design, Morgan et al. discuss some of its features. They decided to have a posted offer market with 30 periods in order to give the subjects time to learn which strategy was the most profitable. They also stress that having computerized buyers would allow them to screen out some effects that are not taken into account in the model, such as inequity aversion effects, thus simplifying sellers' decisions. By eliminating uncertainty related to the behaviour of real buyers, Morgan et al. are able to focus on the strategic interaction among sellers.

According to Morgan et al. their design had a drawback, that of having 30 periods. In such an environment, sellers would have the chance to interact repeatedly and this might have led to collusion. Therefore to be consistent to the model structure, they decided to assign subjects randomly to different groups every period.

After having presented the design and discussed its drawbacks, Morgan et al. discuss the hypothesis and present the results. The results are then compared to the theoretical predictions of the models.

They conclude by noticing that they find support for the main predictions of "clearinghouse" models and more generally they contribute to the understanding of price competition.

To sum up, the strategy Morgan et al. seem to be using is the following. The model says something about the world, in the lab the predictions are confirmed and therefore we are more confident that the model works in the real world.

3.2 “Information cascades in the laboratory” by Anderson and Holt

The experiment that will be discussed in this section is a test of BHW’s model presented previously.

Anderson and Holt start their paper by introducing informational cascades, which as they say, are situations where an individual ignores his/her own private information and adopts the behaviour of other subjects that signals contrasting information. As BHW do, they offer some examples of cascades:

> “…suppose that a worker is not hired by several potential employers because of poor interview performances. Knowing this, an employer approached subsequently may not hire the worker even if the employer’s own assessment is favourable, since this information may be dominated by the unfavourable signals inferred from previous rejections.” (p. 847)

Anderson and Holt stress that, as BHW do, informational cascades can result from rational behaviour. This in fact is the focus of their experiment.

They point out that the way cascades are represented in the model may not be the way they actually form in the real world for several reasons:

1. “Human subjects frequently deviate from rational Bayesian inferences in controlled experiments, especially when simple rule-of-thumb heuristics are available.” (p. 848)
2. “With sequential announcements, decision makers must make inferences about others’ rationality.” (p. 848)
3. “Much of the evidence offered in support of the rational view of cascades consists of anecdotes about patterns in fashion, papers getting rejected by a sequence of journals, the risk of entering the market too early, etc.” (p. 848)

For these reasons, according to Anderson and Holt:

> “Laboratory experiments can provide more decisive evidence on the validity of the rational view of cascades”. (p. 848)

It can be noticed that, as for Morgan et al., the purpose of the experiment is related to some doubts about the validity of a theoretical component of the model. Here the doubts are that real people may actually deviate from Bayes' rule for the reasons Anderson and Holt et al. have discussed. They therefore suggest that this may raise some doubts about whether cascades may form rationally as described in the model. Whether they refer to real cascades or just cascades in the model is not explicitly stated; however, since they refer to doubts as to whether real people can use Bayesian rationality, it seems to me that they refer to real cascades.

Having presented what the aim of their test is, Anderson and Holt go on providing some evidence, drawn mainly from psychological studies, showing that actually people do not behave according to Bayes' rule.

Subjects sometimes use other decision making rules, such as status quo bias, or preferences for conformity. Regarding the first, people might favour the status quo either because they think that the status quo was established on the basis of good information or as a rule-of-thumb. Both cases lead to conformism. However it is difficult in real world decisions to test whether the choice follows Bayesian rationality or the status quo rule of thumb because it is not possible to control for the amount of information people have. If subjects have preferences for conformity, they derive utility just from doing what others do. Nevertheless, in the real world it is difficult to know whether people conform to others' behaviour as a rational response to the information they have or just because they have preferences for conformity. In the lab Anderson and Holt are able to discriminate: (a) between status quo bias and rationality, by controlling for the amount of information subjects have; and (b) between conformism and rationality by using anonymity. In the real world we are not able to do that, and this is a clear advantage of the experiment.

> "It is possible to control information flows in the laboratory by drawing balls from urns and, therefore, to determine whether subjects tend to follow previous decision(s) when it is rational" (p. 848)

The first and second sections of their paper are devoted to the description and implementation of the simple model that they are going to test. There are two events A and B which are equally likely. A signal reveals information about the event, there are 2/3 chances that the signal matches the event and 1/3 that the signal does not. The signal is private but the decision is public. Individuals observe the signal and decide which event is likely to happen, given that signal. The decision process is sequential. The first individual observes the signal and decides which event is more likely. The second individual observes the signal, knows the decisions of the first individual but not the signal, which is private information, then decides which event is more likely according to the information she has. If the first individual's signal is an a, he/she infers that the event is A, and therefore selects it. If the first individual's signal is b, he/she infers that the event is B and therefore chooses it. The second individual only sees which event the first subject has selected, so from it he/she can infer the first subject's signal. Therefore if the second individual's signal matches that of the first one, he/she should decide accordingly, otherwise stick to his/her signal. In this case the probability that the event is A is the same as the probability that the event is B; it is assumed that the person prefers the event that matches his/her signal. The third person in the sequence should ignore her/his private signal if the previous actions match, that is, if the first two individuals have both chosen either A or B starting a cascade. If the first two individuals' actions do not match then the third individual is in the same situation as the first one in the sequence.

The second section of the paper describes the experimental design that is an implementation of the simple model just presented. The main difference between the model and the experiment is that, as for Morgan et al.'s experiment, Anderson and Holt use real subjects. In detail, six subjects participated in each session of 15 periods. At the beginning of each period, a monitor threw a dice to decide which urn was to be used. Both urns had the same probability of being used. In each period, after an urn was selected, each subject, unaware of the urn used, was approached, shown privately a random draw from the urn and asked to decide which urn was the one being used. Once the first subject had decided, the second subject in the sequence was approached and so on. Subjects could only see the previous subjects' decisions but not the draws. Each correct decision was rewarded with $2.

As Anderson and Holt stress in the introduction, individuals often follow heuristics, such as status quo bias or preferences for conformity. The simple model they have tested, allows one to check whether subjects use a status quo bias heuristic vis-à-vis Bayes' rule. However Anderson and Holt are also interested in another heuristic, that is, the counting rule. To explain what is the counting rule I will use the model they implement in order to discriminate between counting rule and Bayes' rule.

The model is similar to the one presented previously except for the composition of the urns. Urn A has 6 balls labelled *a* and 1 ball labelled *b* and urn B has 5 balls labelled *a* and 2 balls labelled *b*. If one observes a *b* signal, since the urns are not symmetric like the ones used in previous part of the experiment, it is more likely that the urn B is the one being used. So counting the decisions made previously is not necessarily rational. Suppose that you observe 2 *a* signals and 1 *b* signal. If you are following a counting rule you will choose urn A, while choosing urn B is the decision that is consistent with Bayes' rule. Since the implications of Bayes' rule and the counting rule do not always coincide in this model, Anderson and Holt are able to discriminate between them.

They found that 1/3 of subjects followed the counting heuristic violating Bayesian rationality.

The final section is devoted to a summary of Anderson and Holt's findings. These are that informational cascades are observed consistently in their experiment. Most of the times cascades form rationally except in some cases where subjects' decisions follow a counting heuristic.

4. What do we learn from experiments that implement models?

In this section I will discuss the methodological strategy adopted by the experimental tests and ask what we can learn from it.

We have seen in the previous sections that the experiments by Morgan et al. and Anderson and Holt use a very similar strategy.

Both experiments justify the purpose of their tests by adducing to some doubts about the empirical validity of a behavioural component of the models. In Morgan et al.'s case the doubt is that MSNE does not work for three main reasons presented before. In Anderson and Holt's case the doubts refer to the ability of subjects to conform to Bayes' rule. These doubts, as I will explain later in more detail, refer to the general validity of those principles rather than to the principles as employed in that specific context, that is the models they test.

Both experiments implement almost completely the model their tests are based on. Obviously, if every feature of the model was implemented, there would be nothing to test (except for the general validity of mathematical theorems). Since all the models' features have been implemented except for the fact that the experiments use real subjects, any divergence between the models' results and the experimental results concern subjects' behaviour. In the case of Morgan et al., this divergence would be between the strategies actually chosen by subjects (given the payoff matrix of the model) and MSNE (the solution concept used in the model). In the case of Anderson and Holt, it would be between the subjects' actual choices and the choices predicted by Bayes' rule.

Neither experiment implements any feature of the target domain that can be identified by the casual evidence presented by the modellers. Morgan et al. use the words 'sellers' and 'profits' in their experiment, but only to make the task more understandable by the subjects. Anderson and Holt do not refer to anything other than urns, balls, predictions and so on. We can therefore conclude that these experimental tests are not tests of hypotheses (derived from or suggested by the models) about the target domain but lab tests of MSNE using a particular matrix and of Bayes' rule using a particular game.

Is it legitimate to test general components widely used in economic modelling in a very specific setting that reproduces a particular model? One way of thinking about this is to suppose counterfactually that the predictions derived from MSNE and Bayes' rule fail in the lab. Does this pose a problem for Varian and Bikhchandani et al.'s models? In general we could say that this poses a problem.

If MSNE fails in a very specific game (the payoff matrix used in the lab) in which it is expected to work, Varian would need to explain why it does not work for subjects but it is expected to work for retail firms. Similarly, if Bayes' rule fails in the specific game derived from Bikhchandani et al.'s model, they would need to explain why they expect that to work for real individuals but not for subjects. In both cases the modellers should refer to some relevant differences between the target domain and the lab domain.

We have seen what problems arise in the case that the predictions derived from the generic component fail in the lab. It is natural to ask what would happen if the predictions on the other hand are confirmed. One obvious answer is that being the lab environment very abstract and tailored on the model, it would be extremely difficult to export the results to the target domain, because of this lack of similarity between the two environments, i.e. the lab environment and the target domain. However this is not the end of the story. One problem of testing the same component by reproducing exactly the model's features, undermines the comparability of tests, being the designs different every time (because different is the model being tested). This difficulty can be tackled if we use the same design (and therefore the same model, if we want to implement a model to test a generic component of it) to understand what were the reasons of its success. But then, if we had to use the same design to test the same generic component, we should be careful in choosing the model. The model should be such that the experimental design is the best possible for the purpose. Then in this case, having in mind the test of a generic component, it may be that the best design is not even based, although it could be, on a model's assumptions.

In general, even if the experiments are not informative about the target domain, they can be informative about the two specific models if there are concerns that the two principles interact with some specific features of the model, thus raising doubts about their validity in this particular context. So we should analyse the model in order to see whether such concerns arise. In Varian's model MSNE is used a solution concept for a n-player normal-form game; however at this level of abstraction, there seems to be no obvious differences between Varian's model and other models that use MSNE. Similarly, Bikhchandani et al.'s model uses Bayes' rule as a general rule for individual behaviour. Many models do the same so it is not clear why this assumption in this model should pose specific problems.

If the two generic components fail in a lab test this, as already pointed out, poses a problem not only for the two models but for the principles themselves. Both Morgan et al. and Anderson and Holt express doubts about the working of MSNE and Bayes' rule respectively not in the specific target domains, that is retail markets and conformity, but as general principles. The reasons why Morgan et al. doubt the MSNE is related to the ability of subjects to compute the equilibrium strategies, because there are no reasons for them to choose that particular equilibrium and so on. These reasons can be adduced to any complex game that uses MSNE as a solution concept. Similarly, Anderson and Holt refer to reasons that suggest that Bayes' rule does not work in contexts that are not specific to informational cascades. They refer to the fact that subjects have to make inferences about other players' rationality and that they may use heuristics as shown by evidence.

Testing MSNE and Bayes' rule as generic components of economic models raises some methodological issues that need discussing. These issues relate to whether it is better to structure those experimental tests around specific models (like Morgan et al. and Anderson and Holt do); or it is better to design those tests around the doubts the have arisen about those particular behavioural assumptions.

Let us consider each experiment in turn.

Morgan et al. justify their experiment by adducing three specific doubts about the empirical validity of MSNE. Then, they structure their test by implementing Varian's model. So what they actually do is to use Varian's model as a framework to test the MSNE component of it. However, Varian's model turns out to be too complicated for the experiment to be a clean test of the three doubts that justified the test in the first place.

Let us make ourselves clear about what I mean by testing those doubts in isolation. For expositional clarity I will repeat here the three doubts that the authors had regarding the validity of the MSN. The first refers to the absence of a positive reason why subjects should follow the equilibrium strategy. The second refers to the ability of subjects to compute difficult equilibrium price distributions. The third refers to subjects' learning ability to reach the equilibrium given the parameters used.

A simple test for the second doubt would be that of changing the degree of computational complexity to check whether that makes any difference to subjects' behaviour. The simpler of the two computational cases would provide also evidence for the first doubt. That is, whether subjects conform to the equilibrium strategy regardless of the fact that they do not have reasons to do that. The third doubt could be tested by varying the parameters of the payoff matrix and having subjects play repeatedly. This would allow one to check whether there is some learning process going on in the expected direction and whether the instability of the equilibrium matters.

Morgan et al. use Varian's model to test those predictions jointly, however if the test fails their design does not allow them to discover which one(s) of the three doubts is(are) causing this failure.

Anderson and Holt experiment is also a test of Bayes' rule. However, as opposed to Morgan et al., theirs is a cleaner test. They use as a framework to test Bayes' rule the very simple model proposed by Bikhchandani et al. Their design allows them to control for the flows of information. In this way they are able to discriminate (at least partially) between cases in which subjects are following heuristics such as status quo bias, preferences for herding, counting rule, and cases in which they are following Bayesian rationality. So in this respect Bikhchandani et al. offer a simple model that can be used in the lab to test a generic component of the model itself.

So, in some cases, models can be used both as explanations of real-world phenomena and as controlled tests of generic theoretical components. From a more general perspective this kind of approach may lead to a proliferation of experiments that test the same theoretical component in different settings (models). As a consequence of that we have a non-systematic study of behavioural assumptions that would not allow for comparability. If the tests' results match the predictions of the principle then there is no problem. The problems arise however, when the results do not confirm the predictions. In this case this strategy may not be very helpful in showing where the theory goes wrong.

What are instead the advantages of organising tests around specific theoretical issues, that is in isolation? In the first place this strategy would allow for a more controlled and systematic test, unlike Morgan et al.'s one. And if the same framework/design is used in many experiments, as for example the prisoner's dilemma, trust and ultimatum games, this would allow for comparability of the results between experiments that use the same framework. It would allows specific issues to be investigated more systematically, leading to wider and consistent knowledge of the subject. In this respect, Anderson et al. use a simple framework that has been used thereafter to study individual behaviour. Here are few examples. Weizsacker (2008) uses data from 13 experiments that use the same framework used by Anderson and Holt to carry out a meta-analysis. His purpose is to test rational expectations. Nöth and Weber (1999) use a modified version of Anderson and Holt design to study how people aggregate private and public information. So although Anderson and Holt experiment is a test of Bayesian rationality built around a model it has also provided a simple framework to study economic behaviour in a systematic way.

5. Experiments that implement real world features

In this section I will refer briefly to approaches that are different from the model-implementing approach just discussed. These approaches will be discussed using two examples: Chamberlin's imperfect experimental market and Schelling's focal points experiments.

5.1 "An experimental imperfect market" – E.H. Chamberlin

Chamberlin tests whether quantities and prices traded are the ones predicted by perfect competition (i.e. supply=demand). In the model of perfect competition it is assumed that the equilibrium in the market is reached as if there was a Walrasian auctioneer. The Walrasian auctioneer is an 'as-if' assumption; like Friedman's assumptions about the leaves, it is false. Real markets do not have a Walrasian auctioneer and therefore this feature of the model is neglected in the experimental design. What Chamberlin is curious about is whether in real markets trade takes place at equilibrium prices and quantities. This is difficult to test in the field, because supply/demand functions are not directly observable (they are normally inferred from behaviour in markets).

Chamberlin begins by referring to real world phenomena that are the results of many forces that operate simultaneously. The social scientist would like to isolate forces in a laboratory setting, however Chamberlin acknowledges that this is precluded to her by the nature of social science. The only way, Chamberlin goes on, she can isolate those forces is theoretically by means of models. However,

"The purpose of this article is to make a very tiny breach in this position: to describe an actual experiment with a 'market' under laboratory conditions to set forth some of the conclusions indicated by it." (p. 95)

This passage shows that the purpose of the experiment is to allow the author to gather some data from a controlled environment and to learn something from the comparison between this and the real world. The experiment is therefore just a tool that lies in between these two extremes: theory on the one hand and real life on the other. (p. 95)

"It [the experiment] was designed to illuminate a particular problem…., viz., that of the effect of deviations from a perfectly and purely competitive equilibrium under conditions (such in real life) in which the actual prices involving such deviations are not subject to 'recontract' (thus perfecting the market), but remain final. (p. 95).

Chamberlin finds the experiment quite instructive:

"…for the many comparisons afforded, both of similarity and of contrast, between the laboratory market and its diverse counterparts in the real economic world."

The design of the experiment is the following. Student participants are either buyers or sellers. They are endowed with one unit of a good and are given induced values for it. The sellers are given the minimum amount they would be willing to accept in exchange for the good and buyers the maximum amount that they would be willing to pay. Induced values are different for every student and define the market demand and supply curves. Students can freely move about in a room, bargaining bilaterally. When contracts are made, the agreed price is written on a blackboard, visible to all subjects. The main characteristic of the design is that it implements both model and target domain features. The model features that have been implemented are that subjects have well-behaved preferences. However the auctioneer is not implemented. To implement this assumption would miss the point of the experiment. Walras' hypothesis is that real market forces work in a way that is similar to the auctioneer as described in the model. He does not claim that a real human being could do what the auctioneer does. The feature similar to the target domain is represented by real people going about and looking for the best bargain.

An interesting feature of the design is that the rules of the trading process are simple and well-defined, although it would be impossible to describe this in a fully-specified mathematical model. This is a strength of the design compared with models. In fact we are able with this experiment to do something that we cannot do with models and simultaneously learn something about the target domain.

So to sum up, some of the features of Chamberlin's experiment represent the model, but some represent the real world directly. The experiment tests the model without implementing it as fully as possible. It is a (flesh-and-blood) model of a real market: it is substituting in part for a mathematical model.

5.2 Schelling's focal points experiments

Schelling starts from a class of real-world problems in which two agents have a common interest in coordinating their behaviour. Players get a positive payoff if both choose the same option. Let us see some examples.

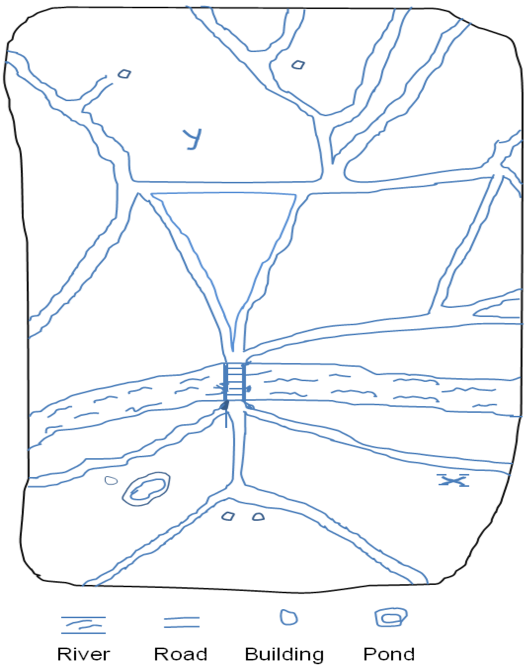


Figure 1: Schelling Map of a Coordination Game

In chapter 3 of "The strategy of conflict" (1960) Schelling analyses several real-world examples. In all cases there is some obstacle to communication, in some others there is a conflict of interest that may render coordination difficult to achieve. However coordination is always the best outcome for both players.

Consider this example based on the map reproduced above (from p. 55 of Schelling)

> ‘Two opposing forces are at the points marked X and Y in a map similar to the one in Fig. 7 [the map above]. The commander of each force wishes to occupy as much of the area as he can and knows the other does too. But each commander wishes to avoid an armed clash and knows the other does too. Each must send forth his troops with orders to take up a designated line and to fight if opposed. Once the troops are dispatched, the outcome depends only on the lines that the two commanders have ordered their troops to occupy. If the lines overlap, the troops will be assumed to meet and fight, to the disadvantage of both sides. If the troops take up positions that leave any appreciable space unoccupied between them, the situation will be assumed ‘unstable’ and a clash inevitable” (p.62)

This is a stylised example of a realistic coordination game where two players have to coordinate. In Schelling’s experiment, more than 50% of the subjects coordinate on the river. It is interesting to notice that people are able to coordinate even if, as in this case, communication is not possible and there is some conflict of interest. Not all the examples presented by Schelling do have conflict of interest. In many of them however the two players have to choose the same options. In some cases for example players have to choose an integer number, in other cases they choose between heads and tails, in others they have to decide the time and the location to meet in some places in New York. The extraordinary thing is that a high percentage of people are able to coordinate, contrary to what game theory predicts. What Schelling wants to show us, using simple examples, is that real subjects are actually able to coordinate on the same option following some principle of salience that conventional game theory treats as irrelevant.

Schelling’s hypothesis is that real agents can solve coordination problems by using some rules that are not well defined. His hypotheses are not derived from a theoretical mathematical model. There is not as yet a conventional theory suggesting what kind of rules people adopt to be successful in coordinating. Game theorists have had so far extreme difficulty in finding mathematical formulations for Schelling's idea. Many attempts have been made (e.g. Lewis, 1969, Bacharach, 1993, Sugden, 1995) but so far there is not a unified theory of focal points. As Sugden and Zamarron (2006) argue, many of these formalizations are not consistent with Schelling's informal theory. This shows that we can run experiments that, as Schelling did, are not necessarily based on any formal theory. Experiments therefore can do in this case more than theoretical models do.

The fact that Schelling's games cannot be fully described and represented in a mathematical model (e.g. the example quoted earlier) should not be seen as a weakness of the experimental design, rather the opposite. In fact this simply shows that these experiments substitute for mathematical models to explore what is of interest to the scientist. In Schelling's case the experiments are models of the world itself. They add something to the discipline allowing the economist to explore phenomena that could not be explored otherwise.

Schelling's experiments have however something in common with theoretical models. As in the case of models, the robustness of experimental results can be tested by varying some of the irrelevant features of the experiments. In this way Schelling has shown that focal points are actually robust in many settings. We are more confident about the results because many experiments carried out by Schelling and others have displayed the same phenomenon; and therefore we are more confident that what we have discovered is a real phenomenon.

6. General implications

In this paper we have seen three different kinds of experiments: model-implementing experiments not related to the target domain; an experiment that implements both part of the model and some features similar to the correspondent features of the target domain; and experiments that do not implement models but implement the target domain using some game-theoretical tools

Model-implementing experiments are not directly linked to the real world. They can be represented in the following way[1]:

[1] The diagram format is mainly based on Guala (2005)

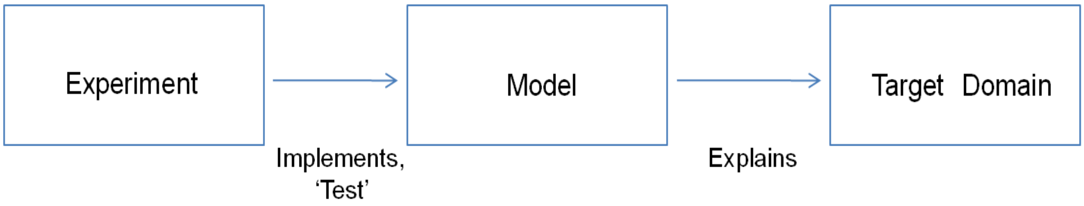


Figure 2: Structure of Experiments that Fully Implement Models

These kinds of experiments are related to the target domain only through the model. If the experiment differs from the model only by using human subjects, it is a test of some behavioural assumption used in the model, but not a test of what the model purports to say about its target.

Some may not agree with this view of model-implementing experiments. Vernon Smith (1982) maintains that experiments can be useful tools for testing theories, since the environment is much more complicated than the one the model formalises.

> "Microeconomic theory abstracts from a rich variety of human activities which are postulated not to be of relevance to human economic behaviour. The experimental laboratory, precisely it uses reward-motivated individuals drawn from the population of economic agents in the socioeconomic system, consists of a far richer and more complex set of circumstances than is parameterized in our theories. Since the abstractions of the laboratory are orders of magnitude smaller than those of economic theory, there can be no question that the laboratory provides ample possibilities for falsifying any theory we might wish to test." (Smith, p.936)

According to Smith then, the advantage of the lab is that it is much simpler than reality but much richer than the theory, thus providing a sound ground for testing models.

Smith's point of view seems to imply that the experiment lies between the model and the real world on a scale of abstractness. The lab is richer than theory but less complicated than the real world. For example the lab has real subjects whose behaviour is not constrained as that of the individuals postulated by the theory. For this reason the lab environment is a good testing ground for models.

Smith's methodological position would support Morgan et al.'s test of Varian's model of sales, while I have argued that this claim is unjustified. It is not enough for the lab to be richer than the theory for the experimental test to teach us something about the real world. It is also required in fact that the elements that make the lab environment richer than the theory are the features of the target domain that we should be implementing, in particular those features, suggested by the model, that makes the lab more similar to the target domain.

In a recent paper Croson and Gächter (2010) discuss how experiments can be used for different purposes to investigate economic issues. My concern here is their argument about how experiments can be used to test theoretical predictions.

> "Theories (models) are, by definition, simplifications of the world. The goal of a theory is to identify and isolate a phenomenon in order to understand its impacts. Ideally, theories yield unique and testable predictions. … Experiments test whether observed behaviour corresponds to the predictions of a particular model." (p. 125)

Croson and Gächter go on by noticing that with observational data it is difficult to test theories.

> "This difficulty stems from a number of sources. First, tests of model predictions using observational data are typically joint hypotheses tests. We need to test jointly whether the assumptions of the theory hold in the field, and whether the predictions of the theory hold in the field. … Experimental procedures in the lab can reduce the jointness of the test. … Of course experiments cannot eliminate the joint hypothesis testing problem entirely. However, the controlled laboratory situation implements the assumptions of the theory as closely as is possible." (p. 125)

We see from these quotes that from a philosophical standpoint, Croson and Gächter view theories as representations of the real world, they are therefore "realist". This, as it should be clear by now, does not make any difference in testing models.

Croson and Gächter see experiments as useful in that they allow the experimenter enough control over the environment (the lab) by implementing all the assumptions. They represent the relationship between theory and experiment in the following way (p. 125):

Theory-----------------Experiment (Lab/Field) ---------------- Observational data

The experiment may tell us something about the world by virtue of the model telling us something about the world. Croson and Gächter seem to hold the same view about model-implementing experiments as Smith's. However, the analysis I have provided in this chapter suggests that in order for the test to tell us something about the target domain, some of it must be implemented in the lab. So experiments that implement all the assumptions are best represented, I maintain, by the diagram shown in p. 134 of this chapter.

Plott (1991) is even more drastic than Smith about the usefulness of experiments, however his position is not much dissimilar from Smith's, in that the target domain box in figure 2 (p.134) seems to be completely removed.

"[There] was this belief suggested that the only effective way to create an experiment would be to mirror in every detail, to simulate, so to speak, some ongoing natural process." (p. 906).

"In other words, the experiment would be dismissed either because it did not mirror some natural process, or because it did. Once models, as opposed to economies, became the focus of research the simplicity of an experiment and perhaps even the absence of features of more complicated economies become an asset. The experiment should be judged by the lessons it teaches about theory and not by its similarity with what nature might happen to have created." (p.906)

Plott seems discouraged by the fact that nature is so complicated that whatever our good intentions are, we will never be able to reproduce it in the lab. However, it has to be said that experiments do not need to reproduce every bit of the phenomenon we want to analyse, as Plott seem to suggest, but just the relevant bits the scientist are interested in investigating and these bits need to be similar to the target domain if we want to learn something about it. So Plott's view about experiment, that is similar to Smith, Croson and Gächter's ones, by suggesting that we need to implement as much as possible of a model, seems to suggest that by testing models we can only learn something about the models but we are hopeless if we want to learn something about the target domain. I claim that experiments that implements models should be judged by what they can tell us about the world, if that is their purpose, as opposed to what Plott maintains[2].

The last two experimental approaches presented in the previous two sections differ from the one that implements models almost completely. While the latter do not represent any feature of the target domain, the former does it. Chamberlin reproduces a market with buyers and sellers that move around making bargains. This kind of market, even though it can be easily implemented in the lab, it is difficult to formalise mathematically. Schelling's experiments are not based on a theory at all. They use some game theoretical tools to implement real situations in order to understand how real people are able to coordinate. In this case the experiment itself can be seen as a model. Mäki (2005) has suggested that experiments can be seen as models and models as experiments. The main difference between the two is that an experiment isolates materially the mechanism while the

[2] Plott's argument shifts the responsability of deciding which features of the real world are relevant entirely on to the theorist. However there seems to be no reason why the experimenter cannot share part of responsability in deciding which features of the real world are worth representing in the lab. This is indeed what happens when experimental methods are used to investigate phenomena that are not (yet) formalized by models (like in Shelling case). In this case, since there is not model the experimenter acts as a theorist,

model does that theoretically. As opposed to Mäki I do not claim that all experiments are models but that some experiments (e.g. Schelling's ones) are models. Morgan (2005) has also compared models to experiments. Both of them, she maintains, have the ability to surprise us if they lead to unexpected results, however experiments are more powerful investigative tools because 'they are made of the same stuff as the real world" (p. 322). Guala (2005) suggests that models and experiments are similar in that they are artificial isolated systems. Artificial because created by the investigator and isolated because they do not include all the features of the real world but purposely only the relevant (or thought to be) ones.

By this direct link from experiments to the target domain we are actually able to learn something about that domain. Chamberlin's experiment uses the theory to get some testable predictions but the theory is not directly linked to the real world. Schelling's experiments are not linked at all with theoretical models, but represent the real world directly.

In Chamberlin's case, the experiment is the link between the model and the real world, being informative about the target domain and not vice-versa, as in the structure showed in fig. 2.

Chamberlin's experiment is a clear example of what the correct experimental strategy to use is if we want to learn something about the real world. The model has only the purpose to indicate which similarities we need to reproduce. Obviously Chamberlin's markets are not real markets but similar to real ones. His experiment is simple and does not lack control for its purposes. In this sense his model is both informative about the model and the real world.

We can summarise the structure of these kinds of experiments in the following way:

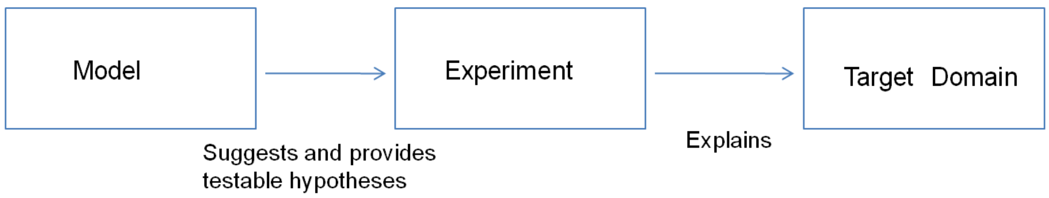


Figure 3: Structure of Experiments that Implement some Features of the Real World and Some Features of the Model[3]

[3]The diagram is from Guala (2005) p. 210-211.

Francesco Guala (2005) in *The methodology of experimental economics* suggests that experiments, as models, functions as mediators.

> 'The study of the target […] can sometimes proceed via the laboratory. Here's a typical route from the theory to the real world: a model is used to give structure to a speculation (a theoretical idea) about the economy.… Then a specific hypothesis is generated from analysing the model, for example by showing what would happen if…certain changes were made to a key variable. The hypothesis, however, is not tested directly on the target.… Hence, a laboratory system is built (an experiment) that can provide an answer to the research question. Then the experimental result is extended to the target by means of the external validity techniques… (p.210)

Experiments "mediate" between the model and the target system. However, Guala acknowledges, for the results to be extended to the target domain there must be some similarities, that we judge relevant, between lab and target. In my view, model-implementing experiments do not share this feature. In particular they are more connected to the model than to the target domain. Chamberlin's experiment, on the other hand, is a good example of an experiment that mediates between the model and the target domain, implementing part of the model and part of the target domain to assure, at least in principle, some external validity of the results.

The model, consistently with what Guala asserts, suggests the experiment and provides the testable hypothesis. The experiment implements the model's features, in order to control for the hypothesis tested, but it is also sufficiently similar to the target domain in order to be informative about it. Chamberlin's experiment however does more than that; it implements a real market which cannot feasibly be represented by a mathematical model. The experiment in this case has an advantage over the model.

Schelling's experiment belongs to that category of experiments that does not implement a model because there is no model available, at least at that time there was not. In this case, the experiment substitutes for the model itself. It is a tool that allows the scientist to investigate aspects of reality that would be inaccessible if the only tool available was that of modelling.

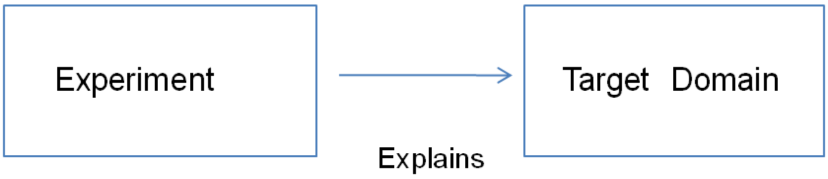


Figure 4: Structure of Experiments that are also Models of the Real World (Schelling Experiments)

Schelling's experiments are not the only instances of experiments that investigate reality without implementing models. In Schelling's case the framework used is borrowed from game theory but there are other cases, such as trust games, that share a similar experimental strategy. There is not in the first place an accepted definition of trust or trustworthiness. Nonetheless, the trust game and investor game (a modified version of a trust game) are widely used to study cooperation. For example Berg et al. (1995) explore reciprocity and the effect of information on cooperation, Croson and Buchan (1999) analyse gender and cultural differences and so on. It is worth noting that the target domain for these games, as in the case of Schelling's experiment, is not well defined, i.e. any situation that resembles the one used in the experiment can in principle be "the" target domain. So for example the players in a trust game can be a buyer or a seller, a bank and a customer, a citizen or a society and so on. So, these kinds of experiments are powerful tools because depending on how we interpret the game they can be applied to any situation that has some relevant resemblance to it.

PART 3: CONCLUSIONS

The main focus of this chapter has been on theoretical economic models that use abstract and unrealistic assumptions and are vaguely connected to the real world. These models' features have given rise to a heated methodological debate. I have given a summary of three different approaches: the instrumentalist, the realist and the fictionalist.

The analysis of two case studies, Varian's model of sales and Bikhchandani et al.'s model of fads, has shown that all three accounts are consistent with those models and the reason is that those models only have a vague and tenuous relationship with the real world. Not surprisingly, these methodological theories provide the same indications about how to test those models. That is, the experiment should be sufficiently similar to the target domain for the test to be able to tell us something about the phenomenon under investigation.

I then turned to experiments that implement models. I have focused on an experimental test of Varian's model and an experimental test of Bikhchandani et al.'s model. We know that for these kinds of experiments to be informative about the target domain, there must be some resemblance between the lab environment and the target domain that is suggested by the models themselves. However, our analysis shows that those experiments turn out to be tests of a single behavioural assumption that is commonly used in economics. So I asked what the advantages of testing assumptions in a model-like environment rather than just testing in isolation are. I concluded that in some cases model-implementing experiments are over-complicated and lack control for being tests of a single assumption. In other cases, however, they inherit the simplicity of the model giving rise to simple designs that allow for enough control to test a single assumption and at the same time provide a framework to other experimentalists wishing to extend the study.

Other experiments, although implementing models, are sufficiently similar to real world phenomena to be informative about the target domain. They however do more than that, they also substitute partially for the model, being the structure they implement too difficult to be represented mathematically. Therefore they are tools that can be used jointly with models in the investigation of the real world.

Lastly, there are experiments that do not implement models but can be thought of models themselves of the real world and are powerful tools to investigate phenomena for which there is not yet a mathematical formalisation. These experiments substitute for models in the scientific investigation.

Models have been widely used as an investigative tool in economics. Since the late 50's experiments have started to be used and became popular in the 80's. I suggest that the reason why experiments are often used as tests of models is probably because economics has for many years been a model-based science. However experiments are tools that can be used not only alongside models but also to integrate models and to substitute for them in the study of economic behaviour, adding something to the discipline that would not be possible otherwise.

Treatment B

Experimental Instructions for Sellers

1. Introduction

This is an experiment in the economics of market decision making. There are five participants in the experiment, one **seller** and four **buyers**. You are the **seller**. Please raise your hand if you have any questions at any point in the experiment.

The experiment is divided into two **phases** (plus some initial practice). Each phase is divided in 10 **periods**. In the experiment you are given an opportunity to earn experimental points. At the end of the experiment your earnings throughout phase 1 and 2 will be added up. You will gain one penny for every 6.5 experimental points you have (one pound for every 650 points), and you will be paid accordingly.

Roles: Each person will be either a buyer or a seller throughout the experiment. As a seller, you will choose how many units of a product you are willing to sell and at what price. The four buyers will then be given the chance to make purchases (if so they wish) from you at this offered price.

Product: In the table below you see an example of product of the kind that you can sell each period.

OUTCOME	PROBABILITY	RETURN
1	35%	50
2	15%	15
3	18%	80
4	22%	139
5	10%	10

Table 1: Example of Product

In this example, each unit of this product will give the *buyer* the chance to earn the following returns at the end of the experiment: 50 points with a probability of 35% (Outcome 1); 15 points with a probability of 15% (Outcome 2); 80 points with a probability of 18% (Outcome 3); and so on. As the *seller*, you *cannot* keep units of the product for yourself: the only way you may be able to make money out of a product is by selling it.

Phases: You will be able to sell **Product A** in phase 1 and **Product B** in phase 2. You can currently view both products on the computer screen; once the experiment starts, the product that is being traded at a given point in time will be displayed on the screen.

Practice: Before phase 1 gets started, you will do two periods just for practice with the example product showed in table 1, to get a better understanding of the experiment. Since these periods are only for practice, they do not count towards final earnings.

2. Market Decisions

The Seller's Decisions. In a given period the seller has the opportunity to sell Product A or Product B (depending on the phase). The decisions that the seller has to make is the **price** at which he or she is willing to sell the product and the **number of units** that he or she is willing to sell at this price.

The Buyers' Decisions. After the seller has made his or her decisions, buyers have the chance to buy the product on sale. Buyers are told the prices at which the product is on sale. Each of them is given an endowment of 390 points every period, and they can use it to buy units of the product if so they wish. First, buyers have to state how many units they are willing to buy from the seller. After they have done this, the computer will randomly select the order in which buyers will purchase. The following system will be followed by the computer:

- The first buyer will purchase the desired number of units if enough units are on sale at the lower price; otherwise he or she will purchase the available number of units.
- Then, the second buyer will purchase the desired number of units if there are enough units left; otherwise he or she will purchase the available number of units.
- Then, it will be the turn of the third buyer, if there are any units left.
- Finally, it will be the turn of the fourth buyer, if there are any units left.

3. Your Earnings

In a given period, as a seller, you may be able to earn points by selling units of the product. Each unit sold has a cost, however. You will be given information regarding the cost of each unit sold. The points that you could earn for each unit are given by the price at which you are selling the unit minus the cost of that unit. Costs are displayed in Table 2:

UNIT	COST
1st & 2nd	5
3rd & 4th	5
5th & 6th	10
7th & 8th	10
9th & 10th	47.5
11th & 12th	50
13th & 14th	52.5
15th & 16th	55
17th & 18th	57.5
19th & 20th	60
21st & 22nd	62.5
23rd & 24th	65
25th & 26th	67.5
27th & 28th	70
29th & 30th	72.5
31st & 32nd	75
33rd & 34th	77.5
35th & 36th	80
37th & 38th	82.5
39th & 40th	85

Table 2: Costs

Table 2 implies that the first unit sold has a cost of 5, the second unit sold has a cost of 5, the third unit sold has a cost of 5, the fourth unit sold has a cost of 5, the fifth unit sold has a cost of 10, the sixth unit sold has a cost of 10, and so on. The **total cost** from selling units is equal to the sum of the costs of producing each unit. For example, the total cost from selling 16 units is equal to 5 (cost on the first unit) plus 5 (cost on the second unit) plus 5 (cost on the third unit) plus 5 (cost on the fourth unit) plus 10 (cost on the fifth unit) plus 10 (cost on the sixth unit) plus 10 (cost on the seventh unit) plus 10 (cost on the eighth unit) plus 47.5 (cost on the ninth unit) and so on, for a total of cost of 470 points.

After the buyers choose the amounts that they are willing to buy, you will be informed about the

1) **number of units** you have been able to sell;
2) **revenue** from selling these units; revenue is equal to price times the number of units sold;
3) **total cost** of the units sold;
4) **profit,** which is equal to the difference between revenue and total cost.

If, for example, you are able to sell 12 units at a price of 70,

- the revenue is equal to $70 \times 12 = 840$ points;
- total cost is equal to

5	(cost of the first unit) +
5	(cost of the second unit) +
5	(cost of the third unit) +
5	(cost of the fourth unit) +
10	(cost of the fifth unit) +
10	(cost of the sixth unit) +
10	(cost of the seventh unit) +
10	(cost of the eighth unit) +
47.5	(cost of the ninth unit) +
47.5	(cost of the tenth unit) +
50	(cost of the eleventh unit) +

= 205 points;

- Profit = revenue – total cost = 840 – 205 = 635 points.

Note that, if you sell units at less than their cost, you will be making losses on these units, and so, if you want to make money, you may wish to make sure to sell at a price which is not below cost.

Your **final earnings** as a seller are paid at the end of the experiment and are given by the sum of the profits made in each period of phase 1 and 2 converted into pounds (every 6.5 points are converted into 1 penny, and so for example 6500 points are worth 10 pounds).

Before starting to take decisions, we ask you to fill the enclosed questionnaire, with the only purpose of checking whether you have understood these instructions. Raise your hand when you have completed the questionnaire.

Experimental Instructions for Buyers

1. Introduction

This is an experiment in the economics of market decision making. There are five participants in the experiment, one **seller** and four **buyers**. You are a **buyer**. Please raise your hand if you have any questions at any point in the experiment.

The experiment is divided into four **phases** (plus some initial practice). Each phase is divided in 10 **periods**. In the experiment you are given an opportunity to earn experimental points. At the end of the experiment your earnings throughout the four phases will be added up. You will gain one penny for every 9.75 experimental points you have (one pound for every 975 points), and you will be paid accordingly.

Roles: Each person will be either a buyer or a seller throughout the experiment. The seller will choose how many units of a product he or she is willing to sell and at what price. The four buyers will then be given the chance to make purchases (if so they wish) from the seller at this offered price.

Product: In the table below you see an example of product of the kind that you can buy each period.

OUTCOME	PROBABILITY	RETURN
1	35%	50
2	15%	15
3	18%	80
4	22%	139
5	10%	10

Table 1: Example of Product

In this example, each unit of this product will give you (as a *buyer*) the chance to earn the following returns at the end of the experiment: 50 points with a probability of 35% (Outcome 1); 15 points with a probability of 15% (Outcome 2); 80 points with a probability of 18% (Outcome 3); and so on. The *seller cannot* keep units of the product for himself or herself: the only way he or she may be able to make money out of a product is by selling it.

Phases: You will be able to buy **Product A** in phases 1 and 2 and **Product B** in phases 3 and 4. You can currently view both products on the computer screen; once the experiment starts, the product that is being traded at a given point in time will be displayed on the screen.

Practice: Before phase 1 gets started, you will do two periods just for practice with the example product showed in table 1, to get a better understanding of the experiment. Since these periods are only for practice, they do not count towards final earnings.

2. Market Decisions

The Seller's Decisions. In a given period the seller has the opportunity to sell Product A or Product B (depending on the phase). The decisions that the seller has to make is the **price** at which he or she is willing to sell the product and the **number of units** that he or she is willing to sell at this price.

The Buyers' Decisions. After the seller has made his or her decision, buyers have the chance to buy the product on sale. Buyers are told the price at which the product is on sale. Each of them is given an endowment of 390 points every period, and they can use it to buy units of the product if so they wish. First, buyers have to state how many units they are willing to buy from the seller. After they have done this, the computer will randomly select the order in which buyers will purchase. The following system will be followed by the computer:

- The first buyer will purchase the desired number of units if enough units are on sale; otherwise he or she will purchase the available number of units.
- Then, the second buyer will purchase the desired number of units if there are enough units left; otherwise he or she will purchase the available number of units.
- Then, it will be the turn of the third buyer, if there are any units left.
- Finally, it will be the turn of the fourth buyer, if there are any units left.

3. Your Earnings

As a buyer, you earn money in two ways:

- By retaining **unspent endowment**. As noted earlier, each buyer is given an endowment of 390 experimental points every period. Each unit bought in a given period will be paid with this endowment. Every unit of the endowment that is not used to buy units in the period is left unspent. This holds true for each period. At the end of the session, the sum of all the points left unspent in phases 1 through 4 is carried out.

- By buying **units of the product**. At the end of the experiment, the computer will add up all the units of Product A and Product B bought by each buyer in phases 1 through 4. The computer will then use the probabilities attached to Product A and Product B outcomes and, given those probabilities, randomly select an outcome for Product A and an outcome for Product B that determine the returns on each unit owned of each product. That is, each unit bought is worth the corresponding return, and the corresponding return is the same for all the units bought of each product. The overall return of Product A is given by the return of Product A times the number of units of Product A bought, and the overall return of Product B is given by the return of Product B times the number of units of Product B bought.

Your **final earnings** as a buyer are equal to the unspent endowment plus the overall return of Product A plus the overall return of Product B. Every 9.75 points you own are converted into 1 penny, and so for example 9750 points are worth 10 pounds.

Before starting to take decisions, we ask you to fill the enclosed questionnaire, with the only purpose of checking whether you have understood these instructions. Raise your hand when you have completed the questionnaire.

Treatment IS1

In this treatment buyers' instructions are the same as treatment B.

Experimental Instructions for Sellers

1. Introduction

This is an experiment in the economics of market decision making. There are five participants in the experiment, one **seller** and four **buyers**. You are the **seller**. Please raise your hand if you have any questions at any point in the experiment.

The experiment is divided into four **phases** (plus some initial practice). Each phase is divided in 10 **periods**. In the experiment you are given an opportunity to earn experimental points. At the end of the experiment your earnings throughout the four phases will be added up. You will gain one penny for every 9.75 experimental points you have (one pound for every 975 points), and you will be paid accordingly.

Roles: Each person will be either a buyer or a seller throughout the experiment. As a seller, you will choose how many units of a product you are willing to sell and at what price. The four buyers will then be given the chance to make purchases (if so they wish) from you at this offered price.

Product: In the table below you see an example of product of the kind that you can sell each period.

OUTCOME	PROBABILITY	RETURN
1	35%	50
2	15%	15
3	18%	80
4	22%	139
5	10%	10

Table 1: Example of Product

In this example, each unit of this product will give the *buyer* the chance to earn the following returns at the end of the experiment: 50 points with a probability of 35% (Outcome 1); 15 points with a probability of 15% (Outcome 2); 80 points with a probability of 18% (Outcome 3); and so on. As the *seller*, you *cannot* keep units of the product for yourself: the only way you may be able to make money out of a product is by selling it.

Phases: You will be able to sell **Product A** in phases 1 and 2 and **Product B** in phases 3 and 4. You can currently view both products on the computer screen; once the experiment starts, the product that is being traded at a given point in time will be displayed on the screen.

In order for you to get a better understanding of the products, you have also been given a sheet which orders outcomes in order of returns; buyers have not been given this sheet, and so they only see the outcomes in the scrambled order presented on the computer display. Both products have an average return of 60 points to buyers, though buyers have not been explicitly told this. Note that the fact that they have an average return of 60 points is not in itself a reason for you to provide a price of 60 (although you are free to choose this price if so you wish).

Practice: Before phase 1 gets started, you begin with six periods of practice in which you are asked to choose whether to buy units of Product A or Product B at a price randomly chosen by the computer, and given an endowment of 390 (fictional) points to do so. This is to get you a better understanding of how the products 'feel' to buyers (but note that buyers do *not* get an equivalent practice).

Afterwards, you will do two periods just for practice with the example product showed in table 1, to get a better understanding of the experiment as a seller. Since these periods (both as a buyer and as a seller) are only for practice, they do not count towards final earnings.

2. Market Decisions

The Seller's Decisions. In a given period the seller has the opportunity to sell Product A or Product B (depending on the phase). The decisions that the seller has to make is the **price** at which he or she is willing to sell the product and the **number of units** that he or she is willing to sell at this price.

The Buyers' Decisions. After the seller has made his or her decisions, buyers have the chance to buy the product on sale. Buyers are told the prices at which the product is on sale. Each of them is given an endowment of 390 points every period, and they can use it to buy units of the product if so they wish. First, buyers have to state how many units they are willing to buy from the seller. After they have done this, the computer will randomly select the order in which buyers will purchase. The following system will be followed by the computer:

- The first buyer will purchase the desired number of units if enough units are on sale at the lower price; otherwise he or she will purchase the available number of units.
- Then, the second buyer will purchase the desired number of units if there are enough units left; otherwise he or she will purchase the available number of units.
- Then, it will be the turn of the third buyer, if there are any units left.
- Finally, it will be the turn of the fourth buyer, if there are any units left.

3. Your Earnings

In a given period, as a seller, you may be able to earn points by selling units of the product. Each unit sold has a cost, however. You will be given information regarding the cost of each unit sold. The points that you could earn for each unit are given by the price at which you are selling the unit minus the cost of that unit. Costs are displayed in Table 2:

UNIT	COST
1st & 2nd	5
3rd & 4th	5
5th & 6th	10
7th & 8th	10
9th & 10th	47.5
11th & 12th	50
13th & 14th	52.5
15th & 16th	55
17th & 18th	57.5
19th & 20th	60
21st & 22nd	62.5
23rd & 24th	65
25th & 26th	67.5
27th & 28th	70
29th & 30th	72.5
31st & 32nd	75
33rd & 34th	77.5
35th & 36th	80
37th & 38th	82.5
39th & 40th	85

Table 2: Costs

Table 2 implies that the first unit sold has a cost of 5, the second unit sold has a cost of 5, the third and fourth units sold also have a cost of 5, the fifth, sixth, seventh and eighth units sold have a cost of 10, the ninth and tenth have a cost of 47.5, and so on. The **total cost** from selling units is equal to the sum of the costs of producing each unit. For example, the total cost from selling 16 units is equal to 5 (cost on the first unit) plus 5 (cost on the second unit) plus 5 (cost on the third unit) plus 5 (cost on the fourth unit) plus 10 (cost on the fifth unit) plus 10 (cost on the sixth unit) plus 10 (cost on the seventh unit) plus 10 (cost on the eighth unit) plus 47.5 (cost on the ninth unit) and so on, for a total of cost of 470 points.

After the buyers choose the amounts that they are willing to buy, you will be informed about the

5) **number of units** you have been able to sell;
6) **revenue** from selling these units; revenue is equal to price times the number of units sold;
7) **total cost** of the units sold;
8) **profit,** which is equal to the difference between revenue and total cost.

If, for example, you are able to sell 12 units at a price of 70,

- the revenue is equal to $70 \times 12 = 840$ points;
- total cost is equal to

5 (cost of the first unit) +
5 (cost of the second unit) +
5 (cost of the third unit) +
5 (cost of the fourth unit) +
10 (cost of the fifth unit) +
10 (cost of the sixth unit) +
10 (cost of the seventh unit) +
10 (cost of the eighth unit) +
47.5 (cost of the ninth unit) +
47.5 (cost of the tenth unit) +
50 (cost of the eleventh unit) +
50 (cost of the twelfth unit) +
= 255 points;

- Profit = revenue – total cost = 840 – 255 = 585 points.

Note that, if you sell units at less than their cost, you will be making losses on these units, and so, if you want to make money, you may wish to make sure to sell at a price which is not below cost.

Your **final earnings** as a seller are paid at the end of the experiment and are given by the sum of the profits made in each period of phases 1 through 4 converted into pounds (every 9.75 points are converted into 1 penny, and so for example 9750 points are worth 10 pounds).

4. Notes on the System

While the two products have the same average return, one of the two products is more complex than the other, because of the larger number of outcomes and because buyers see these outcomes only in scrambled order.

It is up to you to decide whether the greater complexity of one of the two products works to your advantage, disadvantage or neither in making as many profits as possible:

- on the one hand, buyers might dislike product complexity and so be less incline to buy the more complex product unless you lower the price;
- on the other hand, since it may be harder for buyers to figure out the real value of the more complex product (remember that they are not told explicitly their average return nor provided the products sheet), you might be able to sell the more complex product at a higher price.

It is up to you to decide whether either of these possibilities, both or neither is worth taking into account.

Before starting to take decisions, we ask you to fill the enclosed questionnaire, with the only purpose of checking whether you have understood these instructions. Raise your hand when you have completed the questionnaire.

Treatment IS2

The instructions for this treatment are the same as for IS1 with the difference that in IS2 there are two products on sale simultaneously.

.

Products Sheet

These are Product A and Product B, with outcomes ordered by return, from smallest to largest:

Product A

OUTCOMES	PROBABILITY	RETURN
1	50 %	10
2	20 %	65
3	30 %	140

Product B

OUTCOMES	PROBABILITY	RETURN
1	12.5 %	10
2	5 %	11.65
3	5 %	13.85
4	7.5 %	13.9
5	2 %	15.5
6	3 %	17.75
7	7.5 %	19.1
8	3 %	20.75
9	4.5 %	23
10	5 %	59.5
11	2 %	61.15
12	2 %	63.35
13	3 %	63.4
14	0.8 %	65
15	1.2 %	67.25
16	3 %	68.6
17	1.2 %	70.25
18	1.8 %	72.5
19	7.5 %	127
20	3 %	128.65
21	3 %	130.85
22	4.5 %	130.9
23	1.2 %	132.5
24	1.8 %	134.75
25	4.5 %	136.1
26	1.8 %	137.75
27	2.7 %	140

Only sellers have been given this Products Sheet.

Products Sheet

These are Product A and Product B, with outcomes ordered by return, from smallest to largest:

Product A

OUTCOMES	PROBABILITY	RETURN
1	12.5 %	10
2	5 %	11.65
3	5 %	13.85
4	7.5 %	13.9
5	2 %	15.5
6	3 %	17.75
7	7.5 %	19.1
8	3 %	20.75
9	4.5 %	23
10	5 %	59.5
11	2 %	61.15
12	2 %	63.35
13	3 %	63.4
14	0.8 %	65
15	1.2 %	67.25
16	3 %	68.6
17	1.2 %	70.25
18	1.8 %	72.5
19	7.5 %	127
20	3 %	128.65
21	3 %	130.85
22	4.5 %	130.9
23	1.2 %	132.5
24	1.8 %	134.75
25	4.5 %	136.1
26	1.8 %	137.75
27	2.7 %	140

Product B

OUTCOMES	PROBABILITY	RETURN
1	50 %	10
2	20 %	65
3	30 %	140

Only sellers have been given this Products Sheet.

Products Sheet

These are Product A and Product B, with outcomes ordered by return, from smallest to largest:

Product A

OUTCOMES	PROBABILITY	RETURN
2	50%	3
3	20%	66
1	30%	151

Product B

OUTCOMES	PROBABILITY	RETURN
3	12.5%	3
13	5.0%	4.89
26	5.0%	7.41
1	7.5%	7.44
9	2.0%	9.3
24	3.0%	11.85
5	7.5%	13.36
27	3.0%	15.25
22	4.5%	17.8
14	5.0%	59.7
6	2.0%	61.59
23	2.0%	64.11
2	3.0%	64.14
21	0.8%	66
19	1.2%	68.55
25	3.0%	70.06
18	1.2%	71.95
11	1.8%	74.5
15	7.5%	136.2
12	3.0%	138.09
8	3.0%	140.61
16	4.5%	140.64
20	1.2%	142.5
17	1.8%	145.05
10	4.5%	146.56
7	1.8%	148.45
4	2.7%	151

Only sellers have been given this Products Sheet.

Products Sheet

These are Product A and Product B, with outcomes ordered by return, from smallest to largest:

Product A

OUTCOMES	PROBABILITY	RETURN
3	12.5%	3
13	5.0%	4.89
26	5.0%	7.41
1	7.5%	7.44
9	2.0%	9.3
24	3.0%	11.85
5	7.5%	13.36
27	3.0%	15.25
22	4.5%	17.8
14	5.0%	59.7
6	2.0%	61.59
23	2.0%	64.11
2	3.0%	64.14
21	0.8%	66
19	1.2%	68.55
25	3.0%	70.06
18	1.2%	71.95
11	1.8%	74.5
15	7.5%	136.2
12	3.0%	138.09
8	3.0%	140.61
16	4.5%	140.64
20	1.2%	142.5
17	1.8%	145.05
10	4.5%	146.56
7	1.8%	148.45
4	2.7%	151

Product B

OUTCOMES	PROBABILITY	RETURN
2	50%	3
3	20%	66
1	30%	151

Only sellers have been given this Products Sheet.

Treatment IC1

In this treatment there is a computerised seller, therefore there are not instructions for the seller.

Experimental Instructions

1. Introduction

This is an experiment in the economics of market decision making. Please raise your hand if you have any questions at any point in the experiment.

The experiment is divided into four **phases** (plus some initial practice). Each phase is divided in 10 **periods**. In the experiment you are given an opportunity to earn experimental points. At the end of the experiment your earnings throughout the experiment will be added up. You will gain one penny for every 9.75 experimental points you have (one pound for every 975 points), and you will be paid accordingly.

Roles: You will be a buyer throughout the experiment. A computerised seller will choose the price at which it is willing to sell the good. You will then be given the chance to make purchases (if so you wish) from the seller at this offered price.

Product: In the table below you see an example of product of the kind that you can buy each period.

OUTCOME	PROBABILITY	RETURN
1	35%	50
2	15%	15
3	18%	80
4	22%	139
5	10%	10

Table 1: Example of Product

In this example, each unit of this product will give you (as a *buyer*) the chance to earn the following returns at the end of the experiment: 50 points with a probability of 35% (Outcome 1); 15 points with a probability of 15% (Outcome 2); 80 points with a probability of 18% (Outcome 3); and so on.

Phases: You will be able to buy **Product A** in phase 1 and 2 and **Product B** in phase 3 and 4. You can currently view both products on the computer screen; once the experiment starts, the product that is being traded at a given point in time will be displayed on the screen.

Practice: Before phase 1 gets started, you will do two periods just for practice with the example product showed in table 1, to get a better understanding of the experiment. Since these periods are only for practice, they do not count towards final earnings.

2. Market Decisions

The Seller's Decisions. In a given period the seller has the opportunity to sell Product A or Product B (depending on the phase). The decision that the seller has to make is the **price** at which it is willing to sell the product.

The Buyers' Decisions. After the seller has made his or her decision, you have the chance to buy the product on sale. You are told the price at which the product is on sale. You are given an endowment of 390 points every period, and you can use it to buy units of the product if so you wish. You do so by stating how many units you are willing to buy from the seller.

3. Your Earnings

As a buyer, you earn money in two ways:

- By retaining **unspent endowment**. As noted earlier, each buyer is given an endowment of 390 experimental points every period. Each unit bought in a given period will be paid with this endowment. Every unit of the endowment that is not used to buy units in the period is left unspent. This holds true for each period. At the end of the session, the sum of all the points left unspent in phase 1, 2, 3 and 4 is carried out.
- By buying **units of the product**. At the end of the experiment, the computer will add up all the units of Product A and Product B bought by each buyer in phase 1 and 2. The computer will then use the probabilities attached to Product A and Product B outcomes and, given those probabilities, randomly

select an outcome for Product A and an outcome for Product B that determine the returns on each unit owned of each product. That is, each unit bought is worth the corresponding return, and the corresponding return is the same for all the units bought of each product. The overall return of Product A is given by the return of Product A times the number of units of Product A bought, and the overall return of Product B is given by the return of Product B times the number of units of Product B bought.

Your **final earnings** as a buyer are equal to the unspent endowment plus the overall return of Product A plus the overall return of Product B. Every 9.75 points you own are converted into 1 penny, and so for example 9750 points are worth 10 pounds.

Before starting to take decisions, we ask you to fill the enclosed questionnaire, with the only purpose of checking whether you have understood these instructions. Raise your hand when you have completed the questionnaire.

Treatment IC2

The instructions for this treatment are the same as for IC1 with the difference that in IC2 there are two products on sale simultaneously.

APPENDIX B ESSAY1 – PREDICTIONS FOR THE MARKET TREATMENT

It is useful to know what price a profit maximizing monopolist would set under different assumptions regarding buyers' risk and complexity attitudes. To remind the reader, if all agents are neutral to risk, given the seller cost function, a profit maximizing monopolist should sell 20 units making a profit of 495. At that price buyers, who have an endowment of 390, can buy up to 6 units but since the seller does not sell more than 20 they will be rationed, so can buy on average 5 units each.

In this section, we explore 2 cases. A) a share of buyers is risk averse and the rest is risk loving; B) under different assumptions on risk attitudes we also assume that a share of buyers is complexity averse and the rest is complexity exploitable. It is worth noting that, if assume that buyers are not affected by complexity and, as already pointed out in chapter 2, the simple and the complex products are indistinguishable in their degree of riskiness, the price charged for the complex product should not differ from the price charged for the simple one.

CASE A - A SHARE OF BUYERS IS RISK AVERSE AND THE REST IS RISK LOVING

In our experiment there are 4 buyers per market. We therefore consider 4 possible combinations: i) all buyers are risk averse; ii) 3 buyers are risk averse and 1 is risk loving; iii) 2 buyers are risk averse and 2 are risk loving; iv) 1 buyer is risk averse and 3 are risk loving.

Table 1 presents the prices that profits with the price that a profit maximizing seller should charge should her know the exact distribution of buyers' risk attitudes and the profits. As already pointed out above the price charged for the simple product should not differ from the price charged for the complex products since attitudes towards complexity are absent.

i) **All buyers are risk averse**

Prediction. The price that a profit maximize monopolist should charge cannot be higher or equal to 60.

Proof. Risk averse agents by definition prefer the certainty equivalent of a lottery over its expected value. Therefore they are willing to pay a price as high as 59. The monopolist will maximize profits selling 18 units making a profit of 477. Any quantity above 18 would reduce, given the cost function, the profits.

ii) 3 buyers are risk averse and 1 is risk loving

Table 2 shows profits when the price is lower than 60 but equal or above 40. It also shows the price that the monopolist can charge and the maximizing quantity supplied given her cost function. The table also shows the profits that the monopolist makes in case charges prices higher than 60 to the risk loving buyer and the quantity supplied, given the buyer's budget constraint.

In this case there are 5 different scenarios, depending on the degree of risk aversion and risk lovingness of the buyers.

First scenario. Risk averse buyers are willing to buy at price at least as high as 52 or above.

Prediction. Independently of the degree of risk lovingness of the 4^{th} buyer, the profit maximizing monopolist should set a price of 52 or above, but below 60, selling to all buyers, making a profit at most of 477.

Proof. If 3 risk averse agents are willing to buy at least at 52 the profit the monopolist can make is at least 368. Charging a higher price to the risk loving buyer will not get the monopolist a profit higher than 368 (i.e. when the price is 97).

3 Risk averse & 1 risk loving buyers

Quantity	Price<60	Revenues	Profits	Quantity	Price>60	Revenues	Profits	Quantity	Price>60	Revenues	Profits
8	40	320	260	6	61	366	326	4	81	324	304
8	41	328	268	6	62	372	332	4	82	328	308
8	42	336	276	6	63	378	338	4	83	332	312
8	43	344	284	6	64	384	344	4	84	336	316
8	44	352	292	6	65	390	350	4	85	340	320
8	45	360	300	5	66	330	300	4	86	344	324
8	46	368	308	5	67	335	305	4	87	348	328
8	47	376	316	5	68	340	310	4	88	352	332
10	48	480	325	5	69	345	315	4	89	356	336
10	49	490	335	5	70	350	320	4	90	360	340
12	50	600	345	5	71	355	325	4	91	364	344
12	51	612	357	5	72	360	330	4	92	368	348
12	52	624	369	5	73	365	335	4	93	372	352
14	53	742	382	5	74	370	340	4	94	376	356
14	54	756	396	5	75	375	345	4	95	380	360
16	55	880	410	5	76	380	350	4	96	384	364
16	56	896	426	5	77	385	355	4	97	388	368
16	57	912	442	5	78	390	360	3	98	294	274
18	58	1044	459	4	79	316	296	3	99	297	277
18	59	1062	477	4	80	320	300	3	100	300	280

Table 1: Profits when 3 buyers are risk averse and 1 buyer is risk loving

Second scenario. Risk averse buyers are willing to buy at price of 51.

Prediction. The monopolist would set a price of 51 or at least as high as 78.

Proof. If the monopolist sets a price of 51 he should sell all buyers not more than 12 units making a profit of 357. The monopolist can though decide to set a price of 78, if the risk loving buyer is willing to buy at such a high price. In this case she can sell 4 units (given the buyer's budget constraint) making a profit of 360.

Third scenario. The risk averse buyers are willing to buy at a price not higher than 50.

Prediction. In this case a profit maximizing monopolist would set a price of 50 or at least as high as 75.

Proof. If 3 risk averse agents are willing to buy at a maximum price of 50, a profit maximising seller would set a price of 50, supplying no more than 12 units, giving her cost function, and making a profit of 345. Alternatively he can charge a minimum price of 75, depending on the degree of risk lovingness of the other buyer, making a profit at least of 345.

Fourth scenario. The risk averse buyers are willing to buy at a price not higher than 49.

Prediction. In this case a profit maximizing monopolist would set a price of 49 or at least as high as 73.

Proof. If 3 risk averse agents are willing to buy at a maximum price of 49, a profit maximising seller would set a price of 49, supplying no more than 10 units, giving her cost function, and making a profit of 335. Alternatively he can charge a minimum price of 73, depending on the degree of risk lovingness of the other buyer, making a profit at least of 335.

Fifth scenario. The risk averse buyers are willing to buy at a price not higher than 48..

Prediction. In this case a profit maximizing monopolist would set a price of 61 and sell to risk loving buyer.

Proof. If 3 risk averse agents are willing to buy at a maximum price of 48, a profit maximising seller would not make a profit higher than 325, selling at the most 10 units. She therefore sets a price of at least 61 selling to the risk loving buyer no more than 6 units (given the buyer's budget constraint) and making a profit of at least 326.

iii) 2 buyers are risk averse and 2 are risk loving

In this case, we only have one scenario.

Prediction: A profit maximizing monopolist should set a price of at least 61 selling only to the risk loving buyers.

Proof. If the monopolist sets a price of 59 she can only make a profit of 453, selling, given the buyers' budget constraint, at most 12 units. If she decides to sell only to risk loving buyers setting a price of at least 6, she makes a profit of at least 477, selling at most 12 units given the buyers' budget constraint.

iv) 1 buyer is risk averse and 3 buyers are risk loving

Prediction. The profit maximising monopolist should set a price of 61 and sell only to the risk loving buyers.

Proof. If the monopolist sets a price of 61 she would be able to sell 18 units at the most (given the buyers' budget constraint) making a profit of 513. If she sets a price as high as 59 she can sell to all buyers making a profit of 477, selling at most 18 units given her cost function.

C) A SHARE OF BUYERS IS COMPLEXITY AVERSE AND THE REST IS COMPLEXITY EXPLOITABLE.

This is the last set of assumptions we consider. We assume homogeneity in buyers' risk attitudes (considering all possible mixes would lead to an extremely high number of cases). There then three possible cases. All buyers are risk averse, all buyers are risk neutral and all buyers are risk loving. For each risk attitude we consider three possible cases, all buyers are complexity averse, all buyers are complexity exploitable and finally some buyers are complexity averse and some are complexity exploitable. There are overall 9 possible cases. It has to be noticed here that the price of the simple product is not affected, since complexity attitudes are only relevant for complex products.

1) All buyers are risk averse.

In this case a profit maximising monopolist should set a price lower than 60 for the simple product. Let us see how the different complexity attitudes affect the price of the complex one.

i) All buyers are complexity averse.

When all buyers are complexity averse and complexity aversion and risk aversion effects add up, a monopolist should set a price for the complex product that is lower than the price for the simple.

Prediction. The price of the complex product is lower than 60 and lower than the price of the simple product.

Proof. By definition a risk averse subject will pay a lottery less than its expected value. If in addition all subjects are also complexity averse, they will ask for a higher premium to compensate for the complexity aversion.

ii) All buyers are complexity exploitable.

When all buyers are complexity exploitable, a monopolist should set a price for the complex product that is higher than the price for the simple one, if complexity exploitation more than offset risk aversion then the price should be higher than 60.

Prediction. The price of the complex product is higher than the price of the simple one

Proof. By definition a risk averse subject will pay a lottery less than its expected value. If in addition all subjects are also complexity exploitable, the premium they ask as a compensation for the risk they are taking buying the lottery will be reduced by their attitudes towards complexity.

iii) A share of buyers is complexity averse and a share is complexity exploitable.

When some of the buyers are complexity averse and some are complexity exploitable then the final effect on the price of the complex lottery depends on how the two complexity effects interact with each other. Here we need some further conjectures as to how the two complexity effects interact with each other. For simplicity we assume that the two effects have the same intensity, so if one buyer is complexity averse and the other is complexity exploitable then the two effects completely offset each other.

Prediction. If the share of complexity averse subjects is higher than the share of complexity exploitable subjects then the price of the complex lottery should be lower than the price of the simple. So we are back to case i). If the number of complexity averse subjects is lower than that of complexity exploitable ones then the predictions are the same as for case ii). In case the number of the complexity averse subjects is the same as the number of complexity exploitable then complexity attitudes balance out and only risk attitudes matter. So the price of the complex lottery is not different from the price of the simple and both are lower than 60.

Proof. See respectively case 1)-i) and 1)- ii) and chapter 2 for the last one.

2) **All buyers are risk neutral.**

In this case a profit maximising monopolist should set of 60 for the simple product. Let us see how the different complexity attitudes affect the price of the complex one.

i) All buyers are complexity averse.

When all buyers are complexity averse, a monopolist should set a price for the complex product that is lower than the price for the simple, which is 60.

Prediction. The price of the complex product is lower than 60.

Proof. By definition a risk neutral subject will pay for a lottery its expected value. If in addition all subjects are also complexity averse, they will ask for a higher premium to compensate for the complexity aversion. Therefore the price should be lower than 60.

ii) All buyers are complexity exploitable.

When all buyers are complexity exploitable, a monopolist should set a price for the complex product that is higher than the price set for simple one, which is 60.

Prediction. The price of the complex product is higher than 60.

Proof. By definition a risk neutral subject will pay for a lottery its expected value. If in addition to that all subjects are also complexity exploitable, they will be willing to pay more because confused.

iii) A share of buyers is complexity averse and a share is complexity exploitable.

Prediction. If the share of complexity averse subjects is higher than the share of complexity exploitable subjects then the price of the complex lottery should be lower than the price of the simple. So we are back to case i). If the number of complexity averse subjects is lower than that of complexity exploitable ones then the predictions are the same as for case ii). In case the number of the complexity averse subjects is the same as the number of complexity exploitable then complexity attitudes balance out and only risk attitudes matter. So the price for complex lottery is not different from the price of the simple which is 60..

Proof. See respectively case 2)-i) and 2)- ii) and case A).

3) All buyers are risk neutral.

In this case a profit maximising monopolist should set a price higher than 60 for the simple product. Let us see how the different complexity attitudes affect the price of the complex one.

i) All buyers are complexity averse.

When all buyers are complexity averse, a monopolist should set a price for the complex product that is lower than the price for the simple. It is however not possible to say whether the price should be higher or lower than 60. This in fact depends on whether complexity attitudes partially, completely or more than offset risk attitudes.

Prediction. The price of the complex product is lower than the price of the simple one.

Proof. By definition a risk loving subject is willing to pay for a lottery more than its expected value. If in addition to that all subjects are also complexity averse, they will be willing to pay as much given that their complexity aversion works in the opposite direction as their risk lovingness.

ii) **All buyers are complexity exploitable.**

When all buyers are complexity exploitable, a monopolist should set a price for the complex product that is higher than the price set for simple one, which already is more than 60

Prediction. The price of the complex product is higher than 60 and higher than the price of the simple product..

Proof. By definition a risk loving subject is willing to pay for a lottery more than its expected value. If in addition to that all subjects are also complexity exploitable, they will be willing to pay even more given that their complexity exploitation strengthens their risk lovingness.

iii) **A share of buyers is complexity averse and a share is complexity exploitable.**

Prediction. If the share of complexity averse subjects is higher than the share of complexity exploitable subjects then the price of the complex lottery is lower than the price of the simple. So we are back to case i). If the number of complexity averse subjects is lower than that of complexity exploitable ones then the predictions are the same as for case ii). In case the number of the complexity averse subjects is the same as the number of complexity exploitable then complexity attitudes balance out and only risk attitudes matter. So the price for complex lottery is not different from the price of the simple which is higher than 60.

Proof. See respectively case 2)-i) and 2)- ii) for the first two cases. For the third one, i.e. when the number of complexity averse subjects is the same as the number of complexity exploitable then the price is determined on the basis of complexity attitudes. In this case all buyers are risk loving therefore the price that a profit maximizing monopolist should set is higher than 60. If in fact the price is 60, for example 61 the monopolist should not sell more than 20 units, given her cost function, making a profit of 635. If she were to set a price lower than that, for example 60 the profits will be lower, 495 in this example.

Binary Choice Task

Experimental Instructions

This is an experiment in the economics of decision making. It is divided in two stages.

Instructions for Stage 1

In the course of stage 1, over a number of periods you will be asked to choose between lotteries that pay returns in experimental points with given probabilities.

In the table below you see an example of a lottery of the kind a unit of which you can choose each period:

OUTCOME	PROBABILITY	RETURN
1	35%	50
2	15%	15
3	18%	80
4	22%	139
5	10%	10

Table 1: Example of Product

In this example, the lottery will give you the chance to earn the following returns at the end of the experiment: 50 points with a probability of 35% (Outcome 1); 15 points with a probability of 15% (Outcome 2); 80 points with a probability of 18% (Outcome 3); and so on.

If you see a lottery which provides a given return with a probability of 100%, this means that, if you choose this lottery, you will get that return for sure.

Stage 1 Earnings

At the end of the experiment the computer uses the probabilities attached to each lottery you chose and randomly selects a corresponding outcome in each case. This outcome determines the return for the corresponding lottery. Returns from each lottery are added up to determine your overall stage 1 earnings. Every 9.75 points you own are converted into 1 penny, and so for example 9750 points are worth 10 pounds.

Before starting to take decisions, we ask you to fill the enclosed questionnaire, with the only purpose of checking whether you have understood these instructions. Raise your hand when you have completed the questionnaire.

Posted offer Market Task

Instructions for Stage 2

1. Introduction

Stage 2 is divided in twenty periods. Each period you are given an endowment of 650 points, and a computerised seller will choose the price at which it is willing to sell a lottery. You will then be given the chance to buy one or more units of this lottery (if so you wish) from the seller at this offered price.

Lottery: On the computer screen you can view the lottery which is available for you to buy during the stage. It has exactly the same structure as the lotteries which you have been presented with in stage 1 of the experiment.

Practice: Before stage 2 gets started, you will do two periods just for practice with the example lottery shown in table 1 (in the instructions for stage 1). Since these periods are only for practice, they do not count towards final earnings.

2. Your Decision

Each period, after the computerised seller has made his or her decision about the price of the lottery, you have the chance to buy the lottery being sold. You are told the price at which the lottery is being sold. You are given an endowment of 650 points every period, and you can use it to buy units of the lottery if so you wish. You do so by stating how many units you are willing to buy from the seller.

3. Your Earnings

As a buyer, you earn money in stage 2 in two ways:

- By retaining unspent endowment. As noted earlier, each buyer is given an endowment of 650 points every period. Each unit bought in a given period will be paid with this endowment. Every unit of the endowment that is not used to buy units in the period is left unspent. This holds true for each period. At the end of the session, the sum of all the points left unspent in all periods is carried out.
- By buying units of the lottery. At the end of the experiment, the computer will add up all the units of the lottery bought in the 20 periods. The computer will then use the probabilities attached to the lottery outcomes and, given those probabilities, randomly select an outcome for the lottery that determines the returns on each unit owned. That is, each unit bought is worth the corresponding return, and the corresponding return is the same for all the units bought of each lottery.

Your overall earnings from stage 2 are then given by the sum of the returns of each lottery times the number of units of each lottery bought, plus the unspent endowment. Every 9.75 points you own are converted into 1 penny, and so for example 9750 points are worth 10 pounds.

4. Your Overall Experimental Earnings

Your overall experimental earnings will be equal to the sum of your overall earnings from stage 1 and your overall earnings from stage 2.

Before starting to take decisions, we ask you to fill the enclosed questionnaire, with the only purpose of checking whether you have understood these instructions. Raise your hand when you have completed the questionnaire.

APPENDIX D ESSAY2 – DETAILED DISCUSSION OF THE EXPERIMENTAL RESULTS

In the sections 2 and 3 where we have presented a detailed discussion on anchoring and shaping we point out that perceived gain and loss may be relevant for buyers' purchasing and, consequently, expenditure behaviour. In this appendix we present a detailed discussion on how this relates to the results in section 6.1 and to the regression analysis of chapter 3.

Quantity Bought. In phase 2 of the treatment 0 subjects purchase 66% less than in phase 1. In the same phase of treatment 1 they purchase 58% less than in phase 1. This can be explained in terms of perceived loss. In phase 1 subjects form a perceived price (that can be reasonably assumed to be the average price observed in this phase) that use to estimate the value of the good. In treatments 0 and 1 the price is lower in phase 1 than in phase 2, this increase in price is perceived as a loss and therefore subjects tend to buy less. In treatments 3 and 4 the opposite reasoning applies. The price is lower in the first phase of both treatments, therefore the price decrease in phase 2 is perceived as a gain. In treatment 2 the pricing strategy does not change significantly in the two phases, the reference price is therefore the same the same in both phases. If subjects make purchasing decision by contrasting the difference between the actual price and the that, their purchasing behaviour should not differ in the two phases. And this is what the results suggest. The quantity bought is in fact virtually the same.

One may wonder to what extent the results are driven by loss aversion. Let us consider now treatments 1 and 3 (the same applies to treatment 0 and 4). In treatment 1 we observe an increase in the average price of 30%, while in treatment 3 we observe a decrease of 20% with respect to reference price. If price changes are judged with respect to the reference price, if the changes in the price levels of both treatments were the same and additionally if subjects were loss averse then we should observe in treatment 1 an increase in the quantity bought in phase 2 that is greater (in absolute value) than the increase in quantity bought in phase 2 of treatment 3. The results show that the quantity bought in phase 2 increases by approximately 60% in treatment 1 and decreases by more or less the same proportion in treatment 3. If the perceived loss was the same as the perceived gain we would expect the decrease in the quantity bought in treatment 3 to be greater than the increase in the quantity bought in treatment 1. However, since the perceived loss of 30% is greater than the perceived gain of 20%, these results are in contrast with the loss aversion hypothesis. Suppose in fact that the percentage in increase in prices were the same as the percentage increase. In this case, if subjects were loss averse, we should expect a percentage decrease in demand in absolute value greater than the percentage increase in the case of the perceived gain. But if the decrease in absolute value is even greater than that, then the decrease in demand should be even greater. However we observe a symmetric purchasing behaviour with opposite sign in the two treatments. We are assuming that subjects judge any price change relative to the reference price in percentages, but in principle it is possible that the differences are judged in absolute terms. This would be in line with the loss aversion hypothesis. The value function of the prospect theory (Kahneman and Tversky, 1978) is defined over gains and losses relative to a reference point of value 0. So if we apply the same reasoning here, then the average price both in treatment 1 and 3 changes of 30 experimental points (a loss in treatment 1 and a gain in treatment 3). Similarly, it is not clear whether the change in demand should be measured in relative or absolute terms. If we measure it in absolute terms, our results would provide evidence that contradicts the loss aversion hypothesis. In the latter case however, the evidence would be in line with loss aversion. In fact in treatment 3 the number of units bought increase of about 0.7 while in treatment 1 it decreases of about 1.50.

Expenditure. The change in subjects' purchasing behaviour form phase 1 to phase is unsurprisingly reflected in the expenditure behaviour. Let us focus on treatments 1 and 3 (for treatments 0 and 4 the same reasoning applies). In the treatment 1, where a low-medium pricing strategy is implemented, expenditure in phase 2 decreases by about 45%. In treatment 4 where a high-medium pricing strategy is implemented, expenditure increases by about 27%. The perceived loss in treatment 1 leads subjects to spend less than what they have done if the price did not change (i.e. in treatment 2). In treatment 3 on the other hand, the perceived gain leads them to spend more. These results, contrary to what observed for the quantity bought, are not in contrast with the loss aversion hypothesis. However given that prices changes proportionally more in treatment 1, we should expect expenditure to change proportionally more in that treatment than that in treatment 3. However these results could be also due to the fact that the price change is proportionally greater in treatment 1 than it is in treatment 3. So, for this to be conclusive evidence of loss aversion, the change in prices should be the proportionally the same. If on the other hand we consider the absolute changes in prices, that is, 30 experimental units on average in both treatments, these results would provide evidence in favour of the loss aversion hypothesis.

In treatment 2, expenditure is virtually the same in both phases. This is consistent with the perceived gain – perceived loss interpretation. The pricing strategy in phase 2 is the same as it is in phase 1, therefore there is no perceived gain or loss affecting subjects' expenditure behaviour.

Regression analysis. The size of the coefficients T0, T1, T3 and T4 of the regression analysis presented in table 12 of chapter 3 can also be explained in terms of perceived gain and loss in the same way that it has been done for quantity demanded and expenditure. Perceived gain with respect to the reference price leads subjects to buy more in treatments 3 and 4 and to buy less in treatments 0 and 1 relative to treatment 2. However in treatment 1 and 3 the effect is more pronounced than in treatments 0 and 4. As suggested in sections 2 and 3, if subjects' reference price depends on the consistency of the pricing strategy used, then this is not surprising. In treatment 0 and 4 the pricing strategy changes in period 6 of phase 1. We can therefore think of the overall pricing strategy implemented in phase 1 of these treatments less consistent than the pricing strategy implemented in the same phase in treatments 1 and 4. This results in those strategies being less effective in influencing purchasing behaviour. In fact, given that the average price in phase 1 of treatment 0 is higher than in the average price in the same phase of treatment 1, we should expect the perceived gain to be higher in this treatment than that in treatment 1, as a consequence demand should be greater in this treatment than that in treatment 1. However this is not the case. Similarly, in treatment 4, being the average price greater than that in treatment 3, we should also expect the perceived gain greater than that in treatment 3, and this should be reflected in higher sales in treatment 4. However again, given the size of the estimated coefficient, we observe the opposite result. It is worth noting that this is line with the results that we observe in the previous experiment. There we suggest that we observe clearer shaping effects when a consistent pricing strategy is used (i.e. in the IC treatments) as opposed to the treatments that employ subjects as sellers. In this context we can think of shaping effects as being influenced by the variance of prices. If the variance is high it is difficult for subject to form a reference price and therefore to judge any change in price in terms of gains and losses. If the variance is low, that is the pricing strategy is consistent, the reference price is well formed and therefore any change is judged in terms of perceived loss or gain affecting behaviour in the expected way. The size of the coefficient of the variables T0, T1, T3 and T4 shows also evidence consistent with the loss aversion hypothesis. Again it is debatable if this can be interpreted as evidence in favour of this hypothesis. In fact, as discussed before, if the price change is evaluated in percentages terms, then perceived losses and gains are of different size. If, on the other hand, we measure the change in experimental units, then this would be conclusive evidence of loss aversion. In fact, the size of the coefficients T0 and T1 is greater than the size of the coefficients T3 and T4 respectively, showing that subjects are more sensitive to losses than to gains of the same size. The size of the coefficients of the variable T0 and T4 may be thought of to be in contrast with the imprinting account first proposed by Ariely et al (2003). As explained in the main text this may be due to the fact that pricing strategy in the treatments T0 and T4 is less consistent.

As already noticed in chapter 2 the signs of the coefficients of the MarketPeriod variables – statistically significant in the case of MarketPeriod*T1 and MarketPeriod*T3 – suggest that the shaping effect tends to wear off with time. This can be interpreted again in terms of perceived loss and perceived gain. At the beginning of phase 2 the sudden change in the pricing strategy is judged by the subjects relative to reference price induced in the first phase. However, this reference price is slowly upgraded to take into account the new prices, so that later on in phase 2 the reference price increases in the treatments that implement a low-medium pricing strategy and decreases in the treatments where the opposite strategy is implemented. The size of the perceived losses and gains therefore decreases as well, weakening the shaping effects (this is in line with what has been said before about the duration of imprinting).

List of references

Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica* 21, 503-546.

Anderson, L. R. and Holt, C. A. (1997). Information Cascades in the Laboratory. *The American Economic Review* 87, 847-862.

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal* 100, 464-477.

Ariely, D. (2008). The Irrationalities of Product Pricing. *Wall Street Journal Eastern Edition* 252, R2-R2.

Ariely, D., Loewenstein, G. and Prelec, D. (2003). "Coherent Arbitrariness": Stable Demand Curves without Stable Preferences. *The Quarterly Journal of Economics* 118, 73-105.

Ariely, D., Loewenstein, G. and Prelec, D. (2006). Tom Sawyer and the construction of value. *Journal of Economic Behavior & Organization* 60, 1-10.

Bacharach, M. (1993). Variable universe games. In *Frotiers of game theory*(Ed, Binmore, K., Kirman, A., & Tani, P.). MIT Press, 255-275.

Bagwell, K. (1987). Introductory Price as a Signal of Cost in a Model of Repeat Business. *The Review of Economic Studies* 54, 365-384.

Bagwell, K. and Riordan, M. H. (1991). High and Declining Prices Signal Product Quality. *The American Economic Review* 81, 224-239.

Baltagi, B. H. (2005). *Econometrics Analysis of Panel Data.* Colchester: Wiley.

Bardsley, N. C., Cubitt, R., Loomes, G. Moffat, P., Starmer, C., Sugden, R. (2010). *Experimental Economics - Rethinking the Rules.* Oxford: Princeton University Press.

Bass, F. M. (1980). The Relationship between Diffusion Rates, Experience Curves, and Demand Elasticities for Consumer Durable Technological Innovations. *Journal of Business*, *53*(3), pp. S51-S67.

Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997). A Test of the Theory of Reference-Dependent Preferences. *The Quarterly Journal of Economics* 112, 479-505.

Berg, J., Dickhaut, J. and McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10, 122-142.

Bikhchandani, S., Hirshleifer, D. and Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *The Journal of Political Economy*, 992-1026.

Bostrom, J. (March 16, 2005). Too Many Choices May Slow Consumer Electronics Spending. *http://www.pcworld.com*

Brooks, P.; Isoni, A.; Loomes, G. and Sugden, R. (2008). Monetary Valuations in Repeated Markets: Do Prices Matter*?*, *Working Paper.*

Brown J., Hossain, T., John Morgan, J. (2010) Shrouded Attributes and Information Suppression: Evidence from the Field. *The Quarterly Journal of Economics* 125: 859-876

Buchanan, J.M. (1967). Public finance in democratic process. Chapel Hill: University of North

Carolina Press.

Cabral, L. M. B. (2010). Lock in and Switch: Asymmetric Information and New Product Diffusion. *Journal of Industrial Organization*. Forthcoming.

Camelot Group plc. (2008-9). Camelot Stakeholder Report, W. C. G. plc, 2009.

Camerer, C. (1995). Individual Choice. In Kagel, J. H. and Roth, A. E. (Eds.), *Handbook of Experimental Economics*, Princeton: Princeton University Press, pp. 587-703.

Cartwright, N. (1999). *The Dappled World - A study of the boundaries of science.* Cambridge: Cambridge University Press.

Cartwright, N. (2009). The vanity of rigour in economics - Theoretical models and Galilean experiments. In *The Experiment in the History of Economics* (Ed, Fontain, P. L., R.). New York: Routledge. pp.118-135.

Chamberlin, E. H. (1948). An Experimental Imperfect Market. *The Journal of Political Economy* 56, 95-108.

Clotfelter, C.T. (1976) Public Spending for Higher Education: An Empirical Test of Two

Hypotheses. *Public Finance* 31, 177-195.

Conlisk, J., Gerstner, E. and Sobel, J. (1984). Cyclic Pricing by a Durable Goods Monopolist. *The Quarterly Journal of Economics* 99, 489-505.

Consumer, Reports. (2008). Appliance Stores. *Consumer Reports* 73, 44-46.

Cox, J. C. and Grether, D. M. (1996). The preference reversal phenomenon: response mode, markets and incentives, *Economic Theory*, vol. 7, pp. C381–C405.

Croson, R. and Buchan, N. (1999). Gender and Culture: International Experimental Evidence from Trust Games. *The American Economic Review* 89, 386-391.

Croson, R. and Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization* 73, 122-131.

Davies, D.D. and Holt, C.A. (1993) *Experimental Economics*. Princeton: Princeton University Press,.

Dodds, W. B. and Monroe, K. B. (1985). The Effect of Brand and Price Information on Subjective Product Evaluations. *Advances in Consumer Research*, 12(1), pp. 85-90.

Dubourg, W. R., Jones-Lee, M. W. and Loomes, G. (1994). Imprecise preferences and the WTP-WTA disparity. *Journal of Risk and Uncertainty* 9, 115-133.

Ellison, G. and Ellison, S. F. (2005). Lessons about Markets from the Internet. *The Journal of Economic Perspectives* 19, 139-158.

Emery, F. (1970). Some Psychological Aspects of Price. In Bernard Taylor and Gordon Wills, eds.. *Pricing Strategy*.Princeton, N,J.: Brandon/Systems Press,98-111.

Fehr, E. and Gächter, S. (2000). Fairness and Retaliation: The Economics of Reciprocity. *The Journal of Economic Perspectives* 14, 159-181.

Fehr, E. and Tyran, J.R. "Does Money Illusion Matter? *The American Economic Review*, 2001, *91*(5), pp. 1239-62.

____. "Individual Irrationality and Aggregate Outcomes." *The Journal of Economic Perspectives*, 2005, *19*, pp. 43-66.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114, 817-868.

Fitzgerald, K. (1992). Toy Marketers Bank on "Gross" Sales. *Advertising Age* 63, 14-14.

Franciosi, R., Kujal, P., Micheltsch, R., Smith, V.L., and Gang, D. (1995). Fairness: Effect on Temporary and Equilibrium Prices in Posted-Offer Markets. *The Economic Journal* 105, 938-950.

Friedman, M. (1953). *Essays in positive economics.* Chicago: The University Chicago Press.

Garrod, L., Hviid, M., Loomes, G., and Waddams Price, C.(2008). Assessing the Effectiveness of Potential Remedies in Consumer Markets. Office of Fair Trade, London.

Gibbard, A. and Varian, H. R. (1978). Economic Models. *The Journal of Philosophy* 75, 664-677.

Giere, R. N. (1988). *Explaining Science - A Cognitive Approach.* Chicago and London: The University Chicago Press.

Green, D.; Jacowitz, D., E.; Kahneman, D. and McFadden, D. (1998) Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods. *Resources and Energy Economics*, *20*(2), pp. 85-116.

Grether, D. M. and Plott, C. R. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. *The American Economic Review* 69, 623-638.

Guala, F. (2005). *The Methodology of Experimental Economics.* Cambridge: Cambridge University Press.

Hanemann, W. M. (1991). Willingness to Pay and Willingness to Accept: How Much Can They Differ? *The American Economic Review* 81, 635-647.

Hausman, D. M. (1992). *The inexact and separate science of economics.* Cambridge: Cambridge University Press.

Helson, H. (1964). *Adaptation-Level Theory*. New York:Harper & Row.

Highhouse, S., & Yuce, P. (1996). Perspectives, perceptions, and risk-taking behavior. *Organizational Behavior and Human Decision Processes*, 65, 159–167.

Hilty, J. A. and Carnevale, P. J. (1993). Black-Hat/White-Hat Strategy in Bilateral Negotiation. *Organizational Behavior and Human Decision Processes* 55, 444-469.

Holt, C. A. and Laury S. K. (2002). Risk Aversion and Incentive Effects. American Economic 92(5), 1644-1655

Holt, C. and Sherman, R. (1990). Advertising and product quality in posted offer experiments. *Economic Inquiry* 28, 39-56.

Huck, S. and Weizsäcker, G. (1999). Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment. *Journal of Economic Psychology* 20, 699-715.

Hughes, J. (December 4, 2007). Investors Puzzled by Complex Products. *FT.com*, p. 1.

Isoni, A. (2008). Price Sensitivity, "Bad-Deal" Aversion, and the WTA/WTP Disparity. In *Working Paper.*

Andrea Isoni , A., Brooks, P., Loomes, G. and Sugden R. (2010) Do markets *reveal* preferences or *shape* them? *Working paper.*

Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *The Journal of Economic Perspectives* 5, 193-206.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 263-291.

Kerven, A. (August 8, 2001). Study: Product Variety Confusing Consumers. *Cedmagazine.com.*

Klemperer, P. (1995). Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *The Review of Economic Studies* 62, 515-539.

Knetsch, J; Tang, F.-F. and Thaler, R. (2001). The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing? *Experimental Economics, 4*(3), pp. 257-69.

Kotlikoff, L.J. and Rapson, D., (2006). Does it Pay, at the Margin, to Work and Save? Measuring Effective Marginal Taxes on Americans' Labor Supply and Saving . *NBER Working Paper Series, Vol. w12533.*

Krishnan, T. V., Bass, F. M. and Jain, D. C. (1999). Optimal Pricing Strategy for New Products. *Management Science* 45, 1650-1663.

LeBoeuf, R. A. and Shafir E.(2009). Anchoring on the Here' and 'Now' in Time and Distance Judgments, *Journal of Experimental Psychology: Learning, Memory, and Cognition,*
35 (1), 81–93.

Levin, I., P.; Schneider, S, L. and Gaeth, G, J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects." *Organizational Behavior and Human Decision Processes*, 76(2), pp. 149-88.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge MA: Harvard University Press.

Lichtenstein, D. R. and Bearden, W. O. (1989). Contextual Influences on Perceptions of Merchant-Supplied Reference Prices. *The Journal of Consumer Research*, 16(1), pp. 55-66.

Lichtenstein, S. and Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89, 46-55.

Loomes, G. (2005). Modelling the Stochastic Component of Behaviour in Experiments: Some Issues for the Interpretation of Data. *Experimental Economics* 8, 301-323.

Loomes, G., Starmer, C. and Sugden, R. (2003). Do Anomalies Disappear in Repeated Markets? *The Economic Journal* 113, C153-C166.

Lucke, T. and Hogan, J. (2007). The price of Anything Is the Forgone Alternative. In *USA Today*. 62-64.

Mäki, U. (1992). On the Method of Isolation in Economics. *Poznan Studies in the Philophy of the Science and the Humanities* 26, 316-351.

Mäki, U. (2009). *The Methodology of Positive Economics - Reflections on the Milton Friedman Legacy*. Cambridge: Cambridge University Press.

____. (2005). Models Are Experiments, Experiments Are Models. *Journal of Economic Methodology*, 2005, 12(2), pp. 303-15.

____. (2009). On the Method of Isolation in Economics. *Poznan Studies in the Philophy of the Science and the Humanities* 1992, 26, pp. 316-51.

Marquard, D. W. (1980). You Should Standardize the Predictor Variables in Your Regression Models. *Journal of the American Statistical Association 75*(369), pp. 87-91.

Maule, A. J. (1989). Positive and negative decision frames: A verbal protocol analysis of the Asian disease problem of Kahneman and Tversky. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision making* (pp. 163–180). New York: Wiley.

Morgan, J., Orzen, H. and Sefton, M. (2006). An experimental study of price dispersion. *Games and Economic Behavior* 54, 134-158.

Morgan, M. S. (2005). Experiments Versus Models: New Phenomena, Inference and Surprise. *Journal of Economic Methodology,* 12(2), pp. 317 - 29.

Morris, M. H. and Gene, M. P. (1990). *Market-Oriented Pricing.* Westport, CT: Quorum Books.

Morrison, G. C. (2000). WTP and WTA in repeated trial experiments: Learning or leading? *Journal of Economic Psychology* 21, 57-72.

Munley, Vincent G. and Greene, Kenneth V. Fiscal Illusion, the Nature of Public Goods and Equation Specification. *Public Choice*, 1978, *33*(1), pp. 95-100.

Nöth, M. and Weber, M. (2003). Information Aggregation with Random Ordering: Cascades and Overconfidence. *The Economic Journal* 113, 166-189.

Paese, Paul W. (1995) Effects of Framing on Actual Time Allocation Decisions. *Organizational Behavior and Human Decision Processe*, *61*(1), pp. 67-76.

Plott, C. R. (1991). Will Economics Become an Experimental Science? *Southern Economic Journal* 57, 901-919.

Plott, C. (1996). Rational individual behavior in markets and social choice processes: the discovered preference hypothesis, in K.Arrow, E.Colombatto, M.Perleman and C.Schmidt, eds., *Rational Foundations of Economic Behavior*, pp. C225–C50, London: Macmillan.

Putler, D. S. (1992). Incorporating Reference Price Effects into a Theory of Consumer Choice. *Marketing Science,* 11(3), pp. 287-309.

Puviani, Amilcare. Teoria Della Illusione Finanziaria, ISEDI, 1976.

Robin, E D. (1984). Matters of Life and Death: Risks vs. Benefits of Medical Care. New York: Freeman.

Rosenthal, R. W. (1980). A Model in which an Increase in the Number of Sellers Leads to a Higher Price. *Econometrica* 48, 1575-1579.

Rouse, A. (May 21, 2008). Modern Cars Confusing Consumers Says Study. *http:// www.articlealley.com/article_544293_19.html.*

Salop, S. and Stiglitz, J. (1977). Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion. *The Review of Economic Studies* 44, 493-510.

Sarin, R. K. and Weber, M. (1993). Effects of Ambiguity in Market Experiments. *Management Science* 39, 602-615.

Sashegyi, A. I., Brown, S. K. and Farrell, P. J. (2000). Application of a Generalized Random Effects Regression Model for Cluster-correlated Longitudinal Data to a School-based Smoking Prevention Trial. *Am. J. Epidemiol.* 152, 1192-1200.

Sausgruber, Rupert and Tyran, Jean-Robert. Testing the Mill Hypothesis of Fiscal Illusion. *Public Choice*, 2005, *122*(1), pp. 39-68.

Schelling, T. C. (1960). *The Strategy of Conflict.* Cambridge, Massachussets and London, England: Harvard University Press.

Schlee, E. E. (2001). Buyer experimentation and introductory pricing. *Journal of Economic Behavior & Organization* 44, 347-362.

Shafir, E., Diamond, P. and Tversky, A. (1997) Money Illusion*. *Quarterly Journal of Economics*, *112*(2), pp. 341-74.

Sher, S. and McKenzie, C. R. M. (2006). Information Leakage from Logically Equivalent Frames. *Cognition,* 101(3), pp. 467-94.

Shogren, J., F., Shin, S. Y., Hayes, D. J. and Kliebenstein, J. B. (1994). Resolving Differences in Willingness to Pay and Willingness to Accept. *The American Economic Review* 84, 255-270.

Sitzia, S. and Zizzo, D. J. (2009). Does Product Complexity Matter for Competition in Experimental Retail Markets? *Theory and Decision*, 1-18

Sitzia, S. and Zizzo, D. J. (2010). Price High and then Price Low or Price High and then Price Low? *Working Paper.*

Sopher, B. and Gigliotti, G. (1993). Intransitive cycles: Rational choice or random error? An answer based on estimation of error rates with experimental data. *Theory and Decision* 35, 311-336.

Sonsino, D., Benzion, U. and Mador, G. (2002). The Complexity Effects on Choice with Uncertainty - Experimental Evidence. *The Economic Journal* 112, 936-965.

Sonsino, D. and Mandelbaum, M. (2001). On Preference for Flexibility and Complexity Aversion: Experimental Evidence. *Theory and Decision*, 51,. 197-216.

Smith, V. L. (1976). Experimental Economics: Induced Value Theory. *The American Economic Review* 66, 274-279.

Smith, V. L. (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review* 72, 923-955.

Spiegler, R. (2006). Competition over Agents with Boundedly Rational Expectations. *Theoretical Economics* 1, 207-231.

Sugden, R. (1984). Reciprocity: The Supply of Public Goods Through Voluntary Contributions. *The Economic Journal* 94, 772-787.

Sugden, R. (1995). A Theory of Focal Points. *The Economic Journal* 105, 533-550.

Sugden, R. (2000). Credible Worlds: The Status of Theoretical Models in Economics. *Journal of Economic Methodology* 7, 1-31.

Sugden, R. and Zamarrón, I. E. (2006). Finding the key: The riddle of focal points. *Journal of Economic Psychology* 27, 609-621.

Taylor, R. (1979). Medicine out of Control: The Anatomy of a Malignant Technology. Melbourne: Sun Books.

Triano, N. (2001) Table Radio Showdown: Kloss Model One, CSW Model 88, & Bose Wave. *http://www.geek.com/hwswrev/conel/tabrads/index.htm*.

Thaler, R. (1985) Mental Accounting and Consumer Choice. *Marketing Science*, 4(3), pp. 199-214.

____. (1999) Mental Accounting Matters.*Journal of Behavioral Decision Making*, *12(3),* pp. 183-206.

Tversky, A. and Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. Science, 211(4481), pp. 453-58.

____ (1974) Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), pp. 1124-31.

____. (1986) Rational Choice and the Framing of Decisions. *The Journal of Business*, 1986, 59(4), pp. S251-S78.

Van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Clarendon Press.

Varian, H. R. (1980). A Model of Sales. *The American Economic Review* 70, 651-659.

Varian, H. R. (1997). How to Build an Economic Model in Your Spare Time. *American Economist* 41.

von Neumann, J. and Morgenstern 0. (1944). Theory of Games and Economic Behavior, Princeton University Press, Princeton, NJ.

Wagner, Richard E. "Revenue Structure, Fiscal Illusion, and Budgetary Choice." *Public Choice*, 1976, *25*(1), pp. 45-61.

Weizsäcker, G. (2008). Do We Follow Ohters When We Should? A Simple Test of Rational Expectations. In *Working Paper No. 309.* University College London.

Wilson, C. M. and Price, C. W. (2007). Do Consumers Switch to the Best Supplier? *SSRN eLibrary*.

Winer, R., S. A Reference Price Model of Brand Choice for Frequently Purchased Products. *The Journal of Consumer Research*, 1986, *13*(2), pp. 250-56.

von Neumann, J. and Morgenstern 0. (1944). Theory of Games and Economic Behavior, Princeton University Press, Princeton, NJ.

Zeckhauser, Richard, and Jeffrey Liebman (2004), Schmeduling. *unpublished paper,* Harvard
University.

Zizzo, D. (2008). Experimenter demand effects in economic experiments. *Experimental Economics* 13, 75-98.