Reference Points, Biased Beliefs, and Information Avoidance — Essays in Experimental Economics

Lukas Matthias Wenner

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Economics

University College London

April 2016

Declaration

I, Lukas Matthias Wenner, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Chapter 4 of this thesis, *More Effort with Less Pay: On Information Avoidance, Belief Design and Performance* is based on joint work with Steffen Huck and Nora Szech. All three authors contributed equally to this project.

Abstract

This thesis consists of three papers that look at different aspects of how behavioural biases affect individual decision making as well as how they influence strategic interactions. They all make use of laboratory experiments to gain new insights into behaviour that departs from the paradigm of full rationality.

Chapter 2 analyses the effects of reference prices on purchasing behaviour. Offering participants real consumption goods in the lab, I establish that comparisons between the actual purchase price and other possible price realisations affect purchase decisions. Consumers are more likely to buy goods when the price comparisons invoke feelings of making a "gain" rather than a "loss". Furthermore I show theoretically that such behaviour is in contrast to one of the most prominent papers that studies expectation-based reference points, namely Kőszegi and Rabin (2006).

Chapter 3 analyses a market setting in which sellers interact with buyers that have biased beliefs about the characteristics of the product that is being sold. I study whether such buyers can be taken advantage of by sellers through the use of specifically designed pricing structures. Comparing seller profits to the case where they interact with unbiased buyers, however, there is no evidence of exploitation. Buyers are biased in their beliefs, but otherwise sophisticated enough to not suffer from exploitative pricing strategies.

Chapter 4 focuses on the phenomenon that economic agents often prefer to actively avoid instrumental information even though this information comes free. It reports on a real-effort lab experiment where participants have the option to acquire information about their wage. Crucially, many participants engage in information avoidance and this has significant positive effects on their performance.

Acknowledgements

I am extremely grateful to my supervisor Steffen Huck for his invaluable guidance and support during my PhD studies. Throughout the years he provided me with a lot of encouragement for my ideas, many helpful suggestions for improvement, and what felt like a genuine belief in me and the quality of my work. I have learned a lot from him for my future career.

Antonio Cabrales and Syngjoo Choi also provided a lot of support, guidance and very helpful comments at various stages during the PhD. I am grateful to both for having been my second supervisors.

I would also like to thank Kai Barron, Roberta De Filippis, Antonio Guarino, Jose Guerra, Terri Kneeland, Eeva Mauring, Rani Spiegler, Nora Szech, Brian Wallace and Georg Weizsäcker for helpful comments and suggestions on the material in this thesis.

I am also very thankful to my colleagues at UCL who made my time a very enjoyable experience and I am happy that many of them also became good friends during the years we spent together.

I acknowledge financial support from the Economics department at UCL and also thank the Wissenschaftszentrum Berlin for their hospitality and the stimulating research environment during my stay in autumn 2013.

Finally, very personal thanks to my family who supported me throughout these years, and, last but certainly not least, to my girlfriend who has been an amazing source of support, inspiration, and help during the years that I spent writing the thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It is by now widely accepted that the decision making of economic agents often does not adhere to the paradigm of rationality. Instead, behaviour is often better described by accounting for systematic biases in preferences and/or beliefs. Existing work in experimental and behavioural economics has been instrumental and influential in demonstrating the inadequacy of the self-interested, expected utility-maximising, and cognitively unconstrained individual and has argued strongly in favour of alternative theories of economic behaviour. In line with this research, this thesis comprises three papers that each investigate a specific type of bias, that is, each considers a distinct manifestation of behaviour that is at odds with standard models.

However, rather than simply documenting the existence of a bias, all three papers in this thesis directly investigate the implications and explanatory power of theories of boundedly rational behaviour. They thus start from the premise that agents are biased in an ex-ante well-defined way, building on the existing evidence. Making use of the methodology of experimental economics, my contribution to the literature with this thesis can be seen along three dimensions. First, by putting existing models of bounded rationality to a direct test, I aim to provide a better understanding of which model best explains a specific behavioural phenomenon; Chapter 2 does this in the context of reference dependence. Second, I investigate the effects that behavioural biases have on market outcomes and whether these biases can be taken advantage of by unbiased market participants; Chapter 3 looks at this by analysing biased beliefs of buyers in a trade environment. Finally, I study how certain types of non-rational behaviour can have surprisingly positive effects on agents' behaviour; Chapter 4 shows how avoiding information increases performance in a simple real-effort task.

Reference Points. One of the most prominent and most widely studied behavioural biases is reference dependence. This describes the tendency of people to evaluate outcomes with respect to a reference point. Hence, a wage increase that falls short of expectations may trigger very different feelings as compared to the same increase that comes as a complete surprise. Similarly, paying the same price for a product is perceived more positively when it is less than what a consumer expected to pay rather than more. Importantly, as argued by Kahneman and Tversky (1979) who pioneered this line of research, in most cases being in a position worse than a reference outcome induces feelings of loss that are felt more negatively than feelings of gains of the same size are being felt positively. In *Chapter 2* of this thesis, I look at how being exposed to a distribution of possible prices for a consumption good affects purchasing behaviour.

Specifically, I analyse how varying the price distributions affects buying decisions through perceived gains and losses that the agents are exposed to when comparing realised prices with expected prices. Offering subjects a number of real consumption goods in a laboratory experiment, I establish that these expected prices matter for purchasing decisions. I find that subjects are more likely to buy goods when they perceive them as a "good deal", measured by the difference between the realised and the expected prices. This effect is, however, reduced for types of products that have either no, or a very salient retail value outside of the laboratory.

The main theoretical contribution that I make in this chapter is to show that while the behaviour displayed by participants intuitively is very much in line with models of reference dependence, it cannot be explained by the Kőszegi and Rabin (2006) model which has been established as the leading model for expectation-based reference dependence. As I describe in more detail in the chapter itself, the requirement that the model puts on the formation of reference points can lead to predictions that are neither theoretically appealing nor are they supported in the experimental data. This is not to say that the results do not support the idea that consumers' preferences are reference dependent. I show that the behaviour is in line with a model that treats the reference point as directly derived from the price distribution rather than endogenously formed based on the agents' planned behaviour. Broadly speaking, these findings are in line with observed pricing strategies that directly aim to manipulate consumers' reference points.

Biased Beliefs. A large body of evidence shows that consumers often fail to form accurate beliefs about products, (parts of) contracts, or their own future behaviour. Less documented are the effects that such biased beliefs have on market outcomes. An assessment, however, of how consumers are affected by their biases is of utmost importance for the design of policies that aim to make markets work well and protect consumers from exploitation. In the work presented in *Chapter 3*, I design a laboratory experiment that directly tests whether buyers with biased beliefs about the characteristics of the product for sale earn less than buyers who have correct beliefs. Through a novel method that builds on people's tendency to exhibit an "exponential growth bias" (Ensthaler et al., 2015; Stango and Zinman, 2009), combined with a simple feedback manipulation, I am able to induce biased and unbiased beliefs among otherwise comparable groups of subjects.

Models in behavioural industrial organisation predict that firms can find ways to exploit biased beliefs through contracts that are directly targeted at the belief bias, for example by offering bank accounts with high overdraft fees. In doing so, these models typically assume that buyers are strategically naive about their bias and the exploitation motive of the seller. However, as I show in this chapter, it is crucial—both theoretically and for the interpretation of my results—to account for buyers that have biased beliefs, but are strategically sophisticated about the sellers' incentives. Even though I implement a setting with monopoly sellers, they are not able to achieve higher profits when facing buyers with biased beliefs. Moreover, I find evidence confirming that this result is driven by buyers who understand the underlying adverse selection logic of the incentives for exploitation, which allows them to protect themselves from overpaying for the product. These results suggest a more nuanced view on the welfare effects of consumer biases in markets and caution against a too pessimistic view on buyers' abilities to deal with their own cognitive limitations.

Information Avoidance. Classical economic theory makes an unambiguous prediction: when agents are given the opportunity of acquiring information that is instrumental for their behaviour and at the same time for free, they should always acquire such information. A recent literature in economics as well as psychology (see Golman et al. (2016) for a comprehensive survey) documents cases in which this prediction is not upheld. To mention just one example, Oster et al. (2013) show that agents prefer not to get tested for severe diseases even though the test cost is zero or negligible compared to what is at stake.

The experimental results of *Chapter 4* (based on joint work together with Steffen Huck and Nora Szech) document evidence of such information avoidance in a different domain. In a tedious real-effort experiment, participants are paid either a low or a high piece rate for a repetitive task and can decide whether to be fully informed about it or not. Getting to know this piece-rate is costless, nevertheless about a third of participants decide not to obtain it. Importantly, we find that those who work on the task without knowing their wage, perform extremely well. They perform much better than the agents who work for a known piece-rate, and even tend to outperform the sub-group who know that they receive the high wage. Towards the end of the chapter, we present a theoretical explanation for these results. Noting that the results are clearly at odds with standard theories, we show how allowing agents to distort their beliefs as proposed by Brunnermeier and Parker (2005) together with the idea that some agents may "choke" under the pressure of a high wage can fully explain our experimental results.

Chapter 2

Expected Prices as Reference Points — Theory and Experiments[1]

2.1 Introduction

The concept of reference dependent behaviour is one of the most studied departures from expected utility. Introduced by Kahneman and Tversky (1979), the main idea is that outcomes are evaluated against a reference outcome. While earlier work concentrated on the status quo as the reference point, more recent work (most notably Kőszegi and Rabin 2006, 2007) examines the role of expectations in forming reference points. As this paper concentrates on purchasing decisions, it will focus on the way *expected prices* can serve as reference points. Hence, the main idea is that paying a price that is lower than some reference price feels like a gain whereas a price higher than a reference price feels like a

[1]A version of the work presented in this chapter has been published as Wenner (2015). I am grateful for comments from two referees and an associate editor. Some preliminary ideas regarding the experimental design and research question for this chapter have formed part of my dissertation for the MRes degree at UCL (Wenner, 2011)

loss. Along with this comes the concept of loss aversion, the observation that losses have a more negative impact than gains of equal size have a positive impact.

More specifically, consider a buying decision of a consumer who is aware of the distribution of possible prices that he faces for purchasing a good. In other words, he knows the distribution of expected prices for the product that he contemplates buying. These expected prices could be due to a market environment with price dispersion where different firms set different prices and the consumer does not know what price a specific firm sets before he visits the store. Also, one could imagine the case of a monopolist who opts to (credibly) employ a probabilistic pricing strategy. The main question that I ask is whether these expected prices affect the buying decision of the consumer. A natural way to address this is to look at cases where the price realisation (the price faced upon visiting the store) is the same, but the underlying distributions of expected prices are different.

Consider a simple example: A consumer might be in two different situations regarding the distribution of expected prices he faces. In the first situation, he expects a good to be priced either at £0.5 or £1, with equal probability. In the second situation, imagine the same consumer and the same product, but now he expects the prices to be either £1 or £2, again with equal probability. The interesting case now is when, after the resolution of uncertainty (learning the actual price) the price turns out to be £1 in both situations. Is there the possibility that this consumer behaves differently in the two situations, despite the realised price being the same? The focus of this paper will be on examining - theoretically and experimentally - the idea that the expected prices that the consumer faced before learning the realisation serve as reference points. That is, the individuals' preferences depend in some way on the expected prices in the market. A big challenge for the experimental implementation is to disentangle an effect of reference-dependent preference from other potential explanations such as inferences about quality that subjects (consumers) might draw from the prices. In many buying decisions consumers might derive some pleasure from consuming a good whose price is low compared to what they expected, but they might equally well buy because they believe that they get something of a high value, when assessing the good's value by the prices they observe. According to my results the latter effect turns out to be very relevant for my setting, and I thus design treatments that take great care in eliminating this effect, to isolate the pure effect of reference dependence.

Before embedding the situation described above in a theoretical framework, it is important to think about possible implications of the dependence of individual preferences on the expected prices. First, in models of industrial organisation, it has been shown that firms which interact with reference-dependent (and loss averse) consumers employ more rigid pricing strategies, compared to the standard model (Spiegler, 2012; Heidhues and Kőszegi, 2008). As consumers suffer a loss from facing a higher price than expected, firms prefer to set prices that are more similar for different cost levels. Put differently, the markup on the marginal cost is higher for low cost levels than for high cost levels. Second, Heidhues and Kőszegi (2014) provide a rationale based on consumer loss aversion why a monopolist may employ price distributions that consist of one "regular" price together with a continuum of "sale" prices, all below the regular price. Third, the fact that the buying decision at the realised price depends on the whole distribution of prices in a market implies that demand depends on these expected prices and therefore on supply. As outlined by Mazar et al. (2014), ignoring this dependence can lead to biased estimation of demand and welfare.

To study the idea of expectation-based reference points, it is natural to analyse the described situation within the framework of Kőszegi and Rabin (2006) - henceforth KR. Their model makes it very explicit (unlike most previous models of reference dependence) how the reference point held by an agent is formed. The key idea is that the reference point is formed by expectations about outcomes which are determined by one's anticipated behaviour in the future. KR introduce the concept of *personal equilibrium* (PE) which describes the idea that the agent's anticipated behaviour (his "plan") has to be consistent with his actual behaviour. Hence, an agent can only form plans that he knows he will be able to follow through. Applied to buying behaviour, the crucial element of KR's theory is that whenever an agent does not plan to buy at some price ex-ante, this price enters his reference point as spending nothing. Given such a plan, for an agent to be in personal equilibrium, he has to find it optimal to buy the good at the price of £1, but not at £2.

Section 2.2.2 contains the key theoretical result of this paper regarding the predictions of the KR model in this setting. Returning to the earlier example for illustration, their model makes a very strong - and possibly surprising - prediction: whereas one could intuitively think that being faced with £1 and £2 ex-ante makes the price of £1 look more favourable and therefore more attractive for buying compared to when the alternative would have been the lower price of £0.5, I will show that this intuition is not in line with the model of KR. Indeed, their model predicts the opposite effect. The reason for this lies in the nature of reference point formation mentioned before. Any individual in case (£1,£2) who finds it ex-ante optimal to buy at a price of £1 but not for £2, has the reference point "pay £1 with probability one-half, pay nothing with probability one-half". In the other situation, however, the relevant reference point is "pay £0.5 with probability one-half or pay £1 with probability one-half". But then, comparing the price of £1 to what one would have spent had the other price realised, yields the following comparison: When prices are (£0.5,£1), spending £1 feels like a (partial) loss from comparing it to £0.5. However, if prices were expected to be (£1,£2), the consumer compares £1 to the counterfactual outcome of not spending any money, which makes the feeling of a loss even larger. The higher the losses, the less willing he is to buy at the price of £1, which leads to the result stated above. Additionally, this effect is magnified by the attachment that the consumer develops from expecting to buy the good. When he expects to buy at all prices less or equal than £1, he expects to end up with the good for sure when the prices are (£0.5,£1) but only with probability one half when the prices are (£1,£2). As the consumer is loss averse, not buying when he expected to get the good with probability one, leads to a greater negative utility as compared to the case where he only expected to buy with probability one half. This makes buying in case (£0.5,£1) more likely.

In contrast to that, in section 2.2.1, I develop a simple model based on ideas in Thaler (1985) that gives rise to more intuitive predictions. Such a model, which I will call *good deal model*, simply compares the realised price to some measure of the distribution of expected prices, for example the average expected price, or (in the case of only two prices) the non-realised price. As it ignores the KR idea that the expected behaviour at the other prices matters for the reference point, it predicts that consumers who face the price of £1 in the situation where prices are (£1,£2), perceive it as a good deal, whereas when £2 is replaced by £0.5, they perceive it as a rip-off. Hence, they are more likely to buy in the former situation, opposite to what KR predict.

I furthermore show in section 2.2.3 that for settings with more than two prices one obtains a similar discrepancy in the theoretical predictions. This shows that this effect is not restricted to the setting with two prices, but rather is a fairly general result.

Therefore, on the one hand my paper is a specifically designed experimental test for the KR model applied to a consumer framework, where the KR model's predictions are specific and distinguishable from a large class of alternative explanations (including a standard reference independent model). Recent work has applied the KR model to experimental settings of effort provision and endowment effects but apart from the work by Karle et al. (2015), no experiment applies their model to a consumer purchase decision. On the other hand, one can interpret the experimental design more broadly as a test for a distributional dependence of consumer behaviour on expected prices. To my knowledge only the work by Mazar et al. (2014) specifically addresses this question, but contrary to my design, they change the distribution of expected prices in a way that leaves the support fixed across treatments. These two papers will be discussed in more detail in section 2.2.3 as they provide useful empirical results to compare my theoretical predictions against.

More broadly, the experiment adds to a growing experimental literature that tries to assess the relevance of expectation-based reference points. Most work in the literature finds evidence for behaviour predicted by Kőszegi and Rabin (2006, 2007). Abeler et al. (2011) look at effort provision and find that manipulating the expected payment of a repetitive task affects individuals' effort provision. Participants are either paid a fixed amount or a piece rate (with equal probability). In accordance with loss aversion and expectation-based reference points increasing the fixed amount increases effort as to minimise the differences in payments.[2] Ericson and Fuster (2011) use a variant of the classical mug experiment (Knetsch, 1989) where they endow subjects with a lottery

[2]See Gill and Prowse (2012) for similar results in a sequential-move tournament.

about whether they will be able to trade the mug they are given for a pen or not. They find that experimentally increasing the probability of trade increases the likelihood of trade. This is in accordance with KR since a higher ex-ante expectation of being able to obtain the pen shifts the (expectation-based) reference point and - like an endowment effect - makes it harder for an individual to give up the expected ownership. Note that Heffetz and List (2013) cast some doubt on this finding when - in a setting similar to Ericson and Fuster (2011) - they randomise the initial assignment of the mug or the pen and do not find an effect in line with KR and thus question the importance of expectations in these kind of experiments. Non-laboratory evidence is provided by Crawford and Meng (2011) and Fehr and Goette (2007) who show that the labour supply decisions of cab drivers and bike messengers, respectively, can be interpreted as being driven by reference-dependent preferences regarding wage expectations.

The experimental setting used in this paper closely follows the situation described in the introductory example. In a first set of experiments (the goods are a chocolate bar, a pen, and a notepad), I find an effect supportive of the good deal model. Looking at the behaviour at the price that is common across treatments (£1 in the example) subjects are more likely to buy if the other possible price is higher (£2) rather than lower (£0.5). However, a potential confound of these results is that subjects might use the distribution of possible prices to make inferences about the market value and thus exposing subjects to a distribution of possible prices might have induced both reference points as well as lead subjects to infer a different retail value of the good outside of the experiment. This would bias the results in favour of the good deal model and against the predictions in KR. As it turns out, when controlling for these effects by making subjects aware of the price distribution of both treatments (experiment 2), informing them about the production cost of a custom-made chocolate bar (experiment 3), or using a good that has a clear retail value (amazon voucher with a known redemption value, experiments 2 and 3), the results do not show a significant dependence of the buying behaviour on the non-realised price. This suggests, taking all results together, that a change in the distribution of possible prices affects the subjects' perception of the good's retail value (outside of the experiment) more than it directly causes feelings of elation or disappointment when evaluating the actual price draw.

The finding that inferences about the retail price affect behaviour in the experiment, is reminiscent of Plott and Zeiler (2007), who add additional experimental controls to the experiment in Knetsch (1989), addressing the issue that instructions which put strong emphasis on entitlement to the mug once allocated to subjects, might lead to an increase in valuation of the mug. They find that these additional controls eliminate the endowment effect. Hence, my results reinforce the need for special emphasis and care in attributing experimental findings to reference-dependent preferences and loss aversion.

2.2 Theory

As it will lend itself naturally to the experimental implementation, I will consider the following setup. A consumer is assumed to have expectations about the prices he faces for a good and their probabilities. In the simple setting that I am looking at, I concentrate on the case where the consumer knows that there are only two possible prices that can realise, both equally likely. In section 2.2.3, I will also consider more general settings with more than two possible prices. It seems a natural assumption that in purchase decisions, consumers will have expectations about the prices they possibly face. These expectations can be formed through, for example, past buying experience, forecasts about future prices, inferences about firms' pricing strategies, or word-of-mouth via other consumers. Since the theoretical part does not model the formation of these price expectations, they are best thought about as a combination of these factors. Depending on the situation different sources may receive more weight. In line with the experimental setting and the role of stochastic reference points in KR, the theoretical analysis always focuses on a distribution of expected prices rather than one single reference price. I believe that in many cases (a consumer might expect a sale with a certain probability; a consumer knows a past price from his own experience and also learns the price that a friend recently paid) it is plausible to think that all the information that consumers obtain to form expectations, is aggregated into a distribution of expected prices.[3] However, there are clearly situations when one piece of information is particularly salient and we would then expect there to be only a single reference price.

Formally, for prices $p_H > p_M > p_L > 0$, I will analyse consumer behaviour in two different situations. Either the consumer expects the possible prices for the good to be $(\frac{1}{2}, p_L; \frac{1}{2}, p_M)$, that is, p_L and p_M with equal probability of one-half. Call this case *LM.* Alternatively, the consumer faces $(\frac{1}{2}, p_M; \frac{1}{2}, p_H)$. Call this case *MH.* The main interest now lies in the buying decision of a consumer who is faced with a realised price of p_M across the two situations.

[3]The implicit assumption made here is that consumers learn fully learn about the distribution of expected prices at one given point in time. In Kőszegi and Rabin (2009), the authors analyse a dynamic version of the static concept introduced in Kőszegi and Rabin (2006). This model could potentially be useful to study cases where different pieces of information regarding the price distribution in a market arrive over time, as it explicitly models the gain-loss utility from changes in beliefs about future consumption. For my experimental setting, this type of news seems unlikely to play a role which is why I focus on the static version.

2.2.1 The "Good Deal" Model

Assume a consumer derives utility u from consuming the good. His net utility from buying is given by the difference between u and the price p_i he has to pay, and an additional component that evaluates the purchase as to whether buying at p_i is seen as a "good deal" from the viewpoint of the consumer. In order to make such an assessment, the consumer compares the realised price p_i to a reference price $\tilde{p}$. His overall utility from buying is given by:

$$u - p_i + \gamma(\tilde{p} - p_i) \tag{2.1}$$

with $\gamma(x) = \gamma_L \mathbb{1}_{\{x<0\}} x + \gamma_G \mathbb{1}_{\{x>0\}} x$ and $\gamma_L > \gamma_G > 0$, capturing loss aversion. Further, assume that not buying yields utility of zero. The above formulation is an often used way of dealing with reference prices. What I call the good deal model can therefore easily be seen as a model of reference pricing as proposed by, for example, Thaler (1985). In this section where I am looking at the case where there are two possible prices, the reference price $\tilde{p}$ is taken to be the *non-realised price*. When extending the model to more than two prices, it seems natural to take $\tilde{p}$ as the average expected price. While the qualitative predictions are identical, I believe, however, that for two prices the comparison to the other price is a more realistic description of the cognitive process present. The reason is that the non-realised price seems an obvious candidate to compare the current price to.[4]

While this formulation looks similar to the model of "bad-deal aversion"

[4] As the main role of the good deal model is to act as a point of comparison to KR predictions for the specific experimental setting that I present, it does not aspire to be a fully general model that allows predicting behaviour in more general settings. One reason is that I believe that what best determines $\tilde{p}$ can potentially be context specific. For example, when comparing the predictions of the good deal model with KR in section 2.2.3, I use the average of the given price distribution as $\tilde{p}$ because for continuous distributions this arguably captures something more relevant than a particular price in the support, though constructing $\tilde{p}$ from other statistics of the price distribution is clearly also a possibility.

by Isoni (2011), and indeed shares some of the ideas expressed therein, the interpretation of the reference point is different. Unlike Isoni, my specification assumes that the reference price is derived directly from the distribution of possible prices in the market, whereas he defines the reference price as the price consumers expect to trade at and explicitly rules out the case that it is obtained by calculating the average of the price distribution (Isoni, 2011, fn. 9). Modelling the reference point as the price at which one expects to trade has a flavour of personal equilibrium (i.e. dependence on the planned action) to it that I specifically do not want to assume.[5] To derive predictions of this model, equation (2.1) then can be rewritten for the decision whether to buy or not to buy the good at a price of p_M, depending on the other possible price, p_X, with $X \in \{L, H\}$:

$$u - p_M + \gamma(p_X - p_M) \geq 0, \tag{2.2}$$

Thus, when in situation *MH*, the agent gets additional positive utility from comparing p_M to p_H, he thinks that he is making a good deal. However, if in case *LM* where prices were expected to be either p_L or p_M, he perceives the price of p_M as a rip-off, which is detrimental to his overall utility. Thus, for $u - p_M \in [-\gamma_G(p_H - p_M), \gamma_L(p_M - p_L))$ the consumer only buys at p_M if he expected the prices to be either p_M or p_H. Formally:

Proposition 2.1. *For any agent with preferences as specified by the good deal model and prices $p_H > p_M > p_L > 0$, there exists a range of intrinsic valuations u such that a consumer buys at p_M in case* LM *but not in case* MH*. For any*

[5]The specification of consumer preferences in the good deal model is also related to Spiegler (2012). He assumes that the reference price is sampling-based, resembling, for example, market experience. In his model, for each consumer the reference price is randomly drawn from a distribution consisting of all prices in the market. He considers the case where a consumer only experiences losses from unexpectedly high prices, but no gains, i.e $\gamma_G = 0$ in the above notation.

u outside this range, the agent's buying behaviour at p_M is the same for both distributions of expected prices.

Proof. In text. □

This simple setting presented can easily be extended to encompass more general settings within the realm of buying decisions. For example, consider a consumer buying more than one unit of a good. Let $v(q)$ denote his valuation for q units and $T(q)$ the total price for q units. In the same manner as before, I can then define $\tilde{T}(q)$ as the reference price for buying q units and the total utility from buying q units is then given by $v(q)-T(q)+\gamma(\tilde{T}(q)-T(q))$.[6] Hence, the good deal model can handle cases where a consumer buys more than one unit and might face non-linear pricing schedules, as commonly observed, for example, for mobile phone contracts or energy usage.

2.2.2 The Kőszegi and Rabin (2006) Model

In this section, I will introduce the model by KR and then derive its predictions for the setting above. Motivated by the ideas of prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991), but also the lack therein of a reference point that is specified by the model, KR posit that the reference point is given by recent expectations. An agent derives utility from "consumption utility", but also from a psychological component that evaluates the actual outcome with respect to the reference point. Utility is positively affected ("gain") if the actual outcome is better than the reference outcome, but negatively if the actual outcome is worse ("loss"). Importantly, this reference point can often be stochastic. If there is ex-ante uncertainty about the outcome, this will be reflected in the reference point as each potential reference

[6]From this, we obtain equation (2.1) by setting $\tilde{T}(p) = \tilde{p}q$, $T(q) = pq$ and defining $v(1) = u$ and $v(0) = 0$.

outcome is evaluated with its probability of realisation. Moreover, KR suggest a separation of gains and losses for different dimensions. For example, in a purchase decision, there is a "money dimension" that evaluates how the actual price compares to the expectation of how much to pay, but also a "good dimension" evaluating actual ownership of the good with respect to the expectation of whether one expected to obtain the good or not.

I will state the utility function of an agent in its general form in the case where an agent can buy one unit of a product ($b = 1$ if he does, $b = 0$ if he does not) which gives him "intrinsic utility" u. He faces initial uncertainty about the actual buying price, but knows that p_i realises with probability q_i. Then, the reference point is given by $(p^r, u^r) = (p_1^r, ..., p_N^r, u_1^r, ..., u_N^r)$. That is, for each possible price realisation p_i, the reference point specifies whether the agent buys the good, in which case $p_i^r = p_i$ and $u_i^r = u$, or not, in which case $p_i^r = u_i^r = 0$.

$$U(p, b|p^r, u^r) = (u - p)b + \sum_{i=1}^{N} q_i \mu_m(-bp + p_i^r) + \sum_{i=1}^{N} q_i \mu_g(bu - u_i^r) \tag{2.3}$$

Here, the first term is the "classic" (net) consumption utility; the utility from buying the good minus the price that is to be paid. The function $\mu_k(\cdot)$, with $k \in \{m, g\}$ describes gain-loss utility in the money (m) and good (g) dimension. In most applications of KR, it is assumed that $\mu_k(x) = \eta_k x$ if $x \geq 0$, and $\mu_k(x) = \eta_k \lambda_k x$ if $x < 0$. Losses are multiplied by $\lambda_k > 1$, capturing the idea that losses loom larger than equal sized gains. $\eta_k > 0$ measures the relative weight of the gain-loss component in dimension k. I will use this linear specification throughout the main text. However, when proving the result of Proposition 2.2 in the appendix, I will use the more general form of $\mu_k(\cdot)$, satisfying assumptions first stated by Bowman et al. (1999) and also employed by KR. It turns out that it is possible to allow for this more general form, but that an additional restriction on $\mu_m(\cdot)$ is needed. More specifically, the degree of diminishing sensitivity that this more general from exhibits as compared to the linear specification cannot be too large. I will return to this issue below when discussing the workings of the model in more detail. Furthermore, I, unlike KR, allow the gain-loss utility function $\mu_k(\cdot)$ to be different across dimensions. Thus, an agent may feel high losses when paying a higher price than expected, but may be only very little affected by not getting a product that he expected to get, or vice versa.

In what follows, I will state and highlight the key intuition of the main result of this theoretical part, namely that, contrary to the result emerging from the good deal model, KR predict the opposite buying behaviour at the price of p_M:

Proposition 2.2. *For any consumer with KR preferences, linear gain-loss utility, and prices $p_H > p_M > p_L > 0$ with $p_H - p_M \geq p_M - p_L$ and $3p_L \geq p_M$, there exists a range of intrinsic valuations u such that it is a Personal Equilibrium for a consumer to buy at p_M in case* LM *but not in case* MH*. For any u outside this range, the agent's buying behaviour at p_M is the same for both distributions of expected prices.*

Proof. See appendix. □

Due to the nature of personal equilibrium, there are a number of steps necessary to derive this result. I will relegate most of the technical steps into the appendix and in the main text only focus on the steps necessary to understand the intuition behind the proposition.

The concept of personal equilibrium is based on the idea that, given a plan at which prices to buy, it has to be optimal to follow through this plan for each possible price realisation. Therefore, it can often happen that for the same distribution of expected prices, there exists more than one plan that is optimal to follow through, i.e. we have multiple PE. In these cases KR suggest applying the concept of *preferred personal equilibrium* (PPE) as a selection criterion. This amounts to selecting the PE with the highest ex-ante utility.

Consider case LM and the personal equilibrium "buy at p_L and buy at p_M". In this case this is the only PE that implies buying at p_M because "not buy at p_L and buy at p_M" can never be a PE.[7] "Buy at p_L and buy at p_M" therefore is a PE if the agent's utility from buying at p_M and p_L - given this plan as the reference point - is higher than not buying, given the same reference point. That is, in effect we are checking a non-deviation condition from the plan. The reference point is then given - in the money dimension - by the expectation to either pay p_L or p_M with equal probability of one-half, and - in the good dimension - by the expectation to get the good with probability one. One can check that the deviation is more likely to occur when the realised price is p_M, thus the condition can be stated as:

$$\begin{aligned} & U(p = p_M, b = 1 | p^r = (p_L, p_M), u^r = (u, u)) \\ & \qquad \geq \quad U(p = p_M, b = 0 | p^r = (p_L, p_M), u^r = (u, u)) \\ \Leftrightarrow \quad & u - p_M - \frac{1}{2}\eta_m \lambda_m (p_M - p_L) \geq 0 + \frac{1}{2}\eta_m (p_M + p_L) - \eta_g \lambda_g u \qquad (2.4) \end{aligned}$$

This follows directly from the more general form in (2.3). Buying at p_M generates a loss in the money dimension from comparing p_M to p_L which was the expected price with probability one-half. Since the agent buys the good and expected to do so at every price, he experiences neither losses nor gains in

[7] It is straightforward to show that there can never exist a PE that involves buying at a price p, but not at a price $p' < p$. It therefore suffices to consider the three PE, "never buy", "only buy at low price", and "always buy".

the good dimension. The RHS of this condition denotes the utility of the agent in case he deviates from the plan and decides not to buy at p_M. He obtains a consumption utility of zero but registers a gain from not spending the money which he expected to spend under the plan, and a loss from unexpectedly not getting the good.

Now, consider the case *MH* and the possible personal equilibria that involve buying at the price of p_M. Here, the possible cases are either to always buy, or to only buy at the lower price of p_M. In the latter case, buying at p_M and not buying at p_H implies that the agent's reference point in the money dimension is given by "pay p_M with probability one-half, pay nothing with probability one-half", and in the good dimension "get the good with probability one-half". Especially, note that since the agent does not plan to buy at the high price of p_H, this price does not enter his reference point, rather he expects to spend nothing in this case. This is an important difference to the good deal model. There, it was irrelevant what the agent would do at the price of p_H, he would still feel elated from comparing an actual price of p_M to the price of p_H. In KR things are different, and comparing p_M to the counterfactual outcome of spending nothing when the price of p_H is realised, actually feels like a loss. The condition for the PE "only buy at p_M" looks as follows.[8]

$$\begin{aligned} & U(p = p_M, b = 1 | p^r = (p_M, 0), u^r = (u, 0)) \\ & \qquad \geq \quad U(p = p_M, b = 0 | p^r = (p_M, 0), u^r = (u, 0)) \\ \Leftrightarrow \quad & u - p_M + \frac{1}{2}\eta_g u - \frac{1}{2}\eta_m \lambda_m (p_M - 0) \geq 0 + \frac{1}{2}\eta_m (p_M + 0) - \frac{1}{2}\eta_g \lambda_g u \end{aligned} \tag{2.5}$$

[8] Formally, there is another condition, namely that upon realisation of p_H, the agent finds it optimal not to buy. This condition puts an upper bound on u which is not relevant for the comparison here as it only further constrains the existence of equilibria where buying at p_M is optimal in case *MH*.

In the good dimension buying feels like partial gain because with probability one-half the agent expected not to get the good. In the money dimension the agent faces the loss from comparing p_M to zero, as discussed above. Deviating from the plan yields a consumption utility of zero plus a gain in the money dimension from not paying the price as prescribed under the plan, but facing a partial loss from not getting the good.

Rearranging the two conditions yields

$$u - p_M - \frac{1}{2}\eta_m\lambda_m(p_M - p_L) - \frac{1}{2}\eta_m(p_M + p_L) + \eta_g\lambda_g u \geq 0 \tag{2.4'}$$

$$u - p_M - \frac{1}{2}\eta_m\lambda_m(p_M - 0) - \frac{1}{2}\eta_m(p_M + 0) + \frac{1}{2}\eta_g(\lambda_g + 1)u \geq 0. \tag{2.5'}$$

As $p_L > 0$ and $\eta_g\lambda_g > \frac{1}{2}\eta_g(\lambda_g + 1)$, the LHS in (2.4') is larger than the LHS in (2.5') for any value of $\eta_m, \eta_g > 0$ and $\lambda_m, \lambda_g > 1$. Hence, there exists a range of intrinsic valuations u such that it is a personal equilibrium for an agent to buy at the price of p_M in case *LM* but not in case *MH*. Intuitively, this result rests on two forces. First, consider the money dimension. Here, in case *LM* the agent compares the realised price of p_M to the counterfactual price of p_L which is also part of the reference point. The loss generated by this comparison, however, is smaller than the loss from comparing p_M to a price of zero. This is the relevant comparison in the case where the other possible price is p_H but the agent does not plan to buy at this price. Due to loss aversion, this effect dominates the effect that in case *LM* the agent also receives a larger gain from deviating from the plan. In case *LM* he expected to spend $\frac{1}{2}(p_M + p_L)$, whereas in *MH* he expected to spend $\frac{1}{2}p_M$. It is here where we can see that the proposition does not hold for all gain-loss functions $\mu_m(\cdot)$. What we require is that the condition $\mu_m(p_L - p_M) - \mu_m(-p_M) > \mu_m(p_L)$ holds. In the linear case, it is straightforward to see that this is ensured by loss aversion (gain-loss utility has a "kink" at zero), but it might fail in cases where the value function becomes sufficiently flat further away from zero. If in the relevant region the degree of diminishing sensitivity is large, a loss of p_M does not feel that much worse than a loss of only $p_M - p_L$.[9] However, it seems plausible to assume that for the small amounts of money involved in the experiment, the condition will hold.

Furthermore, the effect in the money dimension is enhanced by a so-called *attachment effect* which operates in the good dimension. In the first case with prices p_L and p_H, the ex-ante expectation is to get the good for sure. When prices are p_M and p_H, however, and the agent does not plan to buy at p_H, the ex-ante likelihood of getting the good is only one-half. Thus, if the agent now were to deviate from his plan to buy he would incur a loss in the good dimension as high as his initial expectation of getting the good. Put differently, as he got more attached to obtaining the good, deviating and not buying is more painful than in the case where he expected to get the good with a 50 percent chance in the first place only. It is worth highlighting that these two effects work separately in the two different dimensions and also work in the same direction. Thus, it is clear that by allowing the agent to have different degrees of loss aversion (λ_k) and relative importance of the two dimensions (η_k) - which is something KR do not do - the result is not affected. Moreover, one could even consider an agent that only experiences gains and losses in money, similar to the good deal model, and the Proposition would still hold. This might describe "every day" purchase decisions where it might be harder

[9] A different way of looking at this condition is to see that it can be rewritten as

$$\mu_m(p_L - p_M) + \mu_m(p_M - p_L) - \mu_m(p_M) - \mu_m(-p_M) > \mu_m(p_M - p_L) + \mu_m(p_L) - \mu_m(p_M)$$

The RHS of this equation is positive due to the concavity of $\mu_m(\cdot)$ in the gain domain, whereas the LHS is also positive due to loss aversion (this is assumption A2 which captures loss aversion for large stakes). Hence, the condition might fail if diminishing sensitivity is a stronger force than loss aversion.

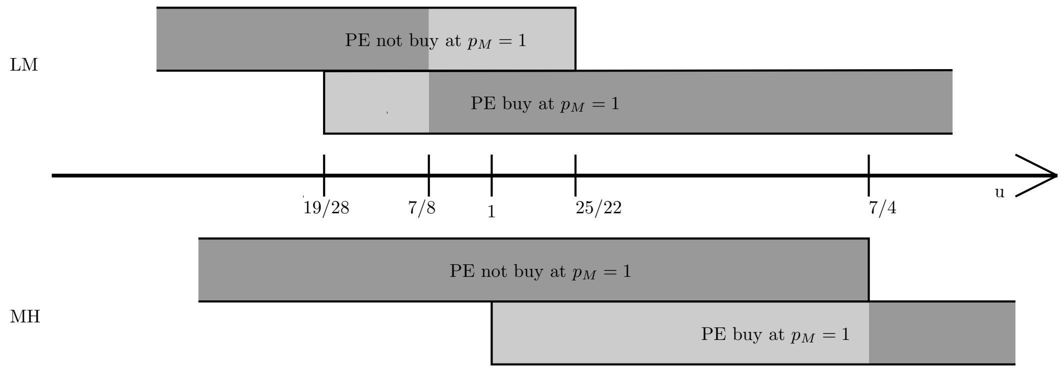
to justify that an agent actually becomes attached to a good while forming a plan.

Intuitively, it might seem as if the result in Proposition 2.2 is mainly driven by the fact that in case *MH* having a plan of only buying at p_M is detrimental to the agent because by not buying at p_H he suffers from the fact that he is not comparing p_M to p_H and not realising the resulting gain. Thus, a natural case to consider is whether the agent might find it worthwhile to form a plan that involves buying at p_M and p_H. In this case, however, for such a plan to be consistent, the agent must find it optimal to buy at p_H as well. It turns out that there could exist cases where an agent would even find it beneficial to buy at p_M and p_H in case *MH* although he would not buy at p_M in case *LM*. This is where one of the conditions mentioned in Proposition 2.2 comes into play. Under the (sufficient) condition that $p_H - p_M \geq p_M - p_L$, such a case can never occur.[10] Intuitively, what this condition ensures is that the loss from comparing p_M to p_L in case *LM* is not too large. If p_M and p_L were far apart, but p_H and p_M very similar, the loss from comparing p_H to p_M would be small enough to make buying at p_H tempting (provided the consumer values gains and losses in money sufficiently).

The appendix takes these considerations further and establishes that for cases in which the PE combination that is driving the result - always buying in case *LM*, never buying in case *MH* - is not unique, applying PPE as the selection criterion does not rule out the desired combination. I further formally establish that the opposite behaviour to Proposition 2.2, the prediction of the good deal model (Proposition 2.1), can not be rationalised by the KR model.[11]

[10]Note that the violation of the condition $p_H - p_M \geq p_M - p_L$, does not necessarily entail the reversal of Proposition 2.2. It rather is the case that depending on the parameter values η and λ both effects (as in Propositions 2.1 and 2.2) could exist. Hence, as under this condition Proposition 2.2 is valid for all parameter values, the predictive power is strongest.

[11]In the appendix, I also relax the assumption of equal probabilities for the two prices. In the latter case this again amounts to having sufficient conditions on prices and probabilities.



This figure shows which Personal Equilibrium is chosen for each of the cases, LM and MH, depending on the value of u. The darker shaded areas indicate the PPE, i.e. the PE with the higher ex-ante utility in case of multiplicity. For $u < \frac{7}{8}$ the individual never buys at $p_M = 1$, for $\frac{7}{8} \leq u < \frac{7}{4}$, she buys only in case LM, and for $u \geq \frac{7}{4}$, she buys in both cases. For this figure, I assume that $p_H = 2$, $p_M = 1$, $p_L = 0.5$, $\lambda_g = \lambda_m = 2.5$, $\eta_g = \eta_m = 1$.

Figure 2.1: Graphical Illustration of Proposition 2.2

Figure 2.1 provides a graphical illustration. I assume $\lambda_g = \lambda_m = 2.5$, $\eta_g = \eta_m = 1$ as well as the prices used for the chocolate bar in experiments 1 and 2, $p_H = 2$, $p_M = 1$, $p_L = 0.5$. Using equations (2.4') and (2.5'), buying at p_M in LM is a PE for $u \geq \frac{19}{28} \approx 0.678$, but in case MH only for $u \geq 1$. Moreover, we see that not buying at p_M is a PE for $u \leq \frac{25}{22} \approx 1.14$ in LM and for $u \leq \frac{7}{4}$ in MH. This is the multiplicity of PE described earlier. The shaded areas in the figure show the use of PPE as the selection criterion. From this it can be seen that $\frac{7}{8} < u < \frac{7}{4}$, constitutes the range of values for u for which the consumer with the utility function parameters as above only buys at p_M if the other price is p_L.

In some parts of the experiment, subjects have to decide at both prices whether to buy or not, that is, before they know which price will turn out to be the actual price, and therefore are exposed to risk when making their buying decisions. Hence, it will be useful to relate the above result to the concepts of

As the experiment only deals with the case of probabilities of one-half, I will not pursue this issue here further, but it should be noted that the above result is not a mere artifact of this specific choice of probabilities.

"unacclimating personal equilibrium (UPE)" or "choice-acclimating personal equilibrium (CPE)", both introduced in Kőszegi and Rabin (2007) to study risk preferences. In CPE the reference point adjusts to the choice of a lottery such that when evaluating the expected utility of a lottery, the lottery itself acts as the reference point. This implies that analysing the setting with the CPE concept is equivalent to the second part of the proof where I establish that no matter what PE combination exists, a comparison of ex-ante utilities rules out the behaviour predicted by the good deal model, and the predictions are thus identical. UPE looks at cases where the reference point does not adjust to the choice of lottery. It can be shown that the conditions needed for UPE are identical to the ones analysed previously. To see this, consider the UPE "buy at p_L, not buy at p_M". The difference to the analysis above is that we now focus on the ex-ante utility associated with this plan and compare it to a deviation from this plan, keeping the initial plan as the reference point. If we, for example, look at a deviation to "never buy", it is clear that since the reference point does not change, nothing changes in the p_M-state and the resulting condition is identical to the case where p_L is the realised price and "buy at p_L, not buy at p_M" the PE under consideration.

In many settings, the predictions made by KR are similar to those made by models of disappointment aversion such as Loomes and Sugden (1986), Bell (1985), or Gul (1991). This is for example, the case for the settings in Abeler et al. (2011) and Gill and Prowse (2012).[12] Models of disappointment aversion typically assume that the reference point is given by the certainty equivalent of the lottery rather than the full distribution as in KR. Considering, as in the previous paragraph, those situations where agents still face price uncertainty, in my setting disappointment aversion, for example formalised as in Loomes and

[12] See, however, Sprenger (2015) for a setting where disappointment aversion and the UPE concept described above can be distinguished.

Sugden (1986), predicts the same behaviour as KR, once I allow the functional form that captures disappointment/elation to share the properties of the gain-loss utility. In the setting with two prices, the resulting expected utilities for a given buying strategy can be expressed in a very similar way to the way it is done using the concept of CPE. However, it should be noted that situations where the buying decision is made after the price realisation cannot be meaningfully analysed by models of disappointment aversion as the choice is essentially riskless.

Finally, I consider the case where intrinsic utility over money is concave rather than linear as assumed so far. Following KR, I denote this function by $m(\cdot)$. While it seems reasonable to assume that a consumer's utility function is approximately linear when the amounts of money involved are as small as in the present experiment, it should be worth noting what happens to the result in Proposition 2.2 when relaxing the linearity assumption. I further denote by w the endowment of a consumer. This can be both thought of in a general specification as her level of wealth, or, more narrowly defined in the context of the experiment, as the amount of money that the subjects can spend in the experiment. For most of the equations derived above and in the appendix, allowing for concave intrinsic utility simply means replacing p_i by $-m(w - p_i)$ and $p_i - p_j$ by $m(w - p_j) - m(w - p_i)$. Since $m(\cdot)$ is strictly increasing, most of the statements still hold. However, it turns out that one of the two conditions stated in Proposition 2.2 is affected by this change. Instead of $3p_L \geq p_M$, prices now have to satisfy $2m(w) - 2m(w - p_L) \geq m(w - p_L) - m(w - p_M)$ which, for a sufficiently concave $m(\cdot)$ is not implied by $3p_L \geq p_M$. Intuitively, if $m(\cdot)$ is very flat for large values, $m(w)$ can be close to $m(w - p_L)$ even if w is much larger than $w - p_L$. As shown in the appendix, this condition is needed to rule out that there exists a consumer who derives the highest ex-ante utility from buying at both prices in MH as well as from buying only at the low price in *LM*. If paying 0 or p_L does not change utility much (due to the extreme concavity in this region) this increases the attractiveness of only buying at p_L in *LM*. Hence consumers exhibiting this type of concavity in $m(\cdot)$ might - provided the corresponding personal equilibria exist - not buy at p_M when in case *LM*, but buy at p_M in case MH which would then be contradicting Proposition 2.2. Hence, it is not possible to allow for any concave $m(\cdot)$; we need it to satisfy the condition stated above.[13]

2.2.3 Comparing the Two Models

Given the stark difference in the predictions of the two models despite them sharing the same basic intuition, it is important to see whether these results extend to other settings. As I will show in the following, the mechanism of the personal equilibrium concept (prices at which the consumer does not buy do not enter the reference point) as the driver of the different predictions between the good deal model and KR, applies to settings with different price distributions in a similar manner. However, it should be clear that the two models do not always make opposite predictions. As a simple example, if we consider a variant of the setting with two prices, now with the two situations being *LM* and *L'M* where in the latter we replace p_L by $p_{L'} < p_L$, then both models predict (though KR requires some assumption on the parameter values when the PE is not unique) that consumers would be less likely to buy at p_M in case $L'M$ than in LM because in both cases the loss in the money dimension from buying at p_M increases in the difference of the two possible prices.

Similarly, we can show that the results obtained by Karle et al. (2015) are

[13]The other condition, $p_H - p_M \geq p_M - p_L$, becomes more slack if it is replaced by $m(w - p_M) - m(w - p_H) \geq m(w - p_L) - m(w - p_M)$, it therefore does not put additional restrictions on $m(\cdot)$.

not only consistent with KR, as the authors demonstrate, but also with the good deal model. In their paper, the authors set up an experiment in which subjects have to choose between two sandwiches (which they can taste beforehand) that differ in their relative prices. Depending on a random draw, one sandwich will be 1 Euro cheaper than the other. The authors show that PE-behaviour as in KR predicts that more loss averse subjects are more likely to choose the cheaper sandwich (in cases when the cheaper sandwich is the one they like less). They confirm this prediction in the data, though find that the data fits a naive-expectations model only slightly worse than an optimal-expectations model that relies on PE. It turns out that (perhaps not too surprisingly, given that both the naive- and optimal-expectation case that the authors look at make the same qualitative prediction) the good deal model can, too, rationalise their evidence: When an agent has to choose between (as in the Karle et al. (2015) experiment) a ham sandwich (which she likes better) and a cheese sandwich (which is 1 Euro cheaper), the good deal model says that the utility from buying the ham sandwich entails a loss of 1 Euro compared to the other possible (cheaper) price, whereas buying the cheese sandwich leads to an additional gain of 1 Euro. The more the loss of 1 affects the consumer negatively (through γ_L) than the gain of 1 (through γ_G) affects her positively, the more likely it is that she chooses the cheaper sandwich (and avoids the loss).[14]

Nevertheless, to illustrate that the tension between the models fleshed out in sections 2.2.1 and 2.2.2, is a relatively general effect, I will look at two examples of price distributions with more than two prices. In the first, possible prices are uniformly distributed over an interval $[a, b]$ and I will analyse the

[14]Formally, buying the ham sandwich yields utility of $u_{ham} - p_{ham} - \gamma_L$ whereas buying the cheese sandwich yields utility of $u_{cheese} - p_{cheese} + \gamma_G$. Buying the cheese sandwich is preferred if $(u_{cheese} - u_{ham}) - (p_{cheese} - p_{ham}) + \gamma_G + \gamma_L \geq 0$, and thus consumers with a higher loss aversion parameter γ_L are more likely to choose the less liked cheese sandwich.

effect of an increase in b on the highest price at which a consumer is willing to buy the good, denoted by $\hat{p} \in [a, b]$. In the second example, I will analyse the setting chosen by Mazar et al. (2014) where the support of the distribution remains unchanged between the two treatments, but half of the mass is concentrated either on the left or on the right end of the distribution.

If $p \sim U[a, b]$, the good deal model predicts that the highest price the consumer buys the good at is the $\hat{p}$ that solves $u - \hat{p} + \gamma(\frac{a+b}{2} - \hat{p}) = 0$, because the average price serves as the reference point. If we increase the upper bound b, the maximum price that a consumer is willing to pay increases. Any price p feels like a better deal the more likely it would have been to obtain a price higher than p. As it turns out, and this chimes well with the result obtained above, KR predict the opposite effect. As b increases, the price $\hat{p}$ which characterises the personal equilibrium to buy at $\hat{p}$ and all lower prices, decreases.[15] The intuitive idea is as follows: If we take $\hat{p}$ as the PE for the distribution $U[a, b]$ and then increase b to b', the likelihood that a price is realised at which the consumer buys, decreases. This means that (i) in the good dimension, there is less of an attachment to the good, i.e. deviating to not buying causes less of a loss and (ii) in the money dimension, mass shifts from prices below $\hat{p}$ to prices above which means that buying at $\hat{p}$ is more often compared to a reference point that is associated with not buying, causing a greater loss. Both effects lead the consumer to choose a lower $\hat{p}$. This argument in the money dimension follows the same intuition as above where the comparison of paying p_M to paying p_L feels less painful than comparing paying p_M to paying 0.

As a second example, consider the setting in Mazar et al. (2014). The

[15]The formal proof follows the proof of Proposition 4 in Kőszegi and Rabin (2004) and I present an adapted version in the appendix. As do they, I make the assumption that the preference parameters η_k,λ_k, and u are such that the PE is unique which allows me to focus on PE only.

authors elicit willingness to pay for a number of goods (and in a number of different settings), but I will focus on their experiment 1 where they elicit the willingness to pay for a travel mug. Subjects are randomly allocated into two groups which differ only in the price distribution used for the BDM-mechanism (Becker et al., 1964). Some participants face a right-skewed distribution which has half of the mass on the lowest price in the support (\$1) and the remaining mass uniformly distributed between \$1 and \$10, whereas the other participants face a left-skewed distribution which has half of the mass on the highest price (\$10) and the remaining mass also uniformly distributed between \$1 and \$10. The good deal model's prediction for this setting is analogous to the previous example. As the average price in the distribution is higher when more mass is on the right, the consumer who faces a realised price between \$1 and \$10 feels more elated if the chances were high that \$10 realised, rather than in the case where \$1 was the most likely price. To see that KR again predict the opposite behaviour, assume that when facing the right-skewed distribution the consumer buys at $1 < \hat{p} < 10$ and all lower prices and that - as assumed in the previous paragraph (see fn.15) - this PE is unique. This means that he expects to buy the good with a probability larger than one half and expects to pay 1 with probability one half for it. Now imagine the same consumer choosing the same $\hat{p}$ when the distribution is left-skewed: the probability of buying is now reduced by one half (the price is now \$10 whenever it was \$1 before and the consumer does not buy at \$10) which reduces the attachment to the mug. Similarly, now instead of comparing "buying at $\hat{p}$" to "buying at 1", the buyer compares with the same probability of one half "buying at $\hat{p}$" to not buying which entails a greater loss. Hence, the consumer will reduce the maximum price he is willing to pay because, firstly, he feels less attached to the mug as he is less likely to buy at $\hat{p}$, and secondly, because he reduces the loss in the money dimension by only buying at lower prices, which are less painful when compared to not buying. The appendix contains a formal proof of this claim. The results in Mazar et al. (2014) are in general supportive of the good deal model. For the mug experiment, the average WTP with the right-skewed distribution was \$2.42, whereas with the left-skewed price distribution, participants were willing to pay up to \$5.08 on average.[16]

2.3 The Experimental Design

I conduct three sets of experiments (the full set of instructions can be found in the appendix) that all have the same general structure, but differ in some aspects. I will start by describing experiment 1 (conducted in March 2012) in detail and then highlight the differences compared to experiments 2 and 3 (conducted in June 2012 and May 2014, respectively). Each experimental session consists of three parts. In the first part, the subjects earn the money that they can then spend in parts two and three. The subjects start by filling out a personality traits questionnaire (Eysenck et al., 1985) consisting of 48 yes/no questions. For this, they are paid £9 which constitutes the money that they can use for the purchase decisions later on. The subjects are given £3 to use in part two and £6 to use in part three. They then move to part two where they are given the opportunity to buy a chocolate bar for a price that is determined by an individual draw of a coloured ball from a bag. The chocolate bar is in front of every subject on his/her desk from the moment they enter the lab. In usual grocery stores it sells for slightly above £2, but the subjects are not informed about this. The subjects are randomly put into two treatments. In the first treatment the possible prices are either £0.5 or

[16]Urbancic (2011) finds similar results examining various types of price distributions underlying the BDM.

£1, whereas in the second treatment the prices are either £1 or £2. The price determination procedure is explained at the beginning of part two on three consecutive screens that are shown to each subject for 60 seconds per screen. This is to force the subjects to think about their buying decision in advance before the resolution of uncertainty. On these screens, every subject is told which of her two possible buying prices for the chocolate corresponds to a blue ball and which one to a red ball. (In both treatments, the blue ball represents the higher of the two prices). An experimenter then puts - visibly for everyone - 5 red and 5 blue balls into a bag and then approaches each subject individually at her desk and asks her to draw a ball from the bag. Every drawn ball is put back into the bag and the result is entered into the required field on the screen. No participant observes the draw of the other participants. Each subject then decides whether to buy at the drawn price or not. If a subject decides to buy, she will be able to take the chocolate with her and the price is subtracted from the budget of £3. Otherwise she keeps all her money. Before moving to part three, each subject is asked some additional questions about her decision and is required to make 6 hypothetical choices between a binary lottery and a fixed payment of zero. In each choice the lottery pays either £6 or £x with $x \in \{-2, -3, -4, -5, -6, -7\}$ in decreasing order. The cutoff value where a subject switches from favouring the lottery to preferring the fixed payment of zero is used to elicit a measure of loss aversion, as previously used in Abeler et al. (2011) and Fehr and Goette (2007), though these papers incentivise these decisions.

In part three the subjects are offered two more goods, this time a notepad (A4, ruled) and a pen with the university's logo on it. These goods are also put in front of the subjects but only after everyone has finished with part two. They are also handed out new instructions because the first set of instructions for parts one and two only states that "the third part will be similar to the second part" and that "the two goods that you can buy then will be shown to you at the beginning of the third part." For the notepad (price in store around £3.50), the prices in the two treatments are £1.50 or £2 in the first and £2 and £3 in the second. The pen (price in the university's shop £1.40) is priced at either £0.5 or £1, or £1 or £2, respectively. For each purchase decision in part three the subjects have £3 at their disposal from part one. For these two decisions I use the strategy method, thus subjects have to make a decision before they learn the price realisation. They are asked to indicate for each price whether they want to buy or not. The actual price is then determined afterwards by the computer and the decision implemented accordingly. The setting where the price is determined before the decision is more in line with the theoretical setting presented, but the practical drawback is that, on average, one loses half the observations, namely all the subjects that do not draw the price of £1. Since subjects do not know in part two the good(s) they can buy in part three and neither is the amount of money they can spend in part three affected by their buying decision in part two, it seems to be reasonable to assume that subjects treat the decisions in parts two and three as independent.

While the setting in experiment 1 is useful as a starting point to see whether different price distributions affect buying behaviour at the same realised price, it is important to disentangle any potential effect on buying behaviour through reference-dependent preferences as in the two models discussed above, from an effect outside of these models. As described in the introduction, such an effect could be, for example, that the intrinsic valuation u is closely linked to the perceived market price of the good outside of the experiment, which in turn is influenced by the price distribution that a subject faces in the experiment. Hence, people might be more willing to buy at p_M in case MH because they infer from the higher average price in the experiment that the market price outside of the experiment is also higher. The next two experiments try to disentangle these channels, which is something that seems to have been done rarely in the reference price literature. Experiments 2 and 3 are designed in such a way that there should be little room for these "retail price inferences" to vary across treatments.

The setup is mostly the same with two notable differences in parts two and three. For experiment 2, the instructions for part two are amended such that every subject is told about all three prices for the chocolate bar. The instructions clearly state the two possible prices a subject faces (for example, either £0.5 or £1) but they also state the prices that the other half of the subjects in the experiment face (accordingly, either £1 or £2). That is, the expected prices are no different than in the experiment conducted before, but now every subject has the same (and complete) information about prices of the chocolate across treatments. Note that the subjects are never exposed to any uncertainty about which treatment they are in. In experiment 3, for part two I replace the branded chocolate used in the previous experiments with a chocolate bar that was custom-made for this experiment (which the subjects are also told in the instructions). The cover of the bar has a picture of the UCL main building printed on it. The subjects are also told (truthfully) that each bar cost the experimenter £3. Given that the chocolate bar now has a higher "production cost" (the previous branded bar cost about £2), I increase the prices to $p_L = 1$, $p_M = 1.5$, $p_H = 2.5$.

In part three, for both experiments, the notepad and the pen are replaced by an amazon.co.uk voucher that has a fixed value of £5. In experiment 2, it is offered to the subjects for either £3 or £3.50 in treatment *LM* or £3.50 or £4.50 in treatment *MH*. In experiment 3, the price $p_H = 4.5$ is replaced by $p_H = 5.5$. Again, the decision is conditional and made before the price realisation. Here, as in experiment 1, the subjects are not informed about both treatments and only see their two possible prices. Using a voucher should eliminate different beliefs about the retail price of the good, though it also creates a more artificial buying decision. Still finding an effect would be a very powerful result supportive of the theory that the expected prices serve as reference points. The higher p_H in experiment 3 is mainly motivated by the prediction of the good deal model that the positive sensation of facing the lower price increases in the difference between the realised price and the non-realised price (or, equivalently, the average price). Hence, as can be seen directly from equation (2.1), the good deal model predicts a stronger effect (higher buying proportion in MH) than before. The choice of a price that is higher than the redemption value of the voucher is an interesting case because we do not expect anyone to buy at that price, which then - thinking about the situation in terms of KR - might make it clearer for consumers that this price should not enter their reference point.

All sessions were run at the UCL-ELSE experimental laboratory with undergraduate students from UCL and there was no restriction imposed regarding their field of study. Subjects additionally received a show-up fee of £5. In total, 223 subjects participated, and each experiment consisted of four sessions. The experimental software used was z-tree (Fischbacher, 2007).

The analysis of the results below also uses some of the data of an earlier pilot study, conducted in January 2012. Here, the subjects were offered the chocolate bar in part three of the experiment, that is, subjects made a conditional decision. The prices were as in the main experiments described above.[17]

[17] In the pilot study the other goods were a USB memory stick (in part two) and the pen (together with the chocolate in part three). For both other goods, the focus was on having

2.4 Results

The results of experiment 1 are summarised in table 2.1-2.5. The results for the notepad - 23% buy at £2 if the other price is £3 and only 3% if the other price is £1.50 ($p = 0.014$)[18] - and the pen - 16% buy at £1 if the other price is £1.50 and only 3% if the other price is £0.50 ($p = 0.026$) - indicate that buying behaviour at p_M is significantly different between the treatments. For the experiment with the chocolate bar the results are less strong. There 52.6% of the subjects buy the chocolate at the price of £1 when the expected prices were £1 and £2, but only 25% buy if £2 is replaced by £0.50. Due to only 39 subjects in the sample that drew the price of £1, this fails to be significant at the 5%-level ($p = 0.105$). Table 2.2 also shows the results from the pilot study in which the chocolate bar was offered in the section where the subjects had to make a conditional decision. Comparing the behaviour across treatments and prices, I confirm that the behaviour in the pilot and in experiment 1 is not significantly different. Pooling the two together, the percentage of subjects buying at £1 in case *LM* (other price £0.5) is 25 % versus 57.1 % in case *MH* (other price £2). For the total of 60 subjects the difference is then significant with $p = 0.017$.

The data from this first experiment seems to provide overall some support for the good deal model and the hypothesis that the distribution of possible prices exerts an influence on the buying behaviour. These results fit well into the large literature in economics and marketing that supports that reference

different probabilities (0.1 vs. 0.9) across treatments while keeping prices constant. I do not report the results here as the present paper solely focuses on the case where prices differ across treatments, keeping probabilities fixed at 0.5. Because of the significantly higher value of the USB memory stick, the total budget of the subjects was £15, of which £10 were earned by filling out the questionnaire and allocated for the decision about the USB memory stick. For the other two goods the subjects were asked to use their show-up fee of £5, hence each decision had an allocated budget of £2.50.

[18] All p-values reported for the 2x2 tables are obtained via a 2-sided Fisher exact test.

other price	# buy at £1	# not buy at £1	
£0.50	5	15	20
£2	10	9	19
	15	24	39
			$p = 0.105$

Table 2.1: Chocolate Bar

other price	# buy at £1	# not buy at £1	
£0.50	3	9	12
£2	6	3	9
	9	12	21
			$p = 0.087$

Table 2.2: Chocolate Bar - Pilot

other price	# buy at £1	# not buy at £1	
£0.50	8	24	32
£2	16	12	28
	24	36	60
			$p = 0.017$

Table 2.3: Chocolate Bar - Pooled

other price	# buy at £2	# not buy at £2	
£1.50	1	35	36
£3	9	29	38
	10	64	74
			$p = 0.014$

Table 2.4: Notepad

other price	# buy at £1	# not buy at £1	
£0.50	1	37	38
£1.50	7	29	36
	8	66	74
			$p = 0.026$

Table 2.5: Pen

prices have a strong influence (see, for example, Mazumdar et al. (2005) for a review). A recent study by Weaver and Frederick (2012) presents a number of ways in which a reference price influences the stated willingness to pay for different goods. For example, in one study the authors elicit buying and selling prices for boxes of candy and provide subjects with different information about the market value (i.e. the price at a theatre versus the price at a normal store), and in another they change the sticker price of a pencil. They find that these changes affect buying and selling prices in the way that a higher reference price typically increases the valuation for the good. Hence the results of experiment

1 present similar evidence in a setting where the reference price manipulation is not done through a change in the price tag or direct information about its market price, but rather through a manipulation of the distribution of possible prices chosen by the experimenter.

Tables 2.6 and 2.8 show the results from the decisions involving the chocolate bar in experiments 2 and 3. In experiment 2, when all subjects are told both price distributions, 20 % of the subjects buy at £1 when the other price is £2, and 12.5 % buy when the other price is £0.50 ($p = 0.672$), thus there is no detectable difference in buying behaviour between the two treatments. In experiment 3, when informed about the production cost of the custom-made chocolate bar, 14 % of the subjects buy at £1.5 when the other price is £2.5 and 29 % buy when the other price is £1 ($p = 0.261$). Again, there is no difference in buying behaviour between two treatments. It should be noted, however, that the latter is the only case where the direction of the effect (though far from significant) is towards the effect that KR predict.[19]

The fact that the results for the chocolate bar from experiment 1 are not replicated in experiments 2 and 3, suggests that most of the effect in experiment 1 cannot be explained by the specific reference point effect in the good deal model. As described in section 2.2.1, the difference in buying behaviour is modelled as caused by the elation from drawing the cheaper of the two prices. This part of the experiment is, however, unchanged, and the predictions should therefore apply equally to all three settings. The fact that they do not, suggests

[19] I tried a number of specifications of a binary choice model that predicts the probability of buying at a given price (distribution) while controlling for the degree of loss aversion and the "personality traits" elicited through the questionnaire. Neither is found to have a significant influence on the buying behaviour. Additionally, I use this data to see - for each good separately - whether there are any notable differences among the subjects in the two treatments. Reassuringly for the randomisation procedure, across all experiments there is only one case out of 28 where the subjects show significant differences (at the 5% level) across treatments. One of the three personality traits - neuroticism - is significantly more prevalent among those who draw a price of p_M in treatment LM than in MH in experiment 2 for the chocolate bar.

that p_L and p_H do not affect the buying decision once there is little room for subjects to make different inferences about the retail price of the chocolate bar.

However, the retail price inference hypothesis makes a clear prediction about how buying behaviour should change from experiment 1 to experiment 2: the proportion of subjects buying at p_M in *MH* should decrease whereas it should increase in *LM*. Giving the subjects in the treatment with £1 and £2 the additional information of a third price of £0.50 makes the proportion of subjects buying drop from 52.6% to 20% ($p = 0.036$). This is clearly in line with the described effect. However, there is no indication that for subjects with prices of £0.50 and £1 - now knowing the third price of £2 - their valuation increased; if anything it drops from 25% to 12.5% ($p = 0.306$).[20] [21]

Looking at the results for the voucher, shown in tables 2.7 and 2.9, there is a higher percentage of subjects buying in treatment MH, but the difference (33.3% vs. 47.4% in experiment 2 and 37.8 % vs. 47.7% in experiment 3) is not significant ($p = 0.322$ and $p = 0.500$). Hence, similar to the results in experiments 2 and 3 for the chocolate bar, I find no effect of the non-realised price on the buying behaviour. Also, changing p_H does not affect buying behaviour significantly. One possible reason that makes it harder to detect an effect in this setting is that the amazon voucher is a good that might be very attractive at a price below its redemption value for regular amazon shoppers, whereas it might be completely unattractive for others who never use

[20]One of the survey questions given to the subjects after the chocolate bar decision in experiment 2 asks them about what they think the chocolate bar costs in a supermarket. Surprisingly, a Wilcoxon ranksum test, reveals that the subjects in treatment MH think the chocolate bar is more expensive compared to the subjects in treatment *LM* ($p = 0.0373$). While the reliability of the (unincentivised) answer should not be exaggerated, the significant difference in responses might reveal that the subjects did not take into account all the information given to them.

[21]As experiments 1 and 2 are three months apart, it could be that there is some seasonal effect that explains the overall lower buying proportions in experiment 2.

other price	# buy at £1.00	# not buy at £1.00	
£0.50	2	14	16
£2.00	4	16	20
	6	30	36
			$p = 0.672$

Table 2.6: Chocolate Bar (Experiment 2)

other price	# buy at £3.50	# not buy at £3.50	
£3.00	10	20	30
£4.50	18	20	38
	28	40	68
			$p = 0.322$

Table 2.7: Amazon Voucher (Experiment 2)

other price	# buy at £1.50	# not buy at £1.50	
£1.00	5	12	17
£2.50	3	19	22
	8	31	39
			$p = 0.261$

Table 2.8: Chocolate Bar - custom made

other price	# buy at £3.50	# not buy at £3.50	
£3.00	14	23	37
£5.50	21	23	44
	35	46	81
			$p = 0.500$

Table 2.9: Amazon Voucher (Experiment 3)

amazon. Thus, I asked in experiment 3 how often subjects buy something from amazon, ranging from "once a week" to "never". Excluding the extreme cases and concentrating on occasional shoppers (who might be more price sensitive), however, does not change the results.[22]

Whereas retail price inferences should not affect buying behaviour differently across treatments in experiments 2 and 3, a more subtle issue is to what extent anchoring could be an explanation for the results. Anchoring is a well-documented phenomenon (though less well-defined theoretically) and describes the effect that various environmental cues can have a strong effect on individual behaviour (Tversky and Kahneman, 1974; Ariely et al., 2003). Hence, it could be that experiments 2 and 3 have eliminated value inference, but not anchoring. Subjects may use the possible prices they face as their main anchor for their valuation, attaching more weight to them than to the additional infor-

[22]Due to a technical problem, this question was only asked in 3 out of the 4 sessions in experiment 3.

mation such as the value of the voucher or the production cost of the chocolate bar. If that were the case then experiments 2 and 3 would still induce a bias against KR and could be a potential reason why I do not find evidence in accordance with their model. However, the information about the voucher value and the production cost are being given to subjects before they get to know the prices. Hence, they constitute the first cue that subjects are exposed to when looking for anchors. At the same time, they arguably are more relevant anchors for valuation than the prices.[23] It is thus rather unlikely that subjects in experiments 2 and 3 hold significantly different anchors that would bias the results against KR. Alternatively, instead of considering anchoring of the valuation, as in the studies cited above, one could argue that the expected prices in itself form an anchor and the realised price is evaluated relative to this anchor. This specific form of anchoring, however, is clearly captured in the good deal model which explicitly models this psychological phenomenon as a reference point effect.

An open question that is relevant for empirical work based on KR is how long it takes for individuals to form their reference point. As discussed in section 2.2.2, according to the theory, forming the reference point involves developing a plan that specifies buying behaviour at each possible price realisation. Thus, a potential reason why I fail to find evidence for the KR predictions could be that subjects did not have sufficient time to develop a plan. There are two reasons why it seems unlikely that the issue of timing has an effect on the results.[24] First, in those parts of the experiment where

[23]There does not seem to be much evidence on how people integrate different anchors. Comparing different anchors, though also only one at a time, Sugden et al. (2013) find that in order to affect willingness-to-pay, an anchor needs to be "plausible". In my experiment, the voucher's redemption value, for example, clearly is a plausible anchor.

[24]While I am not aware of any research specifically answering this question, Gill and Prowse (2012) note that in their real-effort experiment in a tournament setting where a second mover's reference point takes into account the first mover's realised effort, the adjustment of the reference point happens instantaneously. While their setting is clearly in a

subjects have to make a choice at each possible price before they know the actual realisation, subjects are forced to make a plan. The results from these parts (i.e. involving the pen, notepad, and voucher) are equally unsupportive of KR as are the results from the parts involving the chocolate bars. Second, after they have made their decision whether to buy the chocolate bar, subjects are asked whether they would have bought at the other price as well. Here, subjects were specifically given the option to answer "maybe". In total, only 14.8% of subjects indicated that they are not sure. Hence, a large majority of subjects seems to have formed an opinion about what they would do at the other price.

	Experiment 1		Experiment 2		Experiment 3	
	mean	median	mean	median	mean	median
Treatment LM	0.75	0.7	0.78	0.6	0.78	0.5
Treatment MH	0.95	1	0.83	0.80	0.95	1
p-value	0.0972		0.1243		0.0423	

Notes: The p-value is obtained by using a two-sided Wilcoxon ranksum test. In experiment 1 (treatment LM), one subject stated a WTP of 60 which seems implausibly high. My explanation is that (s)he wanted to report 0.6 (i.e. meant 60 pence) and I changed this accordingly. This is consistent with the subject buying the chocolate at the realised price of 0.5 and indicating that (s)he would not have bought at a price of 1.

Table 2.10: Willingness To Pay for the Chocolate Bar

Analysing the hypothetical buying decisions (excluding the participants who answer "maybe"), we see that behaviour is quite similar to the real decisions. In experiment 1, 9.1% would buy in case LM whereas 43.8% would buy in case MH which is significant at the 10%-level ($p = 0.090$). Comparing the behaviour of the hypothetical answers with the real answers, I detect no significantly different behaviour in any of the four groups. Pooling the hypothetical and real answers together for the behaviour at $p_M = 1$ yields a significant

different domain than mine, their results show that it can well be the case that reference point formation happens quite quickly.

difference (19.3% vs. 48.6%, $p = 0.019$). In experiment 2, the hypothetical behaviour shows the buying percentages as 23% vs. 55% which is not significantly different ($p = 0.187$) but there the data reveals different behaviour between the real and hypothetical choices, hence it does not seem appropriate to pool the two together. Even if I do, the result of no significant difference across treatments is confirmed. The same holds for experiment 3, where the hypothetical buying proportions at $p_M = 1.5$ are 9.1% in case LM and 12.5% in MH.

What is more, I also ask subjects - again without providing incentives - to state their willingness-to-pay for the chocolate bar. Table 2.10 shows the results depending on the treatment. In all three experiments the mean and median WTP is higher for those who are in treatment MH, but only in experiment 3, the difference is significant at the 5% level. These results seem to indicate some support for the good deal model but beg the question why this effect does not translate into differences in actual buying behaviour. The reason for this could be that mean and median WTP are significantly lower than p_M in experiments 2 and 3 and therefore there are not enough subjects with an intrinsic valuation in the relevant range. Nevertheless, they are in line with what Mazar et al. (2014) find when eliciting WTP (in an incentivised manner) in their experiments.

2.5 Conclusion

The present paper aims to assess the prevalence of reference dependency in a consumer purchase decision. I derive clear cut predictions of the model by Kőszegi and Rabin (2006) in section 2.2.2 for a simple setting where a subject faces two possible prices for a good which are equally likely to realise as the actual buying price. I believe that I am the first to highlight the discrepancy between the predictions that KR's concept of personal equilibrium makes from a related model that I call "good deal model" which captures a more intuitive notion of reference dependent behaviour in a consumer purchase decision. I show furthermore that the discrepancy between the two models is not restricted to the simple two price setting of the experiment but rather is a fairly general result that also holds for other price distributions. My results of experiment 1 indicate some support for the good deal model. Experiments 2 and 3 explore the results of experiment 1 further and show that in settings in which there is little room for different inferences about the valuation of the goods across treatments, different underlying price distributions do not significantly affect buying behaviour.

The main lessons from this are as follows: First, in accordance with many studies of reference pricing in the marketing literature, I show that reference prices play a significant role in affecting consumer behaviour. The present paper offers some further insight into this topic by credibly creating a distribution of possible prices for a good, thus directly manipulating the prices that a subject in the experiment expects to pay. The results suggest that the non-realised price has an effect on buying behaviour. This effect, however, cannot be fully attributed to elation (disappointment) from a draw of a cheaper (dearer) price, as predicted by the good deal model. Instead, since the non-realised price does not significantly affect buying behaviour neither for the amazon voucher nor the customised chocolate, the most plausible explanation for the results seems to be that for a price of $p_M = 1$, more subjects in MH bought the chocolate than in LM because they inferred a higher retail price from the price distribution. However, as outlined in section 2.4 this conclusion can only be regarded as tentative as there are pieces of evidence (comparison of WTP and buying behaviour in LM in experiment 2) that also do not fully fit this alternative explanation.

Second, the fact that I do not find evidence for personal equilibrium behaviour suggests that individuals are not influenced by their own expected behaviour. That is, unlike predicted by KR, they do not internalise the consequences arising from the anticipated decision not to buy the chocolate bar at the price of £2. This view might be supported by the observation that firms often use sales practices where they present consumers with unreasonably high "standard" prices only to offer the good at a big "discount". KR's theory says that such a practice does not work since consumers anticipate that they do not buy at the standard price and therefore do not feel a gain from the reduced price. The results from experiment 1 show that such practices may work well - especially when subjects may use the price distribution as indicative of the product's value and therefore believe they are buying a more valuable item.

A route that would explore KR further in this respect could be to replace the price p_H with the event that subjects are not able to buy the good at all. By doing so, one would "force" the consumer to anticipate that he will not buy with probability one-half ex-ante. Essentially the data tells us that p_H is a price that almost all subjects regard as too high compared to their valuation (for example, taken all experiments with a chocolate bar together, only 1 out of 54 subjects bought at p_H) but they seem to be unable to discard it from their reference point as predicted by KR. Offering the amazon voucher in experiment 3 at a price above its redemption value was motivated by this idea. As indeed no subject wanted to buy the voucher for £5.50, it seems reasonable to think that subjects in treatment MH did expect ex-ante to obtain the voucher either with probability zero or one half. Since experiment 3 revealed that the price of £5.50 had a very similar effect than the price of £4.50 in experiment 2, it would be interesting to explore in further research whether specifically implementing the event that with some probability subjects cannot buy has a different effect. In the same spirit, it could be worthwhile to see whether consumers are able to "learn" personal equilibrium behaviour. Maybe they are able to learn after a number of purchase decisions that the high price of £2 is not a "relevant" price.[25] However, it should be noted that while these modifications might make detecting an effect as predicted by KR more likely, the predictions of KR are perfectly applicable to the experimental setting chosen in this paper.

[25]It is not straightforward to implement a repeated purchase decision for the same good. Using the strategy method whereby one only implements one of many decisions by randomly selecting one choice at the end does not yield sufficient independence between the decisions.

Chapter 3

Exploiting Biased Beliefs — An Experimental Investigation

3.1 Introduction

When purchasing a product or signing up for a service, a consumer needs to form an accurate assessment of the product characteristics, contractual terms, or his own predicted usage to find the best deal available. An increasingly large literature documents that consumers often struggle with this, and instead hold *biased beliefs* about important aspects of their environment. For example, consumers may neglect or discount the relevance of add-on charges such as shipping costs (Hossain and Morgan, 2006), underestimate borrowing costs on credit cards (Ausubel, 1991), or misperceive future energy costs for cars (Allcott, 2013). Also, consumers may be biased about their ability to avoid overdraft penalties on bank accounts (Stango and Zinman, 2014), unable to accurately forecast their mobile phone usage (Grubb and Osborne, 2015), or their likelihood of exercising (DellaVigna and Malmendier, 2006). What is more, it does not seem far-fetched to think that firms are aware of these limitations of their customers and try to exploit them by designing contracts that directly cater to these biases. Models in behavioural industrial organisation have been developed to formalise these ideas and show how a number of real-world pricing strategies by firms can be explained by the presence of consumers with biased beliefs.[1] For example, DellaVigna and Malmendier (2004) show how firms use flat rate tariffs for gym memberships to exploit consumers' mistaken belief that they will exercise more than they actually do.[2] Grubb (2009) shows how three-part tariffs for mobile phone plans can be explained as the firms' response to consumers having wrong beliefs about their calling behaviour.[3]

The key contribution of this paper is to use a laboratory experiment to shed new light on the question of what effect behavioural biases have on market outcomes. As described above, the failure to accurately perceive product or market characteristics can have important consequences for individual decision making. But (at least) equally important as documenting effects of biases on behaviour is to examine the more general consequences for market outcomes of such misperceptions. Therefore, in this paper, I investigate in a simple trade environment whether there is evidence for negative welfare effects on buyers caused by their biased beliefs. Put differently, can sellers exploit the buyers' false beliefs, that is, do sellers earn higher profits when facing biased rather than unbiased buyers? Such questions are of direct policy relevance. Compe-

[1] For comprehensive syntheses of these and other models, see Spiegler (2011) and Kőszegi (2014)

[2] Eliaz and Spiegler (2006) consider more generally the incentives of sellers to screen consumers according to their time-inconsistency. They use their model to provide explanations for pricing practices such as credit cards with low "teaser rates" followed by higher standard rates aimed at consumers who mis-predict their borrowing behaviour.

[3] These papers mainly concentrate on the incentives of a monopolist to price optimally in the presence of consumer biases, which is also the focus of my paper. In competitive markets, Gabaix and Laibson (2006) and Armstrong and Vickers (2012) analyse settings where consumers underappreciate add-on prices, and Heidhues and Kőszegi (2010) consider borrowers with naive beliefs about their loan-repayment probability.

tition authorities and regulators increasingly care about the role of consumer biases in markets and are interested in whether, for example, there are negative effects of biased beliefs that would warrant interventions such as the mandatory disclosure of product attributes or usage behaviour.

On some level, the opportunities for firms to exploit consumer biases seem straightforward. However, there is an important assumption underlying the exploitation possibilities when consumers have biased beliefs. In order for firms to profitably use, for example, contracts with high add-on fees where buyers underestimate the likelihood that the add-on fee will need to be paid, buyers need to be strategically naive about the sellers' incentives. That is, they cannot become suspicious when being offered such contracts, which would lead them to question their beliefs.[4] In contrast to that, if consumers were to hold biased beliefs but are sophisticated about their bias, such exploitative contracts will not be profitable for sellers and have no negative effect on buyers. Because they are aware that the seller has superior information and a potentially different belief, buyers will never sign such exploitative contracts.[5]

Hence, in this paper, I design an experiment that provides a direct test whether biased beliefs of buyers lead to higher profits for (monopoly) sellers. Unlike field studies, the laboratory environment allows for a direct comparison of an environment where sellers face buyers with biased beliefs with an environment where buyers have equally accurate beliefs as sellers. In the experiment, a seller offers a product to a buyer and either the product itself, or the resulting contract upon acceptance, will induce payoffs that depend on an

[4]The alternative assumption is that they are aware of the belief difference, but maintain that their belief, rather than the firm's is the correct one, they "agree to disagree".

[5]In fact, this result links to work on "No-Trade Theorems" (Milgrom and Stokey, 1982; Tirole, 1982) that provide theoretical arguments as to why there should be no mutually beneficial trade based on differences in private information. As I will show in Section 3.2, when buyers are sophisticated, a similar logic applies to the exploitative (or state-contingent) part of the offered contract.

underlying state. The aim of the design is to I create settings that differ in the degree to which buyers have biased beliefs about the realisation of this state. I can then compare the outcomes with control conditions in which buyers and sellers share the same beliefs. This allows me to directly quantify whether there is evidence of exploitation of biased beliefs.

I employ a novel design to induce systematic belief biases on the buyers' side. Specifically, I use the result established by Ensthaler et al. (2015) that many people have huge difficulties in understanding multiplicative growth processes, especially when these processes are not deterministic. Almost all participants in their study demonstrate an inability to correctly understand the workings of these processes.[6] This type of bounded rationality is a manifestation of the so called "exponential growth bias" (Stango and Zinman, 2009). For my purposes this behavioural bias turns out to be a useful workhorse as the wrong perception of the process promises to be largely mitigated with sufficient feedback. I thus exogenously vary the information given to subjects such that in the treatment condition only sellers receive feedback (and should therefore on average have correct beliefs), whereas in the control condition both buyers and sellers receive feedback. In this way I create belief differences that neither involve deception, nor are caused by different signals drawn from the same underlying distribution. On the contrary, since all subjects have a complete description of the process, it is their bounded rationality that brings about the belief differences. Moreover, there will always be a clear prediction regarding the direction of the belief bias, which is crucial for the experimental design. Finally, buyers in the experiment know that sellers obtain feedback. To this extent, the setting also reflects the observation that in many markets firms do

[6]In the basic setup that I borrow from Ensthaler et al. (2015), subjects are required to develop an understanding about the value of an asset whose value in each period can either increase by 70% or decrease by 60%. Most subjects underestimate how likely it is that over time this asset is worth much less than its starting value.

not only know their products better than consumers, but are also typically better at predicting consumer behaviour, based on the large amounts of data available to them.

The paper reports on two different experiments that, taken together, provide a comprehensive analysis on how biased beliefs affect strategic interactions between buyers and sellers. In the first experiment (the "add-on price experiment"), a seller can design a state-dependent contract by charging an add-on price in addition to the base price. This add-on price only needs to be paid with some (exogenous) probability, which depends on the outcome of the stochastic multiplicative growth process. This process is designed such that biased beliefs lead buyers to underestimate the probability with which the add-on price needs to be paid. Unless buyers are sophisticated enough to realise the sellers' motives, sellers could thus increase profits by using high add-on prices. Interestingly, the main result that emerges from the data is that there is no evidence for exploitation. Sellers do not earn higher profits from interacting with buyers that have biased beliefs, compared to the profits they obtain in the case when buyers and sellers share the same belief. This is despite the fact that the feedback manipulation had the desired effect: by directly eliciting beliefs of all participants, I confirm that buyers who are not exposed to feedback hold biased beliefs in the a priori anticipated direction. Importantly, there is evidence that the no-exploitation result is not driven by the inability of sellers to price optimally, but rather that buyers find ways of protecting themselves. They seem to be sufficiently sophisticated to understand the sellers' incentives. Buyers with biased beliefs have a significantly higher propensity to reject offers with high add-on prices which prevents them from paying more for the product than buyers in the control condition. This leads sellers to set very similar prices in the treatment and the control condition.

In the second experiment (the "insurance experiment"), the product's value may be reduced due to an adverse event. While in general the buyer may suffer from this loss in value once he buys the product, the seller can offer insurance as part of the contract and take the risk off the buyer. I now create a setting where uninformed buyers are expected to hold upward-biased beliefs, thus giving sellers a motive to sell insurance at inflated prices. In the add-on price experiment the (potentially) exploitative contract necessarily lead to a risky outcome for the buyers upon acceptance. In contrast to that, the insurance experiment is designed such that if buyers were to be exploited, this would happen via "safe" contracts that guarantee a fixed payment at the time of acceptance. It can therefore distinguish between buyer sophistication and an aversion to accept risky contracts as an explanation for the results of the add-on price experiment.[7] Exploitative contracts that feature high insurance fees should now be relatively more attractive to buyers who want to avoid being exposed to risky contracts. In line with the explanation put forward previously, however, the results show that sellers do not earn higher profits from buyers with biased beliefs. More specifically, while buyers in the treatment condition pay significantly more for insurance than their counterparts in the control condition, they are compensated through lower prices for the product itself. As it turns out, once buyers are assumed to care sufficiently much about their payoff relative to the sellers' payoff, this trade-off between high insurance fees and lower prices is consistent with the theoretical predictions when buyers are sophisticated.[8]

[7] As I will explain in more detail later, buyers in the treatment condition who do not receive feedback might perceive the contract as more ambiguous than buyers in the control condition. Aversion to such ambiguity might make them more likely to reject offers with a positive add-on price.

[8] The qualifier regarding relative payoffs is important because otherwise we would expect that no insurance should be sold when buyers are sophisticated. The reason for this is that then a simple adverse-selection logic prevails that makes buyer realise that any price at which the seller is willing to sell insurance corresponds to a price that would lead to a loss

The implications of these results are as follows: while both experiments demonstrate that biased beliefs affect market behaviour since decisions by both buyers and sellers differ between treatment and control conditions, there are no statistically significant differential effects on total earnings. To the extent that these results stem from buyers' sophisticated reasoning about exploitation possibilities —and my data provides evidence for that— they imply that in real world markets similar effects might be present, too. If the findings could largely be attributed to sellers' inability to exploit, this would raise a natural question of external validity. However, buyers in the experiment are regularly faced with purchase decisions of comparable monetary value in their daily life. Hence, taken at face value, interventions such as disclosure policies aimed at reducing biased perceptions about product features may not lead to welfare gains for consumers. Looking at the results from a behavioural IO perspective, this suggests that at least in some markets, consumer sophistication may play a larger role than typically assumed, the more so for industries in which the superiority of seller information and beliefs is salient.

The next section provides a general framework for the role of biased beliefs in strategic settings. Section 3.3 describes the design and predictions of the add-on experiment in more detail and I present the corresponding results in section 3.4. Analogously, sections 3.5 and 3.6, respectively, present the insurance experiment, while in section 3.7 I discuss the broader implications of my findings and relate them to the existing literature. Section 3.8 concludes.

for the buyer. However, once relative payoffs matter, using the model by Fehr and Schmidt (1999), I show that sophisticated buyers sometimes buy insurance while being compensated by lower prices for the product.

3.2 Theory

In this section I will briefly describe a simple trade environment with one buyer and one seller (both assumed to be risk neutral) and highlight how differences in beliefs may affect both pricing strategies and profits. Importantly, this section will highlight how the distinction between buyers that are naive about their belief bias differ from sophisticated buyers that are aware of the possibility that their beliefs might be wrong.

Suppose that there is a seller who, for a price p, can sell a product to a buyer who values this product at x. If buyer and seller do not agree on a mutually acceptable transfer, they receive their respective outside options $y^B, y^S \geq 0$. I assume that $x > y^B, y^S$, so selling is always efficient. Denote the highest price that the buyer is just willing to accept, i.e. his willingness to pay by $\hat{p} = x - y^B$. To make things more interesting, assume that there is an underlying state of nature, θ, with $\theta \in \{1, 2\}$. Denote by π^B and π^S the (subjective) probability of buyer and seller, respectively, that state $\theta = 1$ realises. To be concrete, consider the case where the product in question is a personal account and p the monthly account fee. Furthermore, $\theta = 1$ then may describe the event that the account is overdrawn. In this case an additional payment, such as an overdraft fee, would be due. More formally, such a contract involves state-contingent transfers. Denote such transfers (net of the price $\hat{p}$) by $g = (g_1, g_2)$ which consist of a payment from buyer to seller of g_1 if $\theta = 1$ and g_2 if $\theta = 2$. If accepted, payoffs for the seller, u^S, and the buyer, u^B, will be given by:

$$u^S = \begin{cases} \hat{p} + g_1 & \text{if } \theta = 1 \\ \hat{p} + g_2 & \text{if } \theta = 2 \end{cases} \qquad u^B = \begin{cases} x - \hat{p} - g_1 & \text{if } \theta = 1 \\ x - \hat{p} - g_2 & \text{if } \theta = 2 \end{cases} \tag{3.1}$$

I then define $E^j(g) \equiv \pi^j g_1 + (1 - \pi^j) g_2$ for $j \in \{S, B\}$ as the expected value of the transfer g from the perspective of the seller and buyer, respectively. Importantly, since any transfer component that is common across states can be made via p, the reason why there may exist a separate role for transfers that differ across states are belief differences. To this extent, such contracts are what Eliaz and Spiegler (2007, 2009) call "speculative". On a technical note, I constrain the possible contracts g such that $g_1, g_2 \in [\underline{g}, \bar{g}]$ with $-\infty < \underline{g} < \bar{g} < \infty$. Otherwise, agents may want to sign arbitrary large "bets".[9] Before analysing the case with belief differences, consider the case where both buyer and seller have identical beliefs, $\pi^B = \pi^S = \pi$. Since the seller only agrees to the contract g if $E^S(g) \geq 0$ and the buyer only if $E^B(g) \leq 0$, and they agree on the assessment of each contract, the following proposition holds.

Proposition 3.1. *If $\pi^B = \pi^S = \pi$, the only transfers that both buyer and seller will agree to are characterised by $E^B(g) = E^S(g) = 0$.*

Thus, neither party can increase expected profits from using these additional contracts when there are no belief differences. In what follows, I will consider the case where the prior beliefs of buyer and seller differ. I look at two settings that vary in the assumed strategic sophistication of the buyer.

3.2.1 Naive Buyers

From now on, I will call a buyer *naive* if he holds a belief which differs from the seller's belief, $\pi^B \neq \pi^S$, and, furthermore, that he considers his beliefs correct and would not change his belief upon receiving information in contradiction with his belief via the seller's offer. The following proposition states what

[9] In their setting, Eliaz and Spiegler (2007, 2009) get around this undesirable feature by assuming that the state is not verifiable and thus bets cannot directly be based on the state, but on the agents' behaviour. Restricting the type of bets by constraining the payoffs is consistent with the experimental setting.

types of mutually agreeable transfers g both parties will agree on, depending on their beliefs.

Proposition 3.2. *If $\pi^B \neq \pi^S$ and the buyer is naive, both buyer and seller will find it beneficial to agree on a transfer $g \neq (0,0)$ that has $g = (g_1, g_2)$ such that $E_B(g) \leq 0$ and $E_S(g) \geq 0$. The seller's optimal offer then has $E_B(g) = 0$ and $E_S(g) > 0$. Thus, if buyer and seller hold different beliefs, speculative contracts will emerge.*

Proof. The first part is immediate, as either party only signs contracts that make them at least as well off as with $g = (0,0)$. Define $g^o \equiv (0,0)$ and note that clearly, $E_B(g^o) = E_S(g^o) = 0$. Now, consider w.l.o.g. $\pi^B > \pi^S$. Since the buyer thinks $\theta = 1$ is more likely than the seller, there exists a g' such that decreasing g_1^o to g_1' and increasing g_2^o to g_2' leads to $E_B(g') \leq 0$ and $E_S(g') > 0$. Since the seller makes a take-it-or-leave-it offer, he adjusts g_1' and g_2' such that he maximises his expected payoff and the buyer is just indifferent between accepting and rejecting. □

Without taking a stance regarding whose beliefs (if any) are correct, there is little we can say about the welfare effects associated with this additional contract. However, once we assume — and the experiment is specifically designed to reflect this — that the seller's beliefs are correct, we can call these contracts exploitative. Due to their biased beliefs, buyers sign contracts that they (based on the unbiased belief) should not sign and that will create additional profits for the seller.

This result can therefore offer a potential explanation for many contracts that we observe in reality. Firms may prefer to use contracts that require customers to make different payments depending on their behaviour or some event outside of their control, if they have reason to assume that consumers have a wrong perception about these events. In the example above, high overdraft fees will increase the seller's profit if buyers underestimate the probability of maintaining the required balance on their account.

The assumption made in this section posited that both buyer and seller hold some belief without modelling where such a belief (and the associated bias) comes from. While in reality there might be a myriad of different factors responsible for a biased belief about product characteristics or usage behaviour (such as overconfidence, inattention, or too high complexity), my experimental design, as explained below, induces biased beliefs in a specific way. Since I give selective feedback to subjects, one interpretation of this feedback mechanism is that it provides private information to sellers. If buyers thus were to interpret the setting in this way, they would hold beliefs over a distribution of possible seller types, corresponding to different true probabilities. The analysis in this section can then be interpreted as buyers that do not reason about how different seller types choose their offers. In the sense of Eyster and Rabin (2005), they are "fully cursed". The next section analyses the case where buyers take the sellers' private information into account when evaluating offers.

3.2.2 Sophisticated Buyers

I model this game with asymmetric information by assuming that from the perspective of the buyer there are N possible seller types π_i^S with $0 \leq \pi_1^S \leq \pi_i^S \leq \pi_N^S \leq 1$, each type directly corresponding to the true probability. Denote by $\mu(\pi_i^S)$ the probability of each type, so that the buyer's ex-ante expected probability of $\theta = 1$ is given by $\pi^B = \sum_i^N \pi_i^S \mu(\pi_i^S)$. The following proposition shows that there are no exploitation possibilities.

Proposition 3.3. *If buyers are sophisticated, there always is an equilibrium where $g = (0,0)$. Any contracts with $g \neq (0,0)$ that will be accepted, have*

$E^S(g) = 0$.

Proof. Suppose that $g = (g_1, g_2) \neq (0,0)$ is part of a contract that is offered by a seller and accepted by a buyer. Define $\Pi \subseteq \left\{\pi_i^S\right\}_{i=1}^N$ as the set of seller types that, in equilibrium, would offer such a contract. Clearly, any $\pi_i^S \in \Pi$ must have

$$E_i^S(g) \geq 0 \quad \textit{and} \quad \sum_{\pi_i^S \in \Pi} E_i^S(g)\mu(\pi_i^S | \pi_i^S \in \Pi) \leq 0$$

This is because a seller could always offer $g_1 = g_2 = 0$ to guarantee a payoff of zero, and a buyer could always reject. Now, assume w.l.o.g. that $g_1 \geq 0 \geq g_2$, in which case $E_i^S(g)$ is increasing in π_i^S. Note, that if there were to exist a $j < N$ such that $E_j^S(g) \geq 0$, then all $i > j$ must have $E_i^S(g) > 0$ and therefore $\Pi \subseteq \left\{\pi_i^S\right\}_{i=j}^N$. But then it would have to be the case that $\sum_{\pi_i^S \in \Pi} E_i^S(g)\mu(\pi_i^S | \pi_i^S \in \Pi) > 0$ which contradicts the definition of Π. This leaves $i = N$ as the only seller type at which contracts different from $(0,0)$ would arise. To show that such contracts are feasible, note that if $E_N^S(g) \geq 0$ and $E_i^S(g) < 0$ for all $i < N$, then only types $i = N$ offer such contracts which the buyer accepts if $E_N^S(g) \leq 0$. Thus, in this case any g satisfying $E_N^S(g) = 0$ would be accepted by the buyer. Analogous reasoning applies for the case where $g_2 \geq 0 \geq g_1$, in which case only g with $E_1^S(g) = 0$ are feasible. Clearly, $(g_1, g_2) = (0,0)$ will always be an equilibrium because every buyer would accept such an offer, and any deviation from this is either not profitable for any seller, even if it were accepted, or would lead the buyer to update his beliefs such that $\sum_{\pi_i^S \in \Pi} E_i^S(g)\mu(\pi_i^S | \pi_i^S \in \Pi) > 0$ in which case he would reject. □

We see that when buyers are sophisticated, the additional contractual option is not used, unless the seller is of the most extreme type. The underlying logic of the proof is a standard adverse selection/no-trade argument: if there is a type of seller that could make a profit from offering a state-contingent contract that leads to a gain for him in one state and to a loss in the other, he must believe that he gains in expectation, such that the other side of the bet loses. Thus, upon receiving an offer, the buyer updates his beliefs to exclude types that would not find it profitable to offer such bets and then rejects. This leads to full unravelling and sellers do not earn more than they would by just using the state-independent transfer $\hat{p}$. It is also important to note that whenever the equilibrium contract is $g = (0, 0)$, we have a pooling equilibrium and therefore the buyer will not learn the true type.

3.3 Add-On Price Experiment — Design and Predictions

The basic structure of the add-on price experiment is as follows. In each of 20 identical rounds, a participant is randomly allocated the role of a buyer or a seller for the whole duration of the experiment. In each round a seller makes an offer to a buyer that he is paired with. This offer specifies a base price and an add-on price at which the buyer can buy a product with a known monetary value.[10] The difference between the base price and the add-on is that the latter only needs to be paid with an exogenously given probability less than one. In the experiment, a concrete framing for the add-on price is chosen in order to make it simpler for subjects to understand. Specifically, they are told that the seller will choose a price and a repair fee. If the buyer agrees to purchase the product for the price-fee combination chosen by the seller, a random process

[10] I deliberately chose the setting with non-neutrally framed roles of buyers and sellers. Since sellers can make a take-it-or-leave-it offer, the game resembles the well-known ultimatum game (Güth et al., 1982). Hoffman et al. (1994) show that by framing the game as a game between buyers and sellers the predisposition of subjects to share "the pie" equally is reduced. Since the main focus of the paper is not on social preferences per se, in order to isolate the exploitation motive as cleanly as possible, it thus seems sensible to employ a frame that is favourable in this respect.

determines whether the product is faulty, in which case the repair fee needs to be paid in addition to the base price.

Buyers and sellers interact in matching groups of eight (four buyers and four sellers in each) and are randomly re-matched with each other within their group. A participant either participates in the treatment condition or the control condition of the experiment. The two conditions only differ in whether buyers are exposed to feedback or not, more on this below. For each participant, one out of the 20 rounds was randomly selected for payment. In the second part of the experiment, I elicit a measure of risk aversion via a multiple-price list, similar to Holt and Laury (2002), some simple measures of social preferences via four allocation decisions of money between oneself and another randomly chosen person, as in Bartling et al. (2009), and employ Frederick (2005)'s cognitive reflection test.[11] In total, 96 individuals participated in this experiment, in a total of 4 sessions (2 treatment, 2 control), with 24 participants in each session. No subject participated in more than one session. Overall average earnings were £21.02 in the add-on experiment and £27.11 in the insurance experiment, including a show-up fee of £5 and each experiment lasted about 2 hours. Subjects were undergraduate and graduate students from UCL from a variety of disciplines. The experiment was programmed in ztree (Fischbacher, 2007) and recruiting was done via ORSEE (Greiner, 2004).[12]

The main innovation of the design is, in the treatment condition, to introduce differences in beliefs between buyer and seller about the likelihood with which the product is faulty. In order to achieve this, I borrow parts of the setup from Ensthaler et al. (2015). In one experiment in their paper, they ask sub-

[11] A detailed description of these tasks can be found in the appendix.

[12] I also conducted a pilot experiment on a version of the add-on price experiment with 32 subjects. While the design was similar to the add-on price experiment, I did not incentivise the belief elicitation and did not elicit second-order beliefs. Footnote 26 and section B.2.1 in the appendix provide a succinct overview.

jects to predict the value of an asset that starts at an initial value of 10000 and in each of 12 periods can either increase in value by 70% or decrease by 60% where both events are equally likely. Ensthaler et al. (2015) use a procedure that elicits the subjects' belief about the median of the distribution at the end of the 12 periods. Clearly, one decrease cannot be made up for by one increase, so over time the mass of the implied distribution of the asset shifts more and more towards zero. The median is given by $10000(1.7)^6(0.4)^6 \approx 988.67$, which is much lower than subjects think; less than 10% of subjects in their data locate the median below 2000.

In my experiment I frame this process as a "counter" that has a starting value at 10000 which increases or decreases in the same manner as in Ensthaler et al. (2015). Importantly, there is a threshold, set at 1000, which determines whether the product is faulty or not. Should the counter after 12 draws be below 1000, the product will be faulty and the add-on price ("repair fee") has to be paid. In addition, I try to create belief differences between buyers and sellers in the treatment condition by exposing sellers to extensive feedback about the counter. Specifically, each seller sees 20 samples of the evolution of the counter. Ups and downs are depicted by coloured arrows and each simulation clearly states whether - had this been a real round - the repair fee would have to be paid.[13] By clicking through these samples at the beginning of the experiment, subjects should realise that they might have had a wrong idea about the probability of the product being faulty and should revise it upwards. Indeed, Ensthaler et al. (2015) show that with a lot of detailed explanations, the bias can be largely eliminated. In the control condition, both buyers and

[13] I chose this method over simply giving subjects the true probability as I wanted to ensure that they appreciate their initially biased perception. I expected them to gain a better feeling for the situation of the buyers without feedback. Moreover, in both conditions, sellers face buyers that are potentially heterogeneous in their assessment of the probability which makes the two situations more comparable.

sellers get shown the feedback.

In each round, the seller thus chooses the price, p and the repair fee f and the buyer decides, upon observing the offer, whether to accept or to reject. The product has a value of x to the buyer. If the buyer rejects, both buyer and seller earn y, their outside option. In the experiment, $x = \pounds 25$ and $y = \pounds 5$. In addition to non-negativity constraints on p and f, I also impose that the sum of the two cannot be larger than x.

3.3.1 Beliefs

In each round, and also at the very beginning of the experiment, that is before participants are given feedback, I elicit the subjects' belief about the probability that the product is faulty. I also ask buyers (sellers) to guess the average belief of all sellers (buyers) in their group in a given period. For each elicitation task one out of the 21 guesses is chosen at the end, and subjects receive a payment of £2 if their guess is within 5 percentage points of the true value.[14] Apart from getting estimates about subjects' beliefs, these elicitations were also aimed at making subjects aware of potential discrepancies between their assessments before and after receiving feedback. The elicitation of second-order beliefs allows me to quantify the degree to which the subjects are aware of the

[14]Strictly speaking, this elicitation procedure is not incentive compatible, that is subjects may not always find it optimal to state their true belief. If subjects possess distribution over beliefs, they should report the midpoint of the interval of size 0.1 that contains the greatest probability mass. For all unimodal symmetric distributions, this will be the mean, and thus be incentive compatible. I chose this procedure over the well-known quadratic scoring rule (QSR), because it seemed the simplest possible to explain to subjects, who already had quite long instructions to read. (See Charness and Dufwenberg (2006) for a use of the same method) Furthermore, the QSR is not incentive compatible if subjects are risk averse, whereas for the rule I employ, risk aversion does not play a role. In addition, the aim was to keep the potential earnings small relative to the main decisions which would have made the QSR very flat for a large interval around the optimum. In any case, I am mainly interested in differences between beliefs in the control and the treatment and between buyers and sellers, which should further mitigate potential concerns, since the belief elicitation method is held constant across settings.

biased beliefs of others, something that is crucial for these belief differences to be exploited.[15] I can thus formulate the following hypotheses with respect to the belief data.

Prediction 3.1. *(a) In the treatment condition, buyers' first order beliefs about the likelihood of the product being faulty are lower than the sellers' first order beliefs. In the control condition there is no difference.*

(b) In the treatment condition, the sellers' second order beliefs (the beliefs about the buyers' first order beliefs) are lower than their own first order belief. For all other groups there is no difference.

3.3.2 Naive Buyers

In what follows, I will derive hypotheses for the case where (in the treatment condition) buyers have downward biased beliefs and are naive. Also, I will, in the main text, refrain from explicitly modelling social preferences. This is purely to keep notation concise; social preferences, for example in the form of inequity aversion as in Fehr and Schmidt (1999) may play an important role in influencing the buyers' acceptance decisions.[16] However, as I will explain intuitively below and formally in the appendix, accounting for social preferences does not substantially change the predictions. Moreover, I focus on the case where sellers have correct expectations about the buyers' beliefs (second order beliefs), in line with Prediction 1b.

[15] There would be reasons to assume that subjects might struggle to accurately predict the extent of the belief bias of others. Both "hindsight bias" (Fischhoff, 1975) and "curse of knowledge" (Camerer et al., 1989) might lead sellers with feedback to underappreciate the belief difference.

[16] In Fehr and Schmidt (1999), a player's utility when he receives material payoff X_i and his opponent receives X_j is given by

$$U(X_i, X_j) = X_i - \alpha \max\{X_j - X_i, 0\} - \beta \max\{X_i - X_j, 0\}$$

where $\alpha \geq \beta$. α captures the disutility from disadvantageous inequality, whereas β captures the disutility from advantageous inequality.

A buyer accepts the seller's offer if and only if $x-p-\pi^B f \geq y$ and I assume that if he is indifferent, the buyer accepts. Hence, the seller chooses p, f to maximise $p+\pi^S f$, taking into account the non-negativity constraints for p and f and for the buyer's expected payoff. The solution to this problem follows directly from Propositions 3.1 and 3.2 in the previous section. Whenever $\pi^S = \pi^B = \pi$, the maximum that a seller can earn is $x - y$. This can be achieved by setting $f = 0$ and $p = y - x$, which in the notation of the previous section corresponds to $g = (0, 0)$. Under common beliefs and if agents are risk neutral, sellers may use add-on prices $f > 0$, but importantly, they cannot achieve higher profits through that. Given the restrictions on p and f in my setting, the set of optimal contracts is thus given by $f \in [0, \frac{1}{1-\pi}y]$ and $p = x - y - \pi f$. However, if agents are risk averse, the optimal contract has $f = 0$. The appendix formalises this result, but the intuition is straightforward: since the expected value of the offer does not change, any offer with $f > 0$ will be evaluated strictly worse by buyer and seller because it only introduces risk.

In the case where the buyer's belief is biased downwards, the seller can use the fee f to extract more surplus. As buyer and seller disagree over the probability of the product being faulty, the buyer is willing to accept offers that he should not accept were he to hold correct beliefs. In the risk neutral case, the seller will set $f = \frac{y}{1-\pi^B}$ and $p = x - \frac{y}{1-\pi^B}$ to make the buyer indifferent between accepting and rejecting. This will yield an expected payoff of $x - \frac{1-\pi^S}{1-\pi^B}y > x - y$. In the terminology of section 3.2, this offer has $E^S(g) > 0$ and $E^B(g) = 0$ since the seller will offer a speculative bet on the likelihood with which the add-on price has to be paid. Compared to the case with $f = 0$, the seller reduces the price for the good but increases the add-on price such that the buyer receives the same expected payoff (evaluated at his subjective beliefs). Under risk aversion, agents will not always choose such extreme contracts and instead opt for those where $p + f \leq x$ is slack. Importantly, as the appendix shows, all optimal offers will still have $f > 0$. Thus, we can formulate the testable implications for this setting:

Prediction 3.2. *If buyers are naive, sellers in the treatment condition earn a higher expected profit, compared to sellers in the control condition. If agents are risk averse, sellers in the treatment condition will also set higher repair fees.*

In addition, consider the case mentioned above where agents have preferences over relative outcomes as in Fehr and Schmidt (1999) and are therefore negatively affected by unequal allocations.[17] As described above, a buyer with biased beliefs overestimates the subjective expected value of any offer with $f > 0$. Hence, for any given offer p, f, the (perceived) difference between buyer and seller payoff is decreasing in π^B. Therefore, inequity concerns on the buyer side do not eliminate the exploitation motive. Regarding the seller, as long as his aversion against advantageous inequality is not too large, he will choose offers that just meet the buyer's participation constraint, exploiting belief differences where possible.[18]

[17]Their model considers only preferences over deterministic allocations, and cannot without complications be extended into an expected utility framework (Fudenberg and Levine, 2012). The important conceptual issue is whether agents care about ex-ante or ex-post fairness, which in general leads to different preferences over outcomes. Brock et al. (2013) conduct experiments that can differentiate between the two and find that ex-ante fairness has a higher predictive power. I therefore, and because it is analytically more tractable, assume that agents compare expected payoffs.

[18]As in the prediction for the standard ultimatum game for agents with inequity averse preferences, once the seller (proposer) cares more about reducing inequality than about his own material payoff, he will always offer an equal split ($p = \frac{x}{2}$, $f = 0$). This means that such sellers would not want to exploit belief differences, but also that we can easily identify them in the data.

3.3.3 Sophisticated Buyers

Using the theory developed in section 3.2.2, I now analyse the case where buyers who receive no feedback and whose beliefs are downward biased may draw inferences about the true probability from the seller's offers. I will thus concentrate on analysing the treatment condition of the experiment. The control condition could be analysed in a similar manner, but there would be no additional insights. And, indeed the main aim of this section is, as above, to show that if buyers have biased beliefs but are sophisticated otherwise, they do not fare worse than buyers in the control condition.

To show this, I sketch the intuitive argument based on Proposition 3.2.2. A seller can always achieve a payoff of $x - y$ without using the add-on fee, by simply setting $p = x - y$ which the buyer will just accept. This is equal to the payoff that can be achieved with known, identical beliefs. If an equilibrium were to consist of offers with $f > 0$, the buyer will realise that only types with $\pi_i > \pi'$ such that $p + \pi' f = x - y$ would actually benefit from such an offer. But then, the buyer updates his belief to $\pi^B > \pi'$ and thus evaluates the offer as $x - p - \pi^B f$ which is less than y and will thus reject. This adverse selection logic applies to all offers with $f > 0$. The only exception is an equilibrium in which the highest type $i = N$ makes an offer with $f > 0$ that is unprofitable for any other type. The buyer, upon seeing such an offer, updates his belief to $\pi^B = \pi_N$, and accepts any offer that gives him at least y in expectation. Clearly, however, such an offer can never extract more surplus than $x - y$. Even in this knife-edge case, offering $f = 0$ continues to be an equilibrium. By similar arguments a in the analysis of the control condition, if agents are risk averse, offers with $f > 0$ are strictly dominated.

Prediction 3.3. *If buyers are sophisticated, sellers in the treatment condition*

	S, C	B, C	S, T	B, T
Initial belief	0.4046	0.3676	0.4188	0.4034
First-order belief	0.5525	0.5265	0.5277	0.4038
Second-order belief	0.5566	0.5354	0.4830	0.4508
Price	14.34		13.67	
Price (accepted only)	13.58		13.19	
Fee	3.09		3.69	
Fee (accepted only)	2.94		3.19	
Exp. value of offer (av. stated belief)	16.03	9.02	15.62	9.85
Exp. value of offer (true probability)	16.23	8.76	15.93	9.07
Exp. Profit (av. stated belief)	12.52	8.62	11.90	8.89
Exp. Profit (true probability)	12.64	8.39	12.08	8.39
Exp. Profit (true probability, accepted only)	15.39	9.61	15.14	9.86
Acceptance Frequency	0.735		0.700	

S=Seller, B=Buyer, T=Treatment, C=Control

Table 3.1: Add-On Price Experiment: Summary Table of Group Averages

do not earn higher profits than sellers in the control condition. If agents are risk averse, no add-on prices will be used.

As before, the case with Fehr-Schmidt inequity aversion is analysed formally in the appendix, but the same basic logic carries over. The key message behind the prediction is thus that sophisticated buyers realise the danger of exploitation via the fee and reject such offers. Also, observe that since all seller types (with the possible exception of the highest type) set $f = 0$, buyers, even though sophisticated, are not predicted to learn the true belief during the experiment.

3.4 Add-On Price Experiment — Results

In Table 3.1, I present descriptive statistics of all the main outcome variables of the two experiments. I start by analysing the belief data, and then study the outcome variables such as earnings of buyers and sellers as well as offers made.

For the belief data, I estimate the following random-effects specification:

$$\pi_{it} = \alpha + \beta B_i + \boldsymbol{z}_{it}\boldsymbol{\gamma} + u_i + \varepsilon_{it} \tag{3.2}$$

Here π_{it} denotes the (first-order) belief stated by subject i in period t. B_i is a dummy variable equal to one if the subject is a buyer, and zero for sellers. Hence, β is the main coefficient of interest, denoting the difference in beliefs between buyers and sellers. In some specifications, I also include period dummies and individual characteristics in $\boldsymbol{z}_{it}$. u_i is the unobserved individual heterogeneity, assumed to be uncorrelated with (the randomly assigned) B_i and $\boldsymbol{z}_{it}$. The disturbances ε_{it} are assumed to be iid over t and i.[19]

In order to test for treatment differences in outcomes such as profits, offers, or prices, for sellers or buyers comparing treatment and control condition, I use the following individual random effects regression:

$$o_{it} = \alpha + \beta T_i + \boldsymbol{z}_{it}\boldsymbol{\gamma} + u_i + \varepsilon_{it} \tag{3.3}$$

Here $o_{it} \in \{p_{it}, f_{it}, Seller\_Profits_{it}, Buyer\_Earnings_{it}\}$ denotes the outcome of interest and T_i is the treatment indicator. The other variables are equivalent to the ones used in the belief regressions.[20] In both regressions, standard errors are clustered at the matching group level to account for interdependencies among subjects who interact with each other. Subjects were fully informed about the matching protocol, it is therefore unlikely that they thought they were (in)directly interacting with more than the seven other members in their

[19]For the belief regressions, the individual characteristics include a dummy for gender, for whether the individual took (or is taking) an economics or mathematics course during their study, and dummies for their performance in the cognitive reflection test (0 to 3 correct answers).

[20]$\boldsymbol{z}_{it}$ now also includes the subject's risk aversion (measured by their switching point in the multiple-price list task) and four dummies representing the choice in each of the four allocation decisions administered to measure social preferences.

	Belief, Treatment			Belief, Control		
	(1)	(2)	(3)	(4)	(5)	(6)
Buyer	-0.124***	-0.124***	-0.125***	-0.0260	-0.0260	-0.0191
	(0.0372)	(0.0376)	(0.0479)	(0.0517)	(0.0522)	(0.0489)
Constant	0.528***	0.502***	0.490***	0.552***	0.526***	0.517***
	(0.0171)	(0.0150)	(0.0315)	(0.0335)	(0.0541)	(0.108)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.2) separately for the treatment (columns 1-3) and control (columns 4-6). Period controls consist of a dummy variable for each of the periods 2-20. Individual controls include dummy variables for gender, knowledge in economics and mathematics, and cognitive reflection. See footnote 19 and section B.2.3 in the appendix for more details.

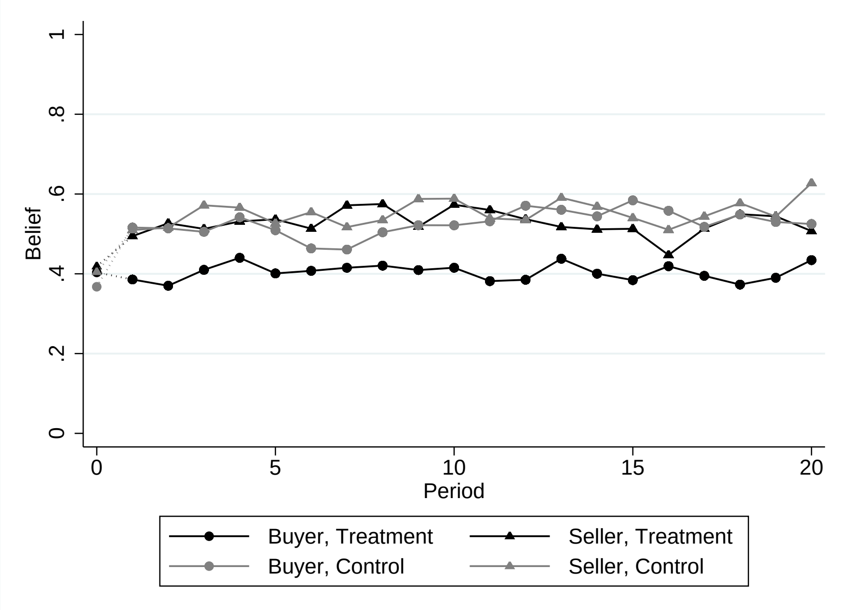
Table 3.2: Add-On Price Experiment: Subjects' Beliefs

group.

3.4.1 Beliefs

I first analyse the data on beliefs from the add-on price experiments. Remember from above that the aim behind employing the stochastic process was to endow buyers in the treatment condition with substantially downward-biased beliefs compared to sellers in the treatment, while there should be no difference in the control condition. Figure 3.1 and columns (1)-(3) in Table 3.2 provide evidence that this indeed is the case. Over all 20 periods, buyers in the treatment condition hold an average belief that is about 0.124 ($p<0.01$) lower than the belief of the sellers they interact with. In the control condition, there is no significant difference between buyers and sellers (Table 3.2, columns (4)-(6)). We can therefore conclude that there are meaningful belief differences which is a necessary condition for exploitative contracts to emerge.[21]

[21] It is worth mentioning that the beliefs of all three groups that received feedback are still below the true probability of 0.6128. One reason for this could be that there is a natural



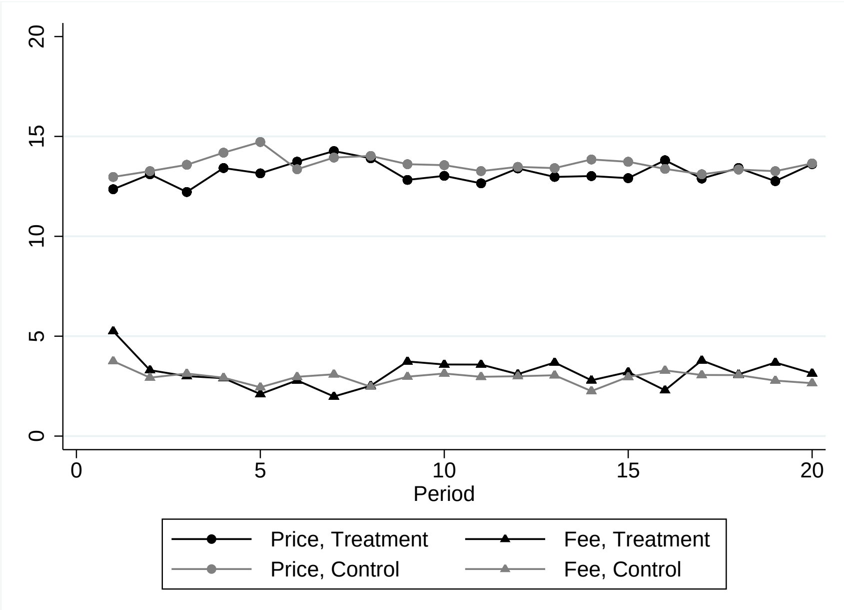
This figure plots the average values of buyer and seller beliefs in each period, separately for the treatment and the control condition. The values corresponding to "period 0" are the beliefs elicited at the start of the experiment before anyone got the feedback in the form of the simulations.

Figure 3.1: Add-On Price Experiment - Beliefs

Result 3.1. *Buyers in the treatment condition hold beliefs that —on average— are significantly lower than sellers' beliefs. In the control condition, there is no significant difference between buyers and sellers.*

In addition, we can investigate to what extent sellers in the treatment condition are able to predict that the buyers they interact with will hold downward biased beliefs. On average, these sellers predict that their own belief is about 0.045 higher than the buyers' ($p<0.1$, t-test, one-sided). This effect is only weakly significant, and there is considerable heterogeneity in subjects. 14 out of 24 sellers have an average difference in beliefs that is positive. There thus seems to be some indication that sellers are to some extent subject to a "curse of knowledge" (Camerer et al., 1989); a substantial fraction fails to realise that buyers might be biased.

tendency for subjects to state beliefs close to 0.5.



This figure plots the average values of accepted prices and fees in each period, separately for the treatment and the control condition.

Figure 3.2: Add-On Price Experiment - Accepted Offers

Figure 3.1 also contains a period 0 which plots the initial belief that subjects reported before receiving feedback. We can see that at the start all group averages are very similar and that therefore the feedback is responsible for the debiasing of subjects. Buyers in the treatment condition on average state an initial belief that does not differ from their average belief over the 20 rounds, whereas all other groups' average belief is significantly higher than their starting belief ($p<0.05$, t-test, one-sided).

3.4.2 Profits

Given these belief differences, the question is how they influence market outcomes. On average, sellers in the control condition set prices that are £0.67 higher and fees that are £0.60 lower, but neither difference is significant. Similarly, when only focusing on accepted offers, prices in the control are £0.39

	Seller Profits (true probability)			Seller Profits (stated belief)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.561	-0.561	-0.352	-0.622	-0.622	-0.355
	(0.582)	(0.588)	(0.516)	(0.595)	(0.601)	(0.527)
Constant	12.64***	12.22***	11.01***	12.52***	12.00***	10.41***
	(0.473)	(1.047)	(1.162)	(0.491)	(1.041)	(1.106)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with seller profits as the dependent variable. Columns 1-3 calculate the expected value of seller profits using the true probability, whereas columns 4-6 use each seller's average stated belief. Period controls consist of a dummy variable for each of the periods 2-20. Individual controls include dummy variables for gender, knowledge in economics and mathematics, and cognitive reflection (footnote 19), as well as subjects' risk aversion and measures of social preferences (footnote 20). See section B.2.3 in the appendix for more details.

Table 3.3: Add-On Price Experiment: Seller Profits

higher, whereas fees are £0.25 lower, which again is not significant.[22] This result is illustrated by Figure 3.2, which plots per period averages of the accepted offers. Moreover, acceptance rates do not differ significantly between the treatment and control condition either. In the control condition 73.5% of offers are accepted, whereas in the treatment condition this is the case for 70.0% of offers.

It then follows that sellers do not earn higher profits from buyers with biased beliefs. To show this, I calculate the expected profit of a seller i in period t as $p_{it} + \pi f_{it}$. For columns (1)-(3) in Table 3.3, I use the true expected probability $\pi = 0.6128$. In columns (4)-(6), I use, for each seller, his average stated probability, $\bar{\pi}_i$, over all 20 rounds instead. Given that the belief is quite volatile (see Figure 3.1) it seems unlikely that these period to period differences

[22]Tables B.2 and B.3 in the appendix contain the corresponding regression results. Note, however, that once including the full set of individual controls, there is a statistically significant treatment effect for the fee ($\beta = 1.537, p < 0.01$)

reflect true changes in beliefs. Rather, it could be attempts of subjects to game the belief elicitation mechanism across rounds, in which case the average would be a more accurate representation of their true belief. In neither case, there is any significant treatment effect, and there is thus no evidence for exploitation.[23] It is even the case that sellers in the treatment condition make slightly lower profits (by about £0.56) but this effect is not significant. Also, the fact that not all sellers in the treatment condition correctly predict the biased beliefs does not automatically mean that we should not see exploitation in the data. We would predict that then at least those sellers that on average predict a positive belief difference earn higher profits than those sellers who do not. However this is not the case. On the contrary, those sellers in the treatment condition who correctly predict the sign of the belief difference earn £0.49 less ($p>0.1$) than the remaining sellers in the treatment condition.

Result 3.2. *Sellers in the treatment condition do not earn higher profits than sellers in the control condition. There is no evidence for exploitation of biased beliefs.*

3.4.3 Buyer Behaviour

It it thus important to understand whether this result is driven by sellers simply being unable to exploit the buyers due to some cognitive limitations, or whether buyers are sophisticated enough to protect themselves from exploitation. We note first that the data shows more positive fees than we would expect, both for the control condition and the treatment condition in the case of sophisticated buyers. While it is unclear what drives this, it allows us to study in more detail the acceptance behaviour of buyers. Because we observe a large variation in the

[23] These results do not change if we use realised profits as the dependent variable, or π_{it}, the stated belief in each round.

fees set, it is possible to see how acceptance behaviour differs across treatment and control, taking into account the variation in the offers.

As a first piece of evidence, we can compare acceptance rates for high fees. While in the full sample there is no significant difference, in the control condition 71.4% of offers with $f \geq 5$ are accepted, but in the treatment condition, this is only the case for 53.2% of offers (p<0.01, Fisher exact test). This is despite the fact that also for the subsample with $f \geq 5$, neither offers nor prices are significantly different between the treatment and the control condition.[24] The buyer's expected value of the offer (using the stated belief) is £0.62 lower (p>0.1) in the control condition, and the price p is £0.37 lower (p>0.1).[25]

Second, I estimate a random effects probit model of the following form:

$$accept^*_{it} = \alpha + \beta T_i + \gamma ES_{it} + u_i + \varepsilon_{it} \tag{3.4}$$

Here, $accept^*_{it}$ is the latent variable underlying the buyer's decision whether to accept a given offer. I assume that if $accept^*_{it} \geq 0$, buyer i accepts the offer in period t, hence $accept_{it} = 1$ in this case and $accept_{it} = 0$ otherwise. T_i is the indicator for the treatment condition, and ES_{it} denotes the subjective expected surplus of a given to the buyer, calculated as $x - p_{it} - \bar{\pi}_i f_{it}$. As before, I use the average belief over all periods for each buyer. The appendix contains additional estimates of this specification that include the controls as specified in footnotes 19 and 20. The coefficient of interest is β which measures any differences in acceptance behaviour between treatments, when controlling for the expected surplus.

[24]The choice of $f \geq 5$ corresponds to the choice made in the probit regression below. There, splitting the sample in offers with low fees ($f \leq 1$) and high fees ($f \geq 5$) led to similar sized subsamples and thus facilitated comparison. The result is robust to other cutoffs, such as 4.5, 6 and 7.

[25]The p-values come from estimating (3.3) on the subsample with $f \geq 5$ (not reported in the paper).

	accept	accept $f \leq 1$	accept $f \geq 5$
	(1)	(2)	(3)
Coefficients			
Treatment	-0.555***	-0.809	-0.926***
	(0.203)	(0.762)	(0.248)
Offer	0.376***	0.730***	0.342***
	(0.0712)	(0.0971)	(0.0635)
Constant	-2.418***	-5.050***	-2.144***
	(0.667)	(1.088)	(0.524)
Marginal Effects			
Treatment	-0.126***	-0.0785	-0.237***
	(0.0440)	(0.0714)	(0.0468)
Offer	0.0852***	0.0709***	0.0874***
	(0.0109)	(0.0189)	(0.00934)
Observations	960	299	245

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows the results of the regression in (3.4). Column 2 only considers the subsample of offers with $f \leq 1$ and column 3 only considers the subsample of offer with $f \leq 5$.

Table 3.4: Add-On Price Experiment - Buyer Acceptance Decision

Table 3.4 shows the coefficients and marginal effects of this regression. The coefficient on the treatment dummy shows that buyers in the treatment condition are 12.6% less likely to accept a given offer after taking into account their beliefs. It thus seems that buyers who did not get feedback show a general predisposition to be more likely to reject a given offer than buyers who got feedback. Columns (2) and (3) in Table 3.4 show the same probit regressions as before, but focus on the subsamples with $f \leq 1$ and $f \geq 5$. We see that the treatment effect on the acceptance decision becomes insignificant for low fees, and becomes much larger for very high fees, in line with this explanation.

The buyers' increased reluctance in the treatment condition to accept sellers' offers is thus only present when there could be meaningful exploitation.[26]

Result 3.3. *Buyers in the treatment condition are more likely to reject an offer with the same subjective expected surplus, especially when these offers consist of high add-on fees.*

Hence, both the regression results as well as the direct comparison of acceptance rates, provide some evidence for sophistication on the side of the buyers. They seem to understand the basic adverse selection logic in that they realise the potential exploitative effect of high fees.[27]

However, it is also conceivable that this effect is due to risk aversion and/or some form of ambiguity aversion on the buyers' side. From the theoretical analysis in sections 3.2 and 3.3, it is clear that in order to exploit belief differences, sellers need to offer risky contracts. If buyers in the treatment condition are more averse to the risk or uncertainty of these contracts than the buyers in the control condition, they might be more likely to reject them, also leading to the behaviour documented above. With respect to risk aversion, Table B.10 shows that buyers in the treatment condition are more risk averse (as measured by the price-list questions) than buyers in the control condition. This might in principle explain the difference in buyer behaviour across treatments, however, when adding the measure of risk aversion as an additional covariate to the regression in (3.4), it does not have a systematic effect on acceptance behaviour,

[26] Table B.1 in the appendix summarises the group averages of the variables in the pilot experiment. The overall picture is relatively comparable, with the exception of the (unincentivised) belief data. The fee tends to be higher in the treatment condition, and the acceptance rates differ significantly, even on the whole sample which echoes the sophistication explanation from above. See section B.2.1 for more details.

[27] A third aspect of the data that is relevant are the second order beliefs of the buyers. Somewhat surprisingly, they (correctly) predict that on average sellers have a higher belief than themselves. It is unclear why they do not update their beliefs accordingly, given that they know that the sellers received the feedback, but at least some buyers seem to entertain the possibility that the sellers' belief may differ.

as can be seen from Table B.5 in the appendix.

Moreover, buyers in the treatment condition might be less certain in their assessment of the probability of the product being faulty than buyers in the control condition. Put differently, since they do not receive the feedback in the form of the simulations, they might perceive the probability associated with having to pay the add-on price as relatively more ambiguous. The aversion to such ambiguity that has been reported in a number of studies (see, for example, Halevy (2007); Abdellaoui et al. (2011) as recent examples or Camerer and Weber (1992) for a survey of earlier evidence) might lead buyers to use a more pessimistic belief to evaluate the offer. This would lead to higher rejection rates in the treatment condition, but need not be reflected in the belief data.

In order to parse these two competing explanations for the empirical findings, the insurance experiment creates a setting in which the exploitative offer for naive buyers does not involve any uncertainty over final payoffs. If buyers in the treatment condition tend to choose the safe option (rejection in the add-on experiment) because of the potentially higher risk or ambiguity of the environment rather than their sophistication, such behaviour should make them susceptible to exploitation in the insurance experiment. The next section describes the setting and predictions in more detail.

3.5 Insurance Experiment — Design and Predictions

The insurance experiment differs from the add-on price experiment in the following way: As before, the product that the buyer can buy will be faulty with probability π. Now, however, whenever this happens, the buyer incurs a fixed loss unless he bought insurance from the seller, together with the product.

Again, the experimental protocol aims to induce different beliefs for buyers and sellers about π in the treatment condition, using the paradigm of participants misunderstanding the multiplicative growth processes. The seller thus makes a take-it-or-leave-it offer, consisting of a price for the product, p and an insurance fee, f. The buyer now has to decide between rejecting the offer altogether, buying the product only, or buying the product plus insurance. In the last case, should the product turn out to be faulty, the seller incurs the loss.

The specific experimental parameters are set such that product is worth $x = \pounds 35$, the potential loss is $L = \pounds 10$ and if the buyer rejects the offer altogether, both buyer and seller receive $y = \pounds 0$. I also impose that the price the seller can set is between £10 and £20 and the insurance fee between £0 and £10. These restrictions ensure that no party can lose money, but they also impose a relatively strict condition on the maximum price relative to the value of the product. The main motivation for this is that, on the one hand, the setting closely matches the previous one (and thus there is an option of not buying the product, as opposed to only deciding about insurance). On the other hand, I want to make it sufficiently attractive for buyers with biased beliefs to buy the product and thus make insurance the important margin for exploitation. The role that this maximal price, $p^{max} = \pounds 20$, plays for the predictions will be explained further below.

3.5.1 Beliefs

The process that determines whether the product is faulty or not, also has a very similar structure as before. Now, the aim is to set up the process such that subjects are expected to overestimate the probability of the product being faulty. To achieve that, starting at 10000, the counter can now 24 times go either up or down, and, as before, an increase by 70% and a decrease by 60%

are equally likely. Whenever the counter ends up at a value above 8000, the product will be faulty. By the same logic as above, the counter will with a high probability take on very low values. For example, the median of the implied distribution of final values will now be at 97.61, and the true probability that the product is faulty is 0.0758. With respect to the beliefs which are elicited exactly in the same manner as before, we can thus formulate the following predictions.

Prediction 3.4. *(a) In the treatment condition, buyers' first order beliefs about the likelihood of the product being faulty are higher than the sellers' first order beliefs. In the control condition there is no difference.*

(b) In the treatment condition, the sellers' second order beliefs (the beliefs about the buyers' first order beliefs) are higher than their own first order beliefs. For all other groups there is no difference.

3.5.2 Naive Buyers

I start by analysing the case where buyers are naive. In line with the belief predictions, they now overestimate the likelihood of the product being faulty. The buyer has to choose between the three options, each yielding the following expected payoff:

$$\begin{aligned} u^B(\textit{product + insurance}) &= x - p - f \\ u^B(\textit{product only}) &= x - p - \pi^B L \\ u^B(\textit{reject}) &= y \end{aligned}$$

It is clear that under risk neutrality the buyer prefers buying insurance over buying only the product if $f \leq \pi^B L$, that is, if the insurance fee is no larger than the expected loss. The buyer will only reject the offer if $y > x - p -$

$\min\{f, \pi^B L\}$. By design, there is a price p^{max} which is the maximum price that any seller can charge for the product. Specifically, in the experiment $p^{max} < x - y - L$ and therefore a buyer always achieves a higher payoff through buying the product than rejecting the offer as a whole. In this case, the margin that is important here for the role of belief differences is the insurance fee f which then determines whether the buyer buys the product with or without insurance. The seller's expected payoff is given by

$$\begin{aligned} u^S(\textit{product + insurance}) &= p + f - \pi^S L \\ u^S(\textit{product only}) &= p \\ u^S(\textit{reject}) &= y \end{aligned}$$

This means that if $p = p^{max}$, the seller will charge $f \geq \pi^S L$ as the insurance fee. Hence, if $\pi^S = \pi^B = \pi$, that is if buyers and sellers both hold the correct belief, the only price at which insurance can be sold is $f = \pi L$. Sellers do not earn any additional profit from selling insurance. In line with the theoretical exposition in section 3.2, the transfers induced by selling or buying insurance will not change expected payoffs.

However, once we consider naive buyers with upward-biased beliefs $\pi^B > \pi^S$, sellers can charge a price for insurance that is higher than the expected loss. In the terminology of section 3.2, the offer of insurance now constitutes the speculative contract that the seller offers to the buyer. This is because for the seller it is profitable to offer an insurance contract with $E_S(g) > 0$. For an appropriately chosen g, the buyer with biased beliefs perceives such an offer as having $E_B(g) \leq 0$, because he overestimates the likelihood with which the product is faulty. The seller will exploit this by setting $f = \pi^B L$ and earn additional profit from the biased beliefs given by $(\pi^B - \pi^S)L$.

For the motivation of this experiment it is important to notice that a buyer who buys insurance receives a sure payoff of $x-p-f$. Hence, contrary to the add-on experiment a buyer who perceives the risk of buying a product that is faulty as ambiguous, would be more willing to buy insurance because it allows him to "sell" the ambiguous part of the contract to the seller. More formally, a buyer would evaluate buying the product without insurance in a more pessimistic way. According to the min-max expected utility (MEU) approach by Gilboa and Schmeidler (1989), or the α-MEU approach by Ghirardato et al. (2004), a buyer overweights the likelihood of the product being faulty which increases the willingness to pay for insurance.[28] Of course, because of random assignment of roles, sellers might have similar ambiguity-averse preferences and would therefore demand a higher price for insurance. But since buyers in the treatment condition do not receive feedback about the process determining π, they arguably perceive the likelihood of the product being faulty as more ambiguous than subjects in the other three roles. Hence, we can formulate the following prediction:

Prediction 3.5. *If buyers are naive, sellers in the treatment condition will earn a higher expected profit, compared to sellers in the control condition. In the treatment condition, buyers will always buy insurance.*

As for the add-on experiment, I show formally in the appendix that allowing for risk aversion and inequity aversion does not change this prediction. To see this intuitively, notice that risk aversion increases the demand for insurance on the buyer side and increases the fee at which the seller is willing to offer

[28]Loosely speaking, in MEU the individual adopts the most pessimistic belief among all possible beliefs. Hence, in the most extreme case a MEU buyer considers the expected loss being given by L. A α-MEU buyer evaluates the expected loss as an α-weighted average between the best and worst outcome, thus αL. For $\alpha > \pi^B$ the individual is ambiguity averse in our setting and in both the MEU and α-MEU model he would be willing to pay even more than $\pi^B L$ for insurance.

insurance, but does not affect the logic behind the prediction. Upward-biased beliefs continue to make buyers in the treatment condition more willing to pay for insurance.[29] Inequity aversion (as modelled in Fehr and Schmidt (1999), see footnote 16) influences a buyer's decision in so far as he might prefer to reject a seller's offer to buy the product even if he obtains a positive payoff but is adversely affected by the resulting payoff inequality. While this requires the seller to reduce the price for the product (in both the treatment and the control condition) in order to continue to sell to the buyer, this does not affect the exploitation possibilities. The optimal price for product and insurance will be higher when buyers have biased beliefs. A buyer with upward-biased beliefs puts more value on the possibility of insuring himself against the possible loss, but also evaluates relative payoffs with respect to this belief and therefore perceives the seller to be worse off than he actually is from taking over the risk.

The next section analyses how the predictions change if buyers are sophisticated. As will become clear, sophistication together with inequity aversion can lead to non-trivial predictions for the experimental setting.

3.5.3 Sophisticated Buyers

As before, using the theoretical argument developed in section 3.2.2, I analyse the case of a sophisticated buyer who treats the situation as a game of asymmetric information, where the true probability corresponds to one of many possible seller types. The following prediction confirms that—unlike in the

[29] As I discuss in the appendix in more detail, risk aversion can, however, lead to insurance being sold at fees that are strictly better for at least one side. The reason is that for preferences that are not of the constant absolute risk aversion (CARA) form the willingness to pay for insurance depends on the wealth of the agents. Since the seller typically earns more than the buyer from the sale of the product itself, decreasing absolute risk aversion (DARA) would predict that buyers value having insurance more than sellers.

naive case—buyers will not be exploited when they are sophisticated:

Prediction 3.6. *If buyers are sophisticated, sellers in the treatment condition do not earn higher profits than sellers in the control condition. Insurance will be sold with probability less than one.*

Clearly, if p is set at p^{max} and a buyer would buy regardless of his belief (as would be the case if he only cared about material payoff), any additional profit for sellers in the treatment condition would have to come from selling insurance at higher prices. In the control condition where the buyer knows the seller's type, insurance cannot be sold at fees above $\pi^S L$, because the assessment of the probability does not differ between buyers and sellers. Knowing that no seller would offer insurance at a price below $\pi^S L$, a buyer has to conclude that any insurance offer f comes from types with $\pi_i \leq \pi'$ where $\pi' = \frac{f}{L}$. But then, his updated belief π^B is smaller than π' and based on this he never accepts such an offer. Again, the only case where a buyer is indifferent between buying insurance, and therefore buying insurance could be part of an equilibrium, is at a price of $f = \pi_1 L$. No other seller but the lowest type would make such an offer and therefore the buyer's belief would be given by $\pi^B = \pi_1$.

However, this case relies on the assumption that buyers always buy the product at p^{max} and at all lower prices. If buyers attach sufficient weight on relative payoffs (Fehr and Schmidt, 1999) this may no longer be the case. If for some belief π^B, a buyer does not buy insurance, his (expected) payoff is given by $x - p - \pi^B L$ and the seller earns p. While in the experiment the buyer's payoff is strictly positive for all $p \leq p^{max}$, once the buyer's utility is negatively affected by the fact that for prices close to p^{max} the seller will earn more than him, he may prefer to reject the offer. More formally, using the notation of Fehr and Schmidt (1999), as introduced in footnote 16, a buyer prefers to accept an offer p only if $p \leq \frac{1+\alpha}{1+2\alpha}(x - \pi^B L)$, where α denotes the weight of relative payoffs in the utility function. Hence, if $p^{max} > \frac{1+\alpha}{1+2\alpha}(x - \pi^B L)$, the equilibrium from above where every type of seller sets $p = p^{max}$ will no longer exist.

The appendix formally establishes how the equilibria can be characterised instead and I provide the necessary intuition here. Most importantly, the main message from the previous analysis of sophisticated buyers carries over: in equilibrium there is no exploitation and sellers are predicted to earn the same as when buyer and seller have the same belief about π. Interestingly, insurance is predicted to be sold in equilibrium, but only with a probability strictly between one half and one, because buyers play a mixed strategy over the decision whether to buy insurance. Intuitively, this can be seen as follows: from above, first instinct suggests an equilibrium where no insurance is sold and all sellers set $p = \frac{1+\alpha}{1+2\alpha}(x - \pi^B L) < p^{max}$ which is the maximum price buyers are willing to accept. However, now the lowest type seller can deviate from this by offering insurance at an extremely low fee and charge $p > \frac{1+\alpha}{1+2\alpha}(x - \pi^B L)$ such that the buyer, for any belief, must prefer to buy insurance. Since the low fee offered leads the buyer to (correctly) revise his belief downwards, he is willing to pay more for the product, making such a deviation profitable for the seller. By the same logic, an equilibrium where insurance will always be sold (like in the naive case) will not exist either. The highest type seller would have an incentive to deviate and not sell insurance.

I can then show that in any equilibrium (there exists a pooling as well as a separating equilibrium) of this game, buyers will randomise over buying insurance. In a pooling equilibrium, buyers receive the same offers as naive buyers with an upward biased belief, but they do not always buy insurance. Compared to the control condition, sophisticated buyers pay more for insurance but are compensated by lower prices for the product.[30]

To summarise, the insurance experiment—like the add-on price experiment—delivers sharply different predictions depending on whether sellers face naive or sophisticated buyers. Since only naive buyers will be exploited by sellers, the insurance experiment allows us to investigate further the role of sophistication of buyers and how this might prevent exploitation. Using a setup where the exploitative offer does not come with any uncertainty over final payoffs for the buyer can give direct insights into whether the result in the add-on price experiment is indeed driven by sophistication. Replicating the no-exploitation result would rule out ambiguity aversion as an alternative explanation of the previous results. In addition, there now is a separate decision between the base product and the add-on which allows a separate investigation into the profitability of the add-on component. In many real-world market add-ons (like insurance, extra services, or upgrades) are indeed sold separately, and the experimental setting reflects this. Finally, the changes made to the multiplicative growth process are expected to lead to even more pronounced belief differences, including differences in second order beliefs, than before. Replicating the no-exploitation result would help alleviate concerns that too small belief differences may have affected the results.

3.6 Insurance Experiment — Results

The analysis of the results follows similar steps as for the add-on price experiment. Table 3.5 provides a summary of the main outcome variables of the

[30]In equilibrium the price is given by $p = \frac{1+\alpha}{1+2\alpha}(x - \pi^B L)$ and the insurance fee by $f = \pi^B L$, whereas in the control condition price and fee can be found by replacing π^B with π^S. Buyers will always buy the product, and buy insurance with probability $\frac{1+\alpha}{1+2\alpha}$. Then, the seller earns $\frac{1+\alpha}{1+2\alpha}(x - \pi^S L)$ in expectation, exactly what he is expected to earn in the control condition when buyers are inequity averse. See section B.1.4 in the appendix for details.

	S, C	B, C	S, T	B, T
Initial belief	0.3909	0.4576	0.3613	0.4614
First-order belief	0.1461	0.2222	0.1837	0.4634
Second-order belief	0.1672	0.2635	0.3024	0.4666
Price	17.79		17.20	
Price (accepted only)	17.74		17.12	
Fee	3.48		4.05	
Fee (accepted only)	2.33		3.36	
Exp. Insurance Profit (av. stated belief)	0.30	-0.04	0.89	0.78
Exp. Insurance Profit (true probability)	0.51	-0.51	1.44	-1.44
Exp. Total Profit (av. stated belief)	16.93	14.11	16.41	12.86
Exp. Total Profit (true probability)	17.15	14.96	16.96	14.07
Exp. Total Profit (true probability, accepted only)	18.29	15.95	18.71	15.53
Acceptance Frequency	0.938		0.906	
Insurance Frequency (cond. on acceptance)	0.349		0.611	

S=Seller, B=Buyer, T=Treatment, C=Control

Table 3.5: Insurance Experiment: Summary Table of Group Averages

experiment. The regression analysis is done by estimating equations (3.2) and (3.3).

3.6.1 Beliefs

Again, I start the analysis by looking at the belief data. In contrast to the add-on experiment, the stochastic process is now set up with the aim to create upward biased beliefs of buyers in the treatment condition, so that exploitation would happen via inflated insurance fees. Table 3.6 and Figure 3.3 confirm that beliefs are biased upwards and also show that the differences are more pronounced than before. Buyers in the treatment condition hold beliefs that are on average 0.280 higher than the beliefs of the sellers they interact with ($p<0.01$). Beliefs in the control treatment show some differences that cannot be explained by the feedback manipulation: buyers' average belief in the control condition is about 0.222, but they face sellers whose beliefs are on average about 0.076 lower ($p<0.05$). As before, despite the feedback, subjects

	Belief, Treatment			Belief, Control		
	(1)	(2)	(3)	(4)	(5)	(6)
Buyer	0.280***	0.280***	0.251***	0.0761**	0.0761**	0.0714*
	(0.0472)	(0.0477)	(0.0632)	(0.0337)	(0.0340)	(0.0385)
Constant	0.184***	0.216***	0.219***	0.146***	0.202***	0.277***
	(0.0343)	(0.0434)	(0.0505)	(0.0160)	(0.0176)	(0.0486)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.2) separately for the treatment (columns 1-3) and control (columns 4-6). The period and individual controls are described in table 3.2.

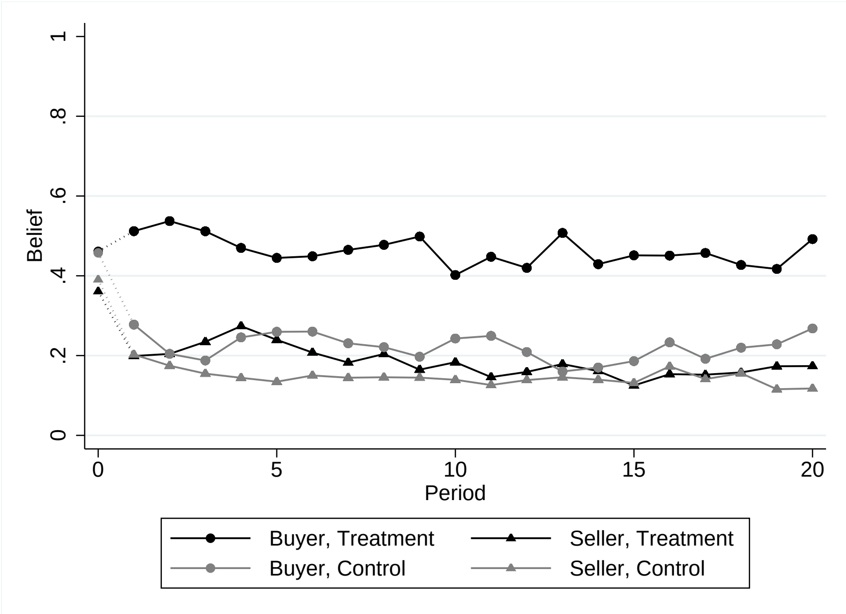
Table 3.6: Insurance Experiment: Subjects' Beliefs

on average still do not get the probability fully correct (they are upward-biased themselves), though as before we note that this does not affect the predictions as the differences and not the levels are important.

Result 3.4. *Buyers in the treatment condition hold beliefs that —on average— are significantly higher than sellers' beliefs. In the control condition, there is a difference in the same direction, but in much smaller magnitude.*

Moreover, we can check whether sellers correctly anticipate that buyers have upward biased beliefs by looking at the difference between first and second order beliefs. We can establish a stronger result than in the add-on price experiment: 21 out of 24 sellers predict the correct sign of the buyers' bias, however, they predict the belief difference to be on average 0.119 (p<0.01, t-test), whereas in reality it is more than double.[31] Also, we can see in Figure 3.3 that, as before in the add-on price experiment, the belief difference can be attributed to the provision of feedback. For all groups but the buyers in

[31]Somewhat puzzlingly, buyers in the control treatment believe that sellers hold a significantly higher belief than themselves (by 0.041, p<0.05, t-test), which is the opposite of what actual beliefs are.



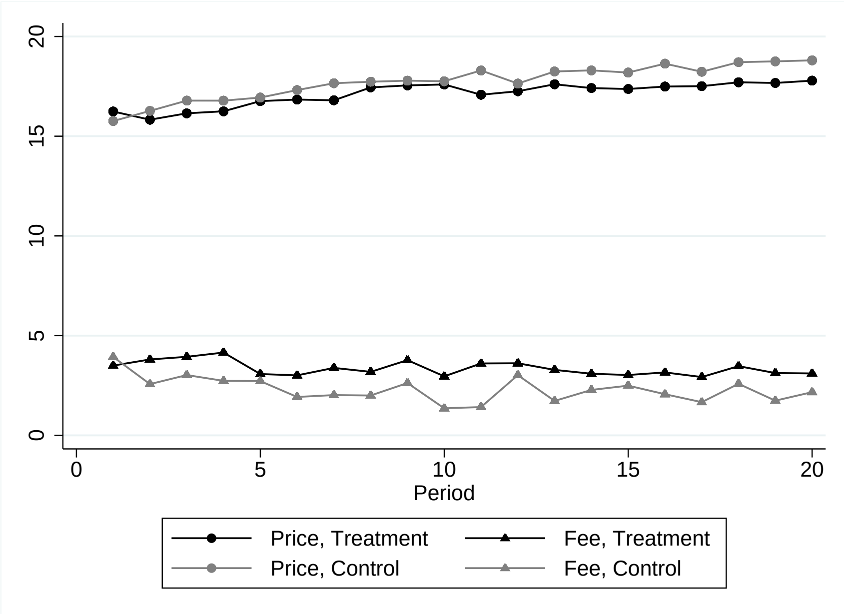
This figure plots the average values of buyer and seller beliefs in each period, separately for the treatment and the control condition. The values corresponding to "period 0" are the beliefs elicited at the start of the experiment before anyone got the feedback in the form of the simulations.

Figure 3.3: Insurance Experiment - Beliefs

the treatment condition, starting beliefs are significantly higher than average beliefs over the course of the experiment ($p<0.01$, t-test, one-sided).

3.6.2 Profits

Moving on to the subjects' decisions on the offers, I start by focusing on the role of insurance. Whereas in the treatment condition 61.1% of accepted offers include insurance, only 34.9% of offers in the control condition do. Buyers pay £0.82 more in the treatment condition ($p<0.05$, see Table B.6 the appendix for the full regression output, including individual controls), and Figure 3.4 visualises this over the 20 rounds of the experiment. Buyers in both conditions who buy insurance pay significantly more for insurance than the expected loss of 0.758. As established before, there is a difference of about 0.076 in stated beliefs between buyers and sellers, even in the control condition. This belief



This figure plots the average values of accepted prices and fees in each period, separately for the treatment and the control condition. The averages for the fees only include those offers where the buyer bought insurance.

Figure 3.4: Insurance Experiment - Accepted Offers

difference could be a possible explanation why we see some insurance also being sold in the control condition.[32]

Columns (1)-(3) in Table 3.7 show that these results imply higher profits made from selling insurance in the treatment condition. Sellers' profits from insurance are calculated as $(f_{it} - 10\pi) * \boldsymbol{I}_{it}$, where $\boldsymbol{I}_{it}$=1 if buyer i bought insurance in period t and zero otherwise. π denotes the true probability, but the results do not change when using the stated (average) belief $\bar{\pi}_i$ (see Table B.8 in the appendix for the corresponding regression output). Sellers make almost three times more money from selling insurance in the treatment condition.

Hence, buyers' biased beliefs have a substantial effect on market outcomes, as indicated by the much higher proportion of insurance being sold. However, as can be seen from columns (4)-(6) in Table 3.7, once we look at total

[32]Another possible explanation is that, as mentioned in footnote 29 and discussed in more detail in section B.1.3 in the appendix, preferences exhibit decreasing absolute risk aversion, creating a market for insurance even when beliefs are the same.

	Seller Profits (Insurance only, true probability)			Seller Profits (total, true probability)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.925***	0.925***	1.170***	-0.188	-0.188	0.0957
	(0.121)	(0.123)	(0.124)	(0.718)	(0.725)	(0.776)
Constant	0.514***	0.818**	0.132	17.15***	16.03***	15.69***
	(0.0266)	(0.332)	(0.537)	(0.409)	(0.451)	(0.995)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with seller profits as the dependent variable. Using the true probability of the product being faulty, columns 1-3 look at the expected seller profits made from selling insurance separately, whereas columns 4-6 analyse potential treatment differences in total expected seller profits. The period and individual controls are described in table 3.3.

Table 3.7: Insurance Experiment: Seller Profits

profits for sellers, there is no significant difference between treatment and control. Including rejected offers (yielding a profit of zero), the difference is even (non-significantly) negative. There are slightly more rejections in the treatment condition (93.8% vs 90.6%, $p>0.1$), but when looking at accepted offers only, profits for sellers are not significantly larger (difference is 0.434, $p>0.1$). This shows that in the treatment condition the price that sellers charge for the product itself has been reduced such that there is no gain in profits for sellers. In the treatment condition, they charge an average price of 17.20, whereas sellers in the control condition charge 17.79.[33] For the buyers, this implies that when focusing on accepted offers only, their overall earnings are also not significantly different between conditions. Using the true probability, they earn £0.39 less ($p>0.1$). If we take the rejections into account, there is a (weakly significant) negative effect on buyer earnings in the treatment condition (differ-

[33] Table B.6 in the appendix shows that the difference between prices, without controlling for period effects or individual characteristics, is not significant, but becomes significant once including these controls.

ence is £0.88, $p<0.1$).[34] Table B.7 contains the complete comparison of buyer earnings across treatments.

Result 3.5. *Sellers in the treatment condition earn higher profits from insurance than sellers in the control condition. However, they do not earn higher total profits. There is no evidence for exploitation of biased beliefs.*

3.6.3 Buyer Behaviour

We thus want to understand whether this result can again be attributed to the sophistication of the buyers. As described in section 3.5.3, the predictions depended on whether buyers' inequity aversion is large enough so that it would force sellers to price the product below p^{max}. Only then could we expect sophisticated buyers to (sometimes) buy insurance. While in both treatments the modal offer indeed is $p = p^{max} = 20$, average prices are substantially lower and offers with $p = 20$ only account for 17.1% of all offers. Hence, if we take this as in indication that buyers are prepared to reject offers with high p, we should analyse the results in light of the predictions that were derived for the case with a large degree of inequity aversion. Therefore, buying insurance in the treatment condition is indeed compatible with sophistication on the buyers' side. Moreover, the empirical frequency of buying insurance is 0.61 which is consistent with Proposition B.1 in the appendix. Also in accordance with this proposition, prices are lower in the treatment condition than in the control condition, and vice versa for the insurance fees. As shown in Table B.6 in the appendix, the general pattern that higher insurance prices are compensated with lower product prices is a clear feature of the data. This suggests that the sophistication of buyers plays an important role here.

[34] In addition to this effect being only weakly significant, it is also the case that one buyer in the treatment condition rejected 70% of all offers, which seems to be the main driver of the effect. Excluding this buyer decreases the difference in profits to £0.48 ($p>0.1$).

Result 3.6. *In the treatment condition, buyers buy insurance with a frequency consistent with the equilibrium prediction for sophisticated buyers. They get compensated for paying more for insurance through lower base prices for the product.*

To further substantiate this result, I will look into the buyer behaviour in more detail. As discussed previously, the equilibrium for the sophisticated buyers involved randomisation over the decision whether or not to buy insurance. While certainly not impossible, it remains a fairly daunting task for buyers to implement such a strategy, especially when faced with different sellers who not always price in the same manner over the twenty rounds of the experiment. Hence, it is interesting to investigate whether there is evidence in the data on the buyer behaviour that allows us further insights in relation to the equilibrium strategies.

A key feature of the equilibrium was that higher insurance fees are compensated by lower prices in the treatment condition. Such an effect is present in the data on prices and fees as a whole (see Table B.6) and is corroborated in the buying behaviour as well: in the control condition, the difference between the price p that buyers pay when in addition they buy insurance compared to when they only buy the product, is £0.14 higher ($p>0.1$). In the treatment condition, however, buyers pay £0.48 less ($p<0.05$) for the product when they buy insurance.[35] This suggests that buyers in the treatment condition are trading off a lower p for purchasing insurance and they are thereby avoiding overpaying for the product and insurance as a whole.[36] Moreover, whenever buyers buy insurance, the total amount spent is not significantly higher in

[35] The full regression output which also accounts for estimates the difference accounting for additional control variables can be found in Table B.9 in the appendix.

[36] It can be checked that this is not a mechanical effect since there is a negative correlation between p and f and the demand for insurance decreases in f.

the treatment condition than in the control condition (the difference is £0.10, p>0.1). This behaviour does not correspond exactly to the mixed strategy described in Proposition B.1, but both the apparent $p - f$ trade-off and the share of insurance contracts less than one can be seen as a simpler heuristic that achieves outcomes similar to the equilibrium predictions.

3.7 Discussion and Related Literature

The growing literature in behavioural IO suggests that when consumers have biased beliefs about product characteristics or product usage, firms can design contracts to take advantage of these biased beliefs. For monopolistic sellers, Grubb (2009) shows how firms design three-part tariffs that exploit consumers' inability to accurately forecast their demand for calling plans. DellaVigna and Malmendier (2004) show that if consumers overestimate (underestimate) their consumption of investment (leisure) goods, firms' optimal contracts use below (above) marginal cost pricing to exploit these biases in predicting future consumption. Eliaz and Spiegler (2006) show in a more general framework how firms can use menus of contracts to screen consumers according to their ability to predict their dynamic inconsistency in preferences and show how naive types can be exploited through speculative contracts. Related to my insurance experiment, Michel (2014) studies how firms sell exploitative warranties to consumers who misperceive the associated return cost.

Sautmann (2013) provides an experimental test for some of these ideas. However, she looks at a principal-agent setting, rather than at a direct IO application. In her paper, biased beliefs come in the form of over- and underconfidence about one's own ability in a moral hazard framework. She first shows theoretically that when an unbiased principal interacts with an overconfident (underconfident) agent, he will optimally distort the incentives towards a contract with more (less) variable pay.[37] She then tests these predictions experimentally (using subjects' beliefs about their performance in a general knowledge quiz) and finds that, in line with the theoretical predictions, overconfident agents earn less than underconfident agents, but most of the principals do not adjust the incentives in the way the theory would predict.

More directly related to the behavioural IO context of my work, though not focusing on biased beliefs, is the paper by Kalaycı and Potters (2011). They look at the incentives of sellers to change the complexity of products (a product is "more complex" the more difficult it is to calculate its value) when interacting with human buyers who make mistakes in evaluating the product as compared to robot buyers who do not. They find that sellers indeed choose more complex products when buyers are human (they "obfuscate") and earn more profits from them, even though there is competition in the market (two sellers). In contrast to that, Sitzia and Zizzo (2011) find no significant effect of product complexity (they use lotteries that vary in the number of outcomes and states) on profits, that is, (monopoly) sellers do not earn more when selling more complex products to buyers.

As argued above, there is evidence in my data that the reason for why sellers do not exploit biased beliefs is because buyers are sophisticated enough to see through the sellers' incentives. Importantly, since a seller faces randomness about which of four buyers he interacts with in a given period, sophisticated buyers might protect naive buyers from exploitation in my setting.[38] For example, consider the add-on experiment. If at least one buyer is sophisticated

[37]For similar models of overconfidence and moral hazard see Santos-Pinto (2008) and De la Rosa (2011).

[38]There is a literature in behavioural IO that analyses the dependencies between naive and sophisticated consumers. Armstrong (2015) discusses the different assumptions about the market structure that determine whether sophisticated consumers benefit from naifs in the market or whether they protect them from exploitation by sellers.

and rejects offers with high fees, it might be more profitable for a seller to reduce the fee of the offer to avoid the rejection of the sophisticated buyer, even though he might forego extra profit if he were to meet a naive buyer.[39]

As mentioned when analysing theoretically the interaction between sellers and sophisticated buyers, the theoretical results are closely related to “No-Trade Theorems” (for example in Milgrom and Stokey (1982) or Tirole (1982)). There has been experimental work that directly tests whether there is indeed no trade solely based on differences in beliefs. Both Carrillo and Palfrey (2011) and Angrisani et al. (2008) find that there is more trade than predicted by theory, though the latter paper finds that over time market outcomes are closer to the theoretical benchmark.[40] Similarly, the paper relates to work documenting that experimental subjects may have problems in understanding the logic of adverse selection (Holt and Sherman, 1994; Samuelson and Bazerman, 1985; Charness and Levin, 2009). In settings resembling a simplified version of Akerlof (1970)’s market for lemons, they document more trade of products than predicted for fully rational agents. In contrast to these papers, my results provide support for the hypothesis that laboratory subjects understand the basic no-trade and adverse selection logic. While an in-depth analysis into the different results seems an interesting direction for future research, at least two possible explanations for this discrepancy come to mind. First, it might be that the explicit framing as a familiar purchase decision in my experiment helps buyers to reason correctly about the sellers’ incentives. Second, most of the aforementioned papers typically require uninformed agents to reason about future events, whereas in my experiment buyers observe the sellers’ of-

[39] This conclusion would, of course, change if sellers were allowed to offer menus of contracts as to screen customers according to their sophistication, which could be an interesting extension of the present study.

[40] Similar results of limited strategic sophistication can be found in the betting games experiments of Brocas et al. (2014), Sonsino et al. (2002), or Rogers et al. (2009).

fers. This might make it easier for buyers to correctly perceive the strategic environment.[41]

Finally, the data on beliefs in itself allows me to gain insights into how well agents can predict the biases of others. Studies on the "hindsight bias" (Fischhoff, 1975) and the "curse of knowledge" (Camerer et al., 1989), document that, when receiving new information, people act as if "they knew it all along" and fail to take on the role of lesser informed subjects by overestimating the information that others have access to. In my experiments, I find that subjects, who receive valuable information about the probability of the product being faulty, take this information into account for their own belief (they update in the correct direction) and the majority of subjects also correctly predicts the sign of the difference between their own belief and the beliefs of the less well informed counterpart. However, in both experiments, there is evidence for the type of information projection referenced above since subjects fail to realise the extent to which beliefs are different, presumably because their second-order beliefs are influenced by the feedback they got (on average, stating their own first-order belief before receiving feedback would have been more accurate).[42] While I argued above that this is unlikely to be the main driver of my results, it is nevertheless an important aspect to consider when thinking more generally about interactions between agents with different degrees of access to information.

[41]In the trading experiments of Ngangoue and Weizsäcker (2015), agents are able to appreciate the informational content of prices better when prices are realised rather than when contingent reasoning is required.

[42]Interestingly, in a recent study, Danz et al. (2015) provide evidence that experimental subjects are, however, able to predict such cases of information projection from others subjects.

3.8 Conclusion

This paper has analysed whether buyers with biased beliefs will be exploited by sellers with better calibrated beliefs. Unlike what models from behavioural IO suggest, I do not find evidence for exploitation. I show that there is reason to believe that buyers are sufficiently sophisticated in anticipating the potential dangers of exploitation. When the product that sellers offer contains a probabilistic add-on price that could potentially be used to exploit belief differences, buyers realise that high add-ons signal exploitative motives, and consequently, reject such offers more often when they hold biased beliefs. In the setting where sellers offer insurance contracts that could be used to overcharge buyers with biased beliefs for insurance, the picture that emerges looks similar. I find that while these buyers significantly overpay for insurance compared to the true expected loss, sellers reduce the base price for the product enough such that there is no effect on overall profits. I show that the data is roughly consistent with an equilibrium in which sophisticated buyers correctly reason about the sellers' exploitation motive.

The broader implications of these findings are as follows: in markets where consumers are likely to suffer from biased perceptions of product characteristics or usage patterns, this does not necessarily mean that they are vulnerable to exploitation. In a time where policy makers such as competition authorities increasingly think about designing policies that account for consumer biases, this paper shows that interventions which aim at debiasing these consumers may not necessarily lead to significant welfare gains for them. A likely reason for this finding is that consumers correctly reason about the incentives of sellers to exploit them and such awareness protects them even without interventions. Moreover, one interpretation of my experimental design is that the control condition is the result of some market intervention such as a mandatory disclosure policy which debiases buyers. In the insurance experiment, such a policy is then shown to decrease the propensity with which buyers buy (and overpay for) insurance. However, this does not change the total amount that buyers pay, because sellers respond by increasing the price for the base product. Thus, unlike papers that focus on the effects of interventions for consumers, keeping firm behaviour constant, my experiment accounts for changes in firm behaviour, i.e. the equilibrium response.[43]

An important question is to understand to what kind of real-world markets these findings apply. As discussed, there is, on the one hand, a case to be made that when buyers have wrong perceptions about their own behaviour rather than some exogenous contingency, they are more optimistic that their own belief is correct and therefore more naive. On the other hand, my findings suggest that it is unlikely that exploitation of consumers is so prevalent that affects all markets where consumers may hold biased beliefs. It would be interesting to extend these results to cases where sellers face a more competitive environment than in my setting and to cases where they may offer different contracts to different buyers, allowing sellers to screen buyers according to, for example, their degree of sophistication.

[43]The result that in these cases debiasing may not be beneficial for consumers echoes findings in Grubb and Osborne (2015) or Handel (2013) who provide counterfactual analyses of eliminating belief biases or inertia, respectively, and find that consumers would even suffer from the resulting equilibrium price increases. Kamenica et al. (2011) provide a theoretical example where —once taking into account firm responses— providing consumers with increased information about their product usage behaviour does not increase their welfare.

Chapter 4

More Effort with Less Pay — On Information Avoidance, Belief Design and Performance[1]

4.1 Introduction

Orthodox economic theory posits that agents have a non-negative willingness to pay for instrumental information, that is, information that may affect their subsequent choices. They have no reason to refuse information that comes for free. In the workplace, for example, agents would want to know their precise wage. Knowing the pay schedule allows agents to adjust their performance optimally, balancing costs and expected rewards.

Recently, this view has been challenged. When anticipations matter, agents may have an incentive to avoid information (see, e.g., Caplin and Leahy, 2001, 2004; Bénabou and Tirole, 2002; Kőszegi, 2003; Brunnermeier and Parker, 2005; Schweizer and Szech, 2014). Optimal expectations will balance the psy-

[1]The work presented in this chapter is based on a joint project with Steffen Huck and Nora Szech. I am grateful to both for the fruitful collaboration.

chological benefits of designed beliefs about the future with the material costs of making suboptimal choices. This recent literature studies how agents actively control their expectations. Optimal beliefs often turn out to be coarse. If an agent decides just for him- or herself, anticipatory utility can provide a reason to avoid information structures that are too precise.

Oster et al. (2013) demonstrate the power of information avoidance and belief design in the context of medical testing for the hereditary Huntington's Disease. They show that a large fraction of people who are at risk shy away from medical testing despite costs of testing being small and behavioral adjustments to test results being large. Their data also show that people who avoid getting tested seem to do pretty well in life. This suggests that information avoidance may be beneficial, at least for a substantial fraction of people. Oster et al. (2013) conclude that a model of belief design in the spirit of Brunnermeier and Parker (2005) captures observed behavior much more accurately than orthodox neoclassical approaches. In the latter, the only reason for information avoidance can be found in prohibitive costs of obtaining information. Furthermore, self-selection into information avoidance may play an important role for the findings as people decide themselves if they want to get tested for the disease or not.[2]

We want to explore whether preferences for information avoidance can also be found in a less extreme but more familiar economic setting: the workplace. We conduct a real-effort experiment (with a strenuous task). Across all treatments, subjects know that the piece rate is either high (1 EUR) or low (0.1 EUR), with equal probabilities. We first establish that information about this piece rate is instrumental. Our data show substantially greater effort and output for the high piece rate compared to the low piece rate when subjects

[2]In a less drastic context, when offering tests for herpes infections, Ganguly and Tasoff (2015) also find a preference for information avoidance in a substantial fraction of subjects.

are informed about their piece rate at the outset. Nevertheless, we observe a sizable fraction of subjects (robustly around one third of all subjects) who prefer not to receive precise and costless information about their piece rate when given the choice. We will refer to these subjects as information avoiders. When asked in a post-experimental questionnaire why they decided to stay uninformed, two main reasons emerge. Some information avoiders state that they did not want to be demotivated by a low piece rate. Others mention their aversion to too much pressure caused by a high piece rate. To the best of our knowledge, this presents the first laboratory evidence for the prevalence of information avoidance in an economic context in which information is instrumental.

In line with Oster et al. (2013), our data further reveal that subjects who avoid information perform highly significantly better than those who decide to learn their piece rate before starting to work. Our context allows us to explore whether self-selection into information avoidance is crucial for this finding. In a no-information treatment, we force subjects into not knowing their piece rate realizations. Performance results are again stunning and not different from those of self-selected information avoiders. Overall, performances under information avoidance are not only higher than performance results under complete information about the piece rates - they also tend to be higher than performance results of subjects who know that they are going to receive the high piece rate for sure. From a designer's perspective, it thus pays to equip subjects with coarse information structures.

As a theoretical explanation, we propose a model in the spirit of Brunnermeier and Parker (2005) incoporating heterogeneous agents.[3] This variation

[3] Answers from our ex-post questionnaire suggest a heterogeneity in fear of choking under pressure. For a homogenous population with some fear of choking, the Brunnermeier and Parker (2005) model captures all our findings straight away.

captures potential differences in the way agents react to performance schemes. Whereas for some agents higher rewards may unambiguously increase motivation and effort, for others such high rewards may have adverse effects. Specifically, when the piece rate is high, some agents may choke under pressure, a phenomenon that has received considerable attention in the psychology literature since Baumeister's seminal (1984) paper and that has more recently also been documented in a number of economic studies (see, for example, Dohmen, 2008; Ariely et al., 2009; Apesteguia and Palacios-Huerta, 2010).

In a population with heterogeneous agents, some uncertainty in incentives may prove superior to any fixed reward system. Coarse information structures may allow different types of agents to adjust their expectations in different ways, according to their personal preferences, and to bias beliefs according to their individual needs. Those who value the motivation-enhancing effect of high wages may bias their beliefs towards optimistic piece-rates, which increases their output. Those who dread choking may bias their beliefs in less drastic ways, thereby enhancing their performance. Optimal incentive design might, thus, make deliberate use of uncertainty. Indeed, tournament incentives which are widely used in the workplace might exploit this very mechanism. Likewise, the coarseness of payment contracts involving shares of the employing company may affect performance results even in workers on lower levels of a company's hierarchy.

Our paper is organized as follows. Section 4.2 describes the design, the procedures of our experiment, and the hypotheses. Section 4.3 presents our results and sets them into the context of the related literature. Section 4.4 proposes a variation of the Brunnermeier and Parker (2005) model allowing for heterogeneous agents, and Section 4.5 concludes.

4.2 Experimental Design

We conduct three main treatments, Full Info, No Info, and Info Choice. These treatments are identical except that information about piece rates varies—exogenously or endogenously.

In all main treatments, subjects know that they receive an either a high (1 EUR) or a low (0.1 EUR) piece rate, with equal probabilities, for working on a tedious task.[4] They have 60 minutes to enter lines of strings, containing numbers, upper case and lower case letters, *backwards* into the computer interface. For each correctly entered string, they are going to receive their piece rate. Each string consists of 60 characters. For example, one of the strings used in the experiment looks as follows.

NXgCX7JHxYZj2cfBSd8JtkYp3LPcyDX8y8NNQhrzJfg22S2ACjC85EQ43B7L

Each task consists of one of these randomly generated strings and all tasks are identical for all subjects.

We vary across treatments how much information subjects have about their piece rate while working on the task. In treatment Full Info, every subject gets to know his or her piece rate immediately, before working. In treatment Info Choice, subjects can choose whether to know their piece rate immediately, or whether they only want to find out about it after they have finished working on the task. Thus subjects can avoid information if they want to. In treatment No Info, subjects are forced into information avoidance, but know the probability distribution over wages. At the end of the experiment, we ask subjects to provide us with some basic demographic information about themselves. In treatment Info Choice we also ask them to state their reasons for

[4]At the time of the experiment, 1 EUR $\approx$ 1.37 USD.

choosing (not) to obtain information about their piece rate before working on the task.

We hypothesized that performance would depend on the piece rate, such that subjects would perform better under the high piece rate than under the low one. Treatment Full Info allows us to find out whether information about the piece rate is indeed instrumental. Furthermore, in line with Oster et al. (2013) and Ganguly and Tasoff (2015), we expected that nevertheless, a substantial fraction of subjects may decide to avoid information about the piece-rate in Info Choice. Further, as suggested in Oster et al. (2013), performance may not suffer from information avoidance. If so, treatment No Info allows us to uncover potential causal effects of self-selection into information avoidance.

We also implement another treatment, Medium Wage, where every subject earns a piece rate of 0.55 EUR. This treatment allows us to disentangle whether subjects stick to their Bayesian prior when avoiding information, or if there is a tendency that a substantial fraction of subjects distort their beliefs behaviorally as suggested in Brunnermeier and Parker (2005). This treatment informs our modeling approach in Section 4.

Implementing the different information structures across treatments. The detailed procedures in our main treatments, Full Info, No Info, and Info Choice, are as follows. When entering the lab, each subject is randomly allocated a red or a black chip by one of the experimenters. Half of the chips are black, the other half red. Each subject is then told to take a seat at a computer terminal where the screen shows a square with the color corresponding to the color of his or her chip.[5] Subjects know that depending

[5]The procedure makes sure that subjects can see that there are chips of two colors and that the procedure is entirely random.

on the color (red or black), they can either earn 0.1 EUR or 1 EUR for each correctly entered string, with equal probability.

In order to determine which color corresponds to a high piece rate and which to a low one, we use the following procedure. We prepared two pieces of cardboard which look identical from the outside when folded, but inside either show a red or a black square. After showing the cardboard pieces (from outside and inside) to all participants, they are folded, secured with paper clips, placed into a small bag and shuffled. Another experimenter then draws one of the two folded cardboard pieces. The color of the drawn piece determines which color is associated with the high wage for this session.

In treatment Full Info, the cardboard is unfolded and the color is revealed to all subjects immediately. Thus subjects know whether they are going to receive the high piece rate, or the low one.

In treatment No Info, the folded cardboard is placed onto a white board at the front of the room where it remains for the whole duration of the experiment and is revealed to all participants once the allowed time for the task (60 minutes) is up. Hence subjects do not know whether they earn the high or the low piece rate when working on the task.

In treatment Info Choice, subjects are asked on their computer screen whether they would like to receive the information about the color now, or wait until the end of the experiment. After clicking the button corresponding to their choice, another screen appears which states the subject's decision. After all subjects have made their choice, the experimenter walks through the lab and privately reveals the color inside the cardboard to those subjects who decided to see it. As in the No Info treatment, the folded cardboard is then placed onto a white board and revealed afterwards. Thus, in this treatment, subjects choose whether they want to know their piece rate beforehand or not.

Then the real effort task starts.

After each string that subjects enter, they learn whether they entered it correctly. They can then click on a button to continue. Subjects are informed about the time that remains. Subjects are not allowed to use any electronic devices, but are each given a copy of a well-known German weekly, called DER SPIEGEL. This magazine has a weekly circulation of more than one million. It contains all sorts of articles, from investigative journalism over reports on politics to articles about scientific discoveries and information on cultural events and sports. Subjects are explicitly told that they can make use of the magazine "...whenever, during the experiment, [they] would like to take a break or pass time". Thus, no subject has to feel obliged to work on the task if he or she prefers to spend their time otherwise.

In the Medium Wage treatment, there is no need for any randomization in the beginning and subjects immediately start working on the task after reading the instructions. In order to keep the context as comparable as possible, we also inform subjects that other participants could earn either 0.1 EUR or 1 EUR for the same task in previous experiments, with equal probability.

In total, our sample consists of 238 subjects. All treatments were run at the WZB-TU Laboratory in Berlin between November 2013 and April 2014. There were no restrictions imposed on the invited participants regarding gender, subject of study, or previous experience with experiments. We used z-tree (Fischbacher, 2007) as the experimental software and ORSEE (Greiner, 2004) to recruit subjects. Participants received a show-up fee of 5 EUR and average earnings over all treatments amounted to 14.29 EUR. Each session lasted 80 to 90 minutes.[6]

[6]We also ran two sessions without the weekly magazine. We report the findings from these sessions as well. Overall, results are quite similar to those in our main treatments, though effects are slightly less pronounced.

4.3 Results

In treatment Full Info, subjects know their piece rate realization before starting to work on the task. Our data show that subjects perform significantly better under the high versus the low piece rate. Subjects working for the low piece rate of 0.1 EUR solve 20.67 tasks on average, whereas subjects working for the high piece rate of 1 EUR solve 26.21 tasks correctly. This difference is significant (p=0.043)[7] and confirms that information about the piece rate is instrumental as hypothesized.[8]

We expected that nevertheless, in treatment Info Choice, a substantial fraction of subjects would avoid information about the piece rate and postpone learning the piece rate to the end of the experiment. This is exactly what we find. 30 out of 95 subjects (31.6%) decide to avoid information. They prefer to learn their piece rate only after they have worked. These subjects thus decided to forgo information that turned out to be of instrumental value as shown in treatment Full Info.[9]

In a post-experimental questionnaire, we ask information avoiders about their motives for not getting informed about their piece rates. There are basically two types of answers that subjects provide. Several subjects tend to argue that they wanted to avoid being demotivated in case of having a low piece rate. Other subjects stressed that they were afraid of the pressure in case of learning that they receive a high piece rate.[10]

[7]Unless indicated otherwise, all p-values are calculated using a Wilcoxon Rank Sum test.

[8]The other possible measure of effort would be the number of attempted tasks. Arguably, our task is prone to errors and for some subjects this measure might more correctly reflect the actual effort put in. Others might choose a more risky strategy and tolerate more errors. The two measures are highly correlated ($\rho = 0.8763$, p-value=0.0000) and the results are very similar.

[9]As a robustness check, we analyze the two sessions that where identical to Info Choice except that subjects did not have access to the magazine. There, 15 out of 44 subjects (34.1%) decided not to acquire information, an effect of almost identical magnitude.

[10]We analyse the subjects' statements in more detail below.

Looking at performance, information avoiders do extremely well. They solve 30 tasks correctly on average, while information receivers solved 21.31 tasks correctly (p=0.0002). Strikingly, information avoiders even tended to outperform the subgroup of subjects in Info Choice who received the information that their piece rate was high (30 versus 25.53, p=0.0573).

Potentially, as the anwers to the ex post questionnaire suggest, information avoiders work under beliefs that motivated their performance in better ways than those subjects who opted for information. Yet as we argued above, treatment Info Choice does not allow us to rule out potential effects of self-selection on unobservables that correlate with performance. For example, high-ability subjects might have self-selected into information avoidance, which would explain the observed performance differential. In order to identify the role of self-selection, we ran treatment No Info in which subjects have no option to learn their piece rates before working. As in the other main treatments, subjects knew that piece rates were either 1 EUR or 0.1 EUR, with equal probability. Yet subjects knew that their piece rate was only disclosed to them at the end of the experiment. Subjects' information structure in No Info is, hence, similar to the one that information avoiders chose in Info Choice, with the only difference that self-selection was not possible in the No Info treatment. Our data reveal that performance results between these two groups of subjects are almost identical and statistically indistinguishable (30 versus 28.02, p=0.3710).[11] The data show that not knowing the piece rate realization enhances performance, even if subjects do not freely opt for this coarse information structure. As self-selection does not lead to statistically different performance results, we can pool the data for the different information structures under which subjects worked. From a designer's perspective,

[11]There is no significant difference either for any of the other two groups that receive information about their piece rates, at any conventional level of statistical significance.

piece rate	Full Info		Info Choice		No Info		Medium Wage	
	mean (s.d.)	median	mean (s.d.)	median	mean (s.d.)	median	mean (s.d)	median
0.1	20.67 (10.49)	22	17.69 (11.37)	17				
	N=24		*N=35*					
0.55							24.66 (9.58)	24
							N=47	
1	26.21 (8.75)	26	25.53 (9.86)	23				
	N=24		*N=30*					
unknown			30 (9.35)	29	28.02 (8.41)	28		
			N=30		*N=48*			
aggregate	23.44 (9.17)	25	24.05 (11.44)	23	28.02 (8.41)	28	24.66 (9.58)	24
	N=48		*N=95*		*N=48*		*N=47*	

Table 4.1: Mean and median performance across treatments

opting for the No Info policy leads to significantly better performance results than Full Info (23.44 versus 28.02, p=0.0274).

Information avoidance enhances performance, even if it is enforced. When asked, subjects argue with controlling their beliefs while working. Both, avoiding demotivation from a potentially low piece rate as well as avoiding too much pressure from a potentially high piece rate are stated as motives for information avoidance. This suggests that there may be a heterogeneity in our subjects, and that, at least for some of them, their performance varies non-monotonically with the wage they expect to receive. If all subjects were to use their Bayesian prior when they do not know their exact wage, their expected wage is 0.55 EUR. If this wage is sufficiently high to avoid demotivation and sufficiently low to prevent choking under pressure, the performance at a fixed wage of 0.55 EUR should not differ much from the performance

piece rate	Pooled		p-value of pairwise test		
	mean (s.d.)	median	0.1	0.55	1
0.1	18.90 (11.03)	19			
0.55	24.66 (9.58)	24	0.0091		
1	25.83 (8.56)	25	0.0003	0.2073	
unknown	28.78 (8.78)	28	0.0000	0.0015	0.0645

Table 4.2: Pairwise comparisons of performance results across treatments

under information avoidance. However, if, for at least some subjects, their performance-maximizing wage is different from 0.55 EUR, information avoidance may allow them to optimally design their expectations as described in Brunnermeier and Parker (2005), away from the Bayesian prior. With room for this kind of behavioral belief distortion, information avoidance may even outperform any kind of fixed reward system in a heterogenous population. In order to understand whether there is evidence for behavioral belief design, we ran treatment Medium Wage. In this treatment, subjects know that they work for a piece-rate of 0.55 EUR. They also know that other subjects worked for either 1 EUR or 0.10 EUR with equal probability, in order to keep the context comparable. We find that subjects solve on average 24.66 tasks correctly in this treatment.

Table 4.1 presents an overview of all our results per piece rate and treatment. Table 4.2 displays the performance results for the different information structures subjects had while working on the task, pooling the data of self-chosen and enforced information structures. We find that the result from above regarding the effort level for the known wages 0.1 EUR and 1 EUR

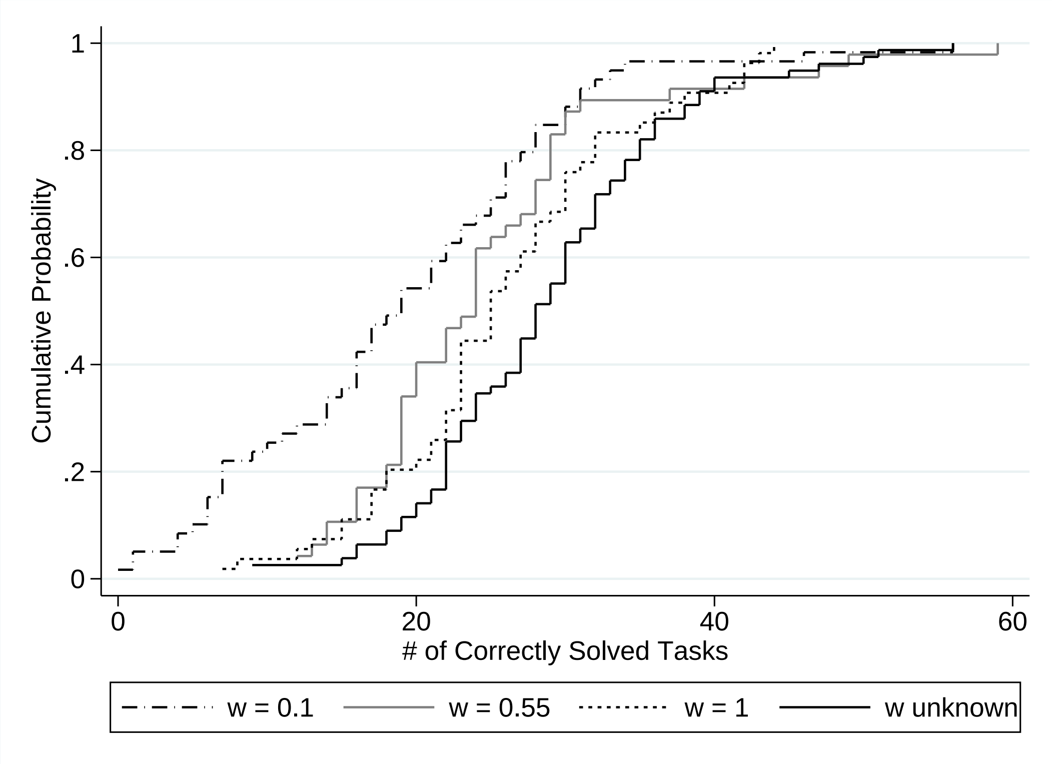


Figure 4.1: CDFs of effort choices, by wage.

carries over: Subjects perform significantly better at the higher wage (18.90 vs. 25.83 correct tasks on average, p=0.0003).[12] On the other hand, for subjects who did not know whether their piece rate is 0.1 EUR or 1 EUR, we obtain an average performance of 28.78 correctly solved tasks. In other words, a participant who does not know his wage solves about 3 more tasks than a subject who knows that he or she receives the high wage of 1 EUR per solved task (p=0.0645).[13] Figure 4.1 plots the empirical distribution functions for the four cases. Visually, the distribution for the case in which the piece rate is unknown almost first-order stochastically dominates the distributions for all

[12]It appears that most subjects did not take longer breaks during the one hour working time. Even under the low piece rate of 0.1 EUR, less than 30 percent of subjects spend more than ten minutes with any task. Looking at the other piece rates (including unknown), no subject paused for a longer interval.

[13]Using attempted tasks as the effort measure we find a similar effect: when the wage is unknown subjects attempt on average 40.51 tasks as compared to 35.46 at the certain wage of 1 EUR (p=0.0052).

other treatments, in which piece rates were deterministic. Performance results under information avoidance are also statistically different from performance results under the secure 0.55 EUR piece rate (p=0.0015). Therefore, while it may be that a fraction of subjects sticked to the undistorted Bayesian expectation of 0.55 EUR under information avoidance, a substantial fraction of subjects are likely to have distorted their beliefs in behavioral ways.

All in all, from a designer's perspective, giving people the choice to avoid information does not harm aggregate performance compared to a full information policy. Subjects who self-select into information avoidance perform extremely well. This result is in line with findings from Oster et al. (2013). Moreover, enforcing information avoidance significantly increases performance results. This effect must be partly due to behavioral belief distortion as described in Brunnermeier and Parker (2005). If subjects mostly tended to stick to the Bayesian prior under information avoidance, performance results would have to be roughly equal to the Medium Wage treatment, which is not the case.

Our data further suggest that a policy of information avoidance tends to outperform paying the high piece rate for sure, even though the latter is much more costly from a designer's perspective. An explanation for this, as put forward by several subjects themselves, could be a fear of choking under pressure in a substantial fraction of subjects. This potential heterogeneity informs our modeling approach in Section 4. A model in the spirit of Brunnermeier and Parker (2005) allowing for heterogeneous preferences can capture our observations. Of course, it may also be that subjects homogeneously distort their beliefs to some moderately optimistic belief, such as a piece rate of 0.80 EUR. If all subjects were to some extent afraid of choking under pressure, performance results could turn out higher than performance results for the high piece rate of 1.00 EUR. As several subjects tended to argue either with choking under pressure or with avoiding potential demotivation from a low piece rate, we will elaborate the case of heterogenous agents in a population in Section 4.

Discussion. We would first like to point out that our results on performance and information avoidance are inconsistent with classical models of expected utility in which anticipatory feelings do not matter. Consider a utility function that is separable in monetary payoffs and effort costs, and make the natural assumption that the cost of effort is increasing. Then, irrespective of the agents' risk attitudes, the optimal level of effort is predicted to be higher in the case where the agent works for a fixed piece rate compared to the case where he or she faces a lottery over this piece rate and any lower piece rate.

At the end of the experiment, after subjects finished working on the task, we asked them to state the reasons for obtaining (avoiding) information about their wage. Among those who decided to obtain information, the vast majority of subjects (49 out of 65) state that they wanted to know their piece rate in order to adjust their effort accordingly. These subjects thus acknowledge the instrumental role of information for them.[14] A minority of people (13 out of 65) explicitly state preferences for information, e.g. "curiosity", as their reason for obtaining information. Looking at subjects who avoided information, a more diverse picture emerges. On the one hand subjects state that by not knowing their wage they wanted to ensure that they would be sufficiently motivated in light of the risk of receiving a low wage (7 out of 30), and on the other hand subjects reveal that they are afraid of being under too much pressure when knowing for certain that their wage is high (8 out of 30).[15] Interestingly,

[14]A typical statement was, for example, "I wanted to know whether putting in effort would be worthwhile, if the wage would have been low, I would not have bothered".

[15]For the first group, a typical statement reads "with the hope of the higher payment, I wanted to keep up my motivation", whereas those potentially choking under the pressure of

another group of subjects (9 out of 30) explicitly state that both potential demotivation as well as pressure of a high wage influenced their decision, or that they felt they could work best without being influenced too much by a certain wage. All these answers point towards an important role for belief design in our setting. Subjects consciously reflect on they way their performance will be affected by the wage distribution and deliberately choose to avoid information in order to optimize their performance.

Related findings in the literature. Our finding that roughly a third of participants choose to avoid instrumental information does not have a precedent in the literature in so far as settings as the workplace are concerned. Oster et al. (2013) document evidence for information avoidance in the domain of health outcomes. The authors find that only 7 percent of individuals that can be classified as being at-risk of contracting the hereditary Huntington Disease decide to undergo testing that provides them with certainty about their health status. This effect of information avoidance is significantly larger than ours, but unlike in our setting obtaining information is not costless. Furthermore, effects on anticipatory feelings are arguably much more drastic than in our setting, such that information avoidance may become even more attractive.

In a laboratory study Ganguly and Tasoff (2015) give subjects the option to avoid being tested for herpes at a cost of 10 USD. Depending on the type of virus they are tested for, 5.2% to 15.6% of individuals are willing to give up 10 USD in order not to be tested. While in this setting information about the health status is instrumental, Ganguly and Tasoff (2015) also look at demand for non-instrumental information. Giving subjects the option to learn (or avoid) information about potential monetary payments at some cost, they

a high wage gave answers like "I chose not to learn the color in order to take the pressure off myself. I would have been even more error-prone".

find that some subjects are willing to pay money to avoid learning about the outcome of a lottery whereas others are willing to pay for early resolution of uncertainty (each of the two groups roughly comprises between 30% and 40% of the sample).

Eil and Rao (2011) find evidence for information avoidance when they elicit subjects' willingness-to-pay of learning their relative rank in terms of IQ and attractiveness. The authors show that subjects who believe themselves to be below average in a category are willing to pay for not learning their true rank in the ability distribution. However, in their setting when subjects were given the option to learn their rank, this information did not have an impact on their earnings in the experiment, and is thus non-instrumental.

Likewise, Eliaz and Schotter (2010) and Falk and Zimmermann (2014) analyze the demand for non-instrumental information. Both studies find evidence that subjects have a preference for obtaining non-instrumental information. In Eliaz and Schotter (2010), a majority of subjects choose to pay a fee to obtain information that will only alter their confidence about their decision, but not their decision itself. In Falk and Zimmermann (2014), subjects can choose the timing when to find out about whether they are to receive small electrical shocks and a large majority chooses to find out immediately, and can thus be classified as "curious", i.e. having a strict preference for information. Compared to our setting, these results indicate that we might be underestimating the magnitude of the effect of anticipatory utility and the desire for belief design. For subjects in our sample with a high degree of curiosity (we can identify some of them via their answers in the post-experimental questionnaire, see above) or desire for confidence, these motives might outweigh the benefits from being able to optimally design their beliefs when the piece rate is unknown.

In relation to our findings in the domain of effort choices, Ariely et al. (2009) find in experiments conducted in the U.S. and in India for a variety of different tasks that high incentives might backfire and reduce performance. For the majority of the tasks administered in their study, performance varies non-monotonically with the compensation offered; moderate incentives typically deliver the highest performance level. Dohmen (2008) and Apesteguia and Palacios-Huerta (2010) document a similar effect for professional athletes. Building on Baumeister (1984), all these studies identify the phenomenon that individuals may choke under too much pressure induced by high rewards. Therefore, the fear of choking under pressure that a fraction of subjects expressed, in our study, may be rather plausible.

Shen et al. (2015) document that in certain real-effort situations a small reward that is uncertain and either higher or lower (e.g. 1 USD or 2 USD with equal probability, or a smaller versus a larger amount of candy) may generate better performance results than the fixed higher reward (e.g. 2 USD). In all the settings studied by Shen et al. (2015), overall rewards are small. The authors suggest that the uncertainty about these rewards may increase subjects' excitement, and subsequently, their motivation with which they engage in the task. While this may be true when stakes are low, we consider it unlikely that excitement alone drives our results, in which overall stakes are rather high (about 30 EUR if the piece rate is 1 EUR). Various papers show that most agents are risk-averse in economic settings, and try to avoid overly risky lotteries (e.g. Holt and Laury, 2002). As discussed above, in our ex-post questionnaire, many subjects argue that they decided to avoid information in order to prevent demotivation from a low piece rate, while others stated that they wanted to avoid feelings of pressure. Excitement from the lottery does not seem to play a major role in the explanations subjects gave. In the next section, we adapt the model of Brunnermeier and Parker (2005) on belief design in order to account for potentially heterogeneous goals in behavioral belief design across agents.

4.4 Theory

All our results can be captured by the model of behaviorally distorted beliefs of Brunnermeier and Parker (2005) (BP henceforth). In the following, we are going to elaborate on the case of a heterogenous population of agents in which some agents are afraid of choking under pressure if the piece rate, or wage w henceforth, is high. We are motivated to do so as in our ex-post questionnaire, there seemed to appear a heterogeneity in preferences across subjects. For a homogenous population in which all agents are a bit afraid of choking under pressure, all our findings can be explained by the BP model directly.

In their model, agents optimally choose their beliefs as to balance benefits from anticipatory emotions and costs in decision making due to biased beliefs. In our experimental setting, subjects either know the wage they are working for (in treatments Full Info, Medium Wage, and if they opted for information to be revealed in Info Choice) or they face uncertainty about whether it is the high or the low wage (No Info and Info Choice, if they decided to stay uninformed). Whereas the former case leaves no room for manipulation of beliefs, in the latter subjects might hold subjective beliefs that do not treat the high and the low wage as being equally likely (which would be the Bayesian prior).

In our setting, an agent derives utility from the payment she receives for solving tasks but has to bear the cost of effort. We model effort directly as the number of correctly solved tasks, e, and assume risk neutrality throughout.

The (expected) payment is we, where w is the (expected) wage. As described in the previous section, we aim to develop a model that can capture the possibility of choking under pressure, i.e. the phenomenon that an agent's performance might be adversely affected if the (expected) wage for the task is high. We therefore allow the cost of effort not only to depend on e, but also (potentially negatively) on w for these agents. In the specific case we look at below, such a cost function delivers an optimal effort level that is hump-shaped in the wage, i.e. effort is maximized at an intermediate wage. In cases where there is uncertainty about the wage, we interpret w as the expected wage, potentially distorted by optimal belief design by the agent. Assuming additive separability of monetary payments and effort costs the agents' consumption utility is then given by $u(e, w) = we - c(e, w)$.

At time 0, subjects in treatment Info Choice decide whether to learn their wage or not. Subjects in the other treatments either know their wage by default or are forced into information avoidance. At time 1, subjects decide how much effort to exert when working on the task and they experience *anticipatory utility* based on their expected consumption utility which materializes at time 2. Following BP, we assume that agents who do not know their wage can optimally design their beliefs $\pi \in [0, 1]$, where π denotes the belief that the wage is high, w_H, rather than low, w_L (where $w_H > w_L > 0$).[16] The chosen belief affects anticipatory utility but has no direct effect on consumption utility. Indirectly, however, it does affect consumption utility through the agent's effort choice which will be based on the subjective wage $w(\pi) = \pi w_H + (1 - \pi) w_L$.

Following BP, agents' overall well-being is given by a weighted sum of

[16] Note that allowing for the possibility that beliefs can be freely (and therefore in a non-Bayesian manner) chosen is what makes this model more suitable for our purposes than, for example, Bénabou and Tirole (2002). In addition, the latter model focuses on how agents forget or suppress bad news through selective memory, rather than optimally distorting beliefs in the presence of uncertainty.

anticipatory utility and consumption utility where the relative weight of anticipatory utility is denoted by $\delta \geq 0$. Given a belief π and the corresponding expected wage $w(\pi)$, our agent chooses effort e to maximize $w(\pi)e - c(e, w(\pi))$, i.e. the anticipated consumption utility. Expecting the effort choice $e^*(w(\pi))$, the agent maximizes overall well-being by choosing the belief π that balances anticipatory feelings and final consumption utility. In our setting, the belief choice directly corresponds to a choice of the subjective wage. Hence, we can therefore suppress the dependence on π and allow the agent to choose $w \in [w_L, w_H]$ directly. Defining $\bar{w} = \frac{1}{2}(w_H + w_L)$, our agent then maximizes

$$U(w|\bar{w}) = \delta \left[we^*(w) - c(e^*(w), w)\right] + \bar{w}e^*(w) - c(e^*(w), \bar{w}). \tag{4.1}$$

The optimal choice of beliefs has to consider the following trade off: an agent may distort beliefs away from $\bar{w}$ in order to increase her anticipatory utility. By being more optimistic about her odds to be paid the high wage, she manipulates herself into exerting more effort since effort is determined by the subjective expected wage. However, choosing a belief different from $\pi = 0.5$ may come at a cost because the agent will exert more effort than what is optimal given $\bar{w}$. In general, choking and non-choking agents differ in their choice of how much to distort their beliefs. Under our assumptions on the cost function, by being more optimistic, non-choking agents unambiguously increase their anticipatory utility and the more they care about it the more optimistic they will be. Agents who choke at high wages, however, will prefer to distort their beliefs less in order to work at a (subjective) expected wage that is lower than w_H.

It is worth noting that the wage that enters the cost function differs between the anticipatory utility term and the consumption utility term. When the agent experiences "true" consumption utility, her cost of effort is given by the undistorted expected wage $\bar{w}$. Nevertheless, when anticipating these costs we assume that they are based on the subjective wage w.

In order to demonstrate how this variation of BP can account for our experimental findings, we formalize the above intuition using the family of cost of effort functions given by

$$c(e, w) = \frac{1}{2}\alpha e^2 + \gamma e f(w)$$

with $\alpha > 0$, $\gamma \geq 0$, and $f(w) > 0, f'(w) > 0, f''(w) > 0$.

It is straightforward to see that for a given w, optimal effort is given by

$$e^*(w) = \frac{1}{\alpha}\left(w - \gamma f(w)\right).$$

Optimal effort e^* is strictly concave in the (expected) wage and if γ is sufficiently large, e^* is decreasing for sufficiently large w. We add two more assumptions which essentially impose that γ is not too high. We assume that $w_L > \gamma f(w_L)$ and $w_H > \gamma f(w_H)$ so that optimal efforts e^* are positive over $[w_L, w_H]$. Moreover, we assume $\gamma f'(w_L) < 1$ so that optimal efforts are increasing in wage for small wage levels, though not necessarily for high wages.

As desired, our model delivers an optimal effort level that can be hump-shaped in the wage. Specifically, we define $\hat{w}$ as the unique wage that maximizes effort. If an interior solution $\hat{w} < w_H$ exists, it solves $\gamma f'(\hat{w}) = 1$. This means that for any agent with $w_H > \hat{w}$ effort is maximal at a wage level lower than w_H. We think of these agents as choking under the pressure of a wage that is too high. To simplify the exposition below, we also assume that $\hat{w} > \bar{w}$. If there is no $\hat{w}$ solving $\gamma f'(\hat{w}) = 1$, we know that e^* is increasing and $\hat{w} = w_H$. Define $\hat{\gamma}$ as $\hat{\gamma} = 1/f'(w_H)$. Then, the agents who do not choke under pressure, that is, those with $\hat{w} = w_H$, can be characterized as the agents with $\gamma \leq \hat{\gamma}$. These agents always exert more effort as the wage increases.

Before presenting the solution for the full optimization problem, let us consider an agent who only cares about anticipatory utility. This agent will choose w to maximize $we^*(w) - c(e^*(w), w)$. By the envelope theorem, the first order condition for this problem is given by

$$e^*(w)\frac{de^*(w)}{dw} = 0. \tag{4.2}$$

Since e^* is positive, the unique maximizer of anticipatory utility is the wage $\hat{w}$ which maximizes effort. This holds both for the interior solution of the choking agents and for the corner solution $w = \hat{w}$ of the non-choking agents.[17]

Next, consider the other extreme, an agent who does not care about anticipatory utility at all, $\delta = 0$. He or she chooses the wage w to maximize $\bar{w}e^*(w) - c(e^*(w), \bar{w})$, which yields the following first order condition:

$$\frac{de^*(w)}{dw}\left(e^*(\bar{w}) - e^*(w)\right) = 0.$$

It is straightforward to check via the second derivative that $w = \bar{w}$ maximizes the objective function. However, for choking agents there might be a second maximizer due to the hump-shaped nature of effort in wage, namely the effort level that induces the same effort level as $\bar{w}$, if it exists. For our purposes it is immaterial which one of the two the agent chooses. A "standard" agent who is not affected by anticipatory utility chooses the same effort level as an agent who faces a sure wage of $\bar{w}$, like in our Medium Wage treatment. Thus, for

[17]It is straightforward to check that this equation has more than one solution, but that only $w = \hat{w}$ is indeed a maximum, provided that $\hat{w} < w_H$. Hence, for choking agents, the wage that maximizes anticipatory utility is interior and equal to the effort-maximizing wage, whereas for non-choking agents the optimal wage is w_H since for $\gamma < \hat{\gamma}$ sufficiently low the LHS in (4.2) is positive everywhere on $[w_L, w_H]$.

$\delta = 0$ our model nests the neoclassical model since under risk neutrality effort choices should only depend on the average wage.

Putting these two effects together for intermediate values of δ, we obtain the FOC of the full objective function in (4.1) as:

$$\frac{de^*(w)}{dw}\left(e^*(\bar{w}) - (1-\delta)e^*(w)\right) = 0.$$

Analyzing this expression allows us to derive the solution w^* to the maximization problem.

Proposition 4.1. *Define* $\delta^* = 1 - \frac{e^*(\bar{w})}{e^*(\hat{w})} \in (0, 1]$. *(a) For all* $\delta < \delta^*$, *the optimal wages* w^* *are implicitly defined through the equation*

$$e^*(\bar{w}) = (1-\delta)e^*(w^*)$$

All solutions induce the same effort and satisfy $w^* > \bar{w}$. *(b) For* $\delta \geq \delta^*$, $w^* = \hat{w}$.

Proof. Using the specific functional form, the first order condition for maximizing (4.1) can be written via equation (4.4) as

$$\frac{de^*(w)}{dw}\left(e^*(\bar{w}) - (1-\delta)e^*(w)\right) = 0.$$

Solutions of this condition are either $\hat{w}$, since $\frac{de^*(w)}{dw}\,|_{w=\hat{w}} = 0$, provided it exists, or the w that solve the term in brackets. In order to solve the term in brackets, $e^*(w)$ must be larger than $e^*(\bar{w})$ and δ must not be too large. The boundary value δ^* is derived from $e^*(\bar{w}) = (1-\delta^*)e^*(\hat{w})$. A solution w^* which sets $e^*(\bar{w}) - (1-\delta)e^*(w)$ to zero exists whenever $\delta \leq \delta^*$. Since e^* is continuous and increasing over $[0, \hat{w}]$, there is always a solution $w^* \in [\bar{w}, \hat{w}]$. There may be further solutions $w^* \in [\hat{w}, w_H]$ which induce the same wage.

Consider the second derivative, given by

$$-(1-\delta)\left[\frac{de^*(w)}{dw}\right]^2 + \frac{d^2e^*(w)}{dw^2}\left(e^*(\bar{w}) - (1-\delta)e^*(w)\right).$$

If the solution to the problem is given by a solution of $e^*(\bar{w}) = (1-\delta)e^*(w^*)$, this is a maximum because such a solution can only exist for $\delta < \delta^* < 1$, so that the second derivative is negative.

Given the definition of $\hat{w}$, $e^*(w)$ is maximized at $\hat{w}$ if $\hat{w} < w_H$. $\hat{w}$ sets the first term in the second derivative to zero. Furthermore, for any $\delta > \delta^*$, $\frac{d^2e^*(w)}{dw^2}|_{w=\hat{w}} < 0$ and the term in brackets then is positive. We thus have a maximum here as well. For $\delta < \delta^*$, $\hat{w}$ will be a minimum because in this case $e^*(\bar{w}) - (1-\delta)e^*(\hat{w})$ will be negative. For $\delta = \delta^*$, the second derivative is zero, but for $\varepsilon > 0$, the first derivative at $\hat{w} - \varepsilon$ is positive and at $\hat{w} + \varepsilon$ it is negative. In the case where $\hat{w} = w_H$, and there is no w to satisfy $e^*(\bar{w}) = (1-\delta)e^*(w^*)$, the first derivative in (4.4) will be strictly positive and thus $w^* = w_H$. □

The proposition shows that in the case where $\gamma < \hat{\gamma}$ (i.e. where the agent does not choke and therefore $\hat{w} = w_H$) a sufficiently large δ implies that the agent will choose $\pi = 1$, that is, he or she chooses effort and receives anticipatory utility under the (distorted) belief that the piece rate will be w_H with probability 1. These non-choking agents then exert the same effort in the case where they do not know their wage and in the case where they have found out that their wage is w_H per task solved. Non-choking agents with a lower value of δ, will adopt interior $\pi \in [0.5, 1)$ because they care relatively more about the upward distortion of consumption utility induced by over-exerting effort.

Agents that choke under pressure, as represented by a positive value of γ, do not distort beliefs in such an extreme way. Since their optimal level of effort is strictly below $e^*(w_H)$, they will, provided they care enough about anticipatory utility (δ is large enough), distort beliefs only up to the point where they exert the maximum level of effort, $e^*(\hat{w})$. Hence, the model delivers the result that these choking agents exert a strictly higher effort in the case where they do not know their piece rate, compared to where they know for sure that they will be paid according to w_H.

Corollary 4.1. *Consider a group of agents consisting of two types of agents with different parameters $\gamma_2 > \hat{\gamma} > \gamma_1$. If $\delta \geq \delta^*$ then average effort (= average number of correctly solved tasks) of the group will be higher when the agents do not know whether their wage is w_L or w_H than in the case when they all know that that their wage will be w_H.*

Proof. Under the assumptions on γ_1 and γ_2 (which ensures that the effect of choking is sufficiently large as to guarantee that agents of type 2 ("choking agents") put in maximal effort at a wage lower than w_H) and δ (that anticipatory utility is large enough), $w_1^* = w_H$ and $w_2^* = \hat{w}$ holds for the case of unknown wages. Furthermore, agents of type 1 exert effort of $e_1^*(w_H)$ whereas agents of type 2 exert effort of $e_2^*(\hat{w})$. Under a known wage of w_H, the respective effort levels are given by $e_1^*(w_H)$ and $e_2^*(w_H) < e_2^*(\hat{w})$. Choking agents thus exert higher effort when the wage is unknown whereas standard agents exert the same effort level as for a known high wage, proving the statement. □

Our variant of BP can also explain why in treatment Info Choice some agents choose not to obtain information about their wage. An agent who decides whether to learn her wage faces a tradeoff between optimally choosing her effort after having eliminated uncertainty about the wage, but also forgoes the opportunity to benefit from being able optimally to choose her belief and benefit from distorting anticipatory utility upwards. Formally, an agent decides not to learn the wage if

$$\delta\left[w^*e^*(w^*) - c(e^*(w^*), w^*)\right] + \bar{w}e^*(w^*) - c(e^*(w^*), \bar{w}) \geq$$
$$\frac{1}{2}(1+\delta)\left[w_H e^*(w_H) - c(e^*(w_H), w_H) + w_L e^*(w_L) - c(e^*(w_L), w_L)\right] \quad (4.3)$$

Proposition 4.2. *There exists a $\hat{\delta} \geq 0$ such that all agents with $\delta > \hat{\delta}$ prefer not to know whether their wage is w_H or w_L.*

Proof. Dividing both sides by $(1+\delta)$ ensures that the RHS of the condition in (4.3) stays constant once we increase δ. The LHS is then simply a weighted average between anticipatory utility and consumption utility. Using a standard envelope theorem argument, we then see that increasing δ strictly increases the LHS because $w^*e^*(w^*) - c(e^*(w^*), w^*) \geq \bar{w}e^*(w^*) - c(e^*(w^*), \bar{w})$. Hence, for large enough δ, inequality (4.3) is satisfied. □

To conclude this section, consider again the experimental results described in the previous section. An orthodox model of effort choice without anticipatory utility or choking ($\delta = \gamma = 0$), would predict that for all treatments where the wage is known, average effort is increasing in the wage, $e^*(w_L) < e^*(\bar{w}) < e^*(w_H)$ and that under risk neutrality agents who do not know their wage choose $e^*(\bar{w})$. Also, we should not see anybody rejecting the information about the wage. Our results do not conform to this. Introducing anticipatory utility can remedy this: agents who do not know their wage, optimally choose more optimistic beliefs, inducing themselves to exert more effort. If they value anticipatory utility sufficiently much, they will actively choose to stay uninformed. Observe, however, that such a model predicts that the highest effort level chosen by the uninformed subjects is at most $e^*(w_H)$. It cannot explain our finding that in the aggregate subjects with an unknown wage perform better than $e^*(w_H)$. Hence, in order to fully explain our results, we must incorporate the concept of "choking" into the BP model. We posit that some agents' productivity may be highest at a wage strictly below w_H. For a known wage these agents' performance may still be as in the standard model, but their effort will be highest at a subjective wage $\bar{w} < w < w_H$. Not knowing the wage might therefore induce them to be most productive. Moreover, if they care enough about anticipatory utility, they will also choose not to learn the wage before they start working.

4.5 Conclusion

We considered a real-effort task in which agents can choose to receive instrumental information about their piece rate before starting to work. Our data show that more than 30 percent of subjects deliberately decide to forgo learning their piece rate, revealing a preference for information avoidance. Furthermore, agents avoiding information achieve considerably better performance results than agents opting for information.

In order to uncover potential causal effects of self-selection on performance results, we run a treatment in which subjects are forced to stay uninformed about their piece rate. Performance again turns out to be higher than under the certain high wage. Information avoidance, even if enforced instead of chosen, significantly enhances performance. Moreover, performance under information avoidance tends to be better than performance under the certain high piece rate. This is at odds with basic predictions from orthodox economic theory.

Looking into reasons why subjects choose to avoid information suggests that there may be different types of subjects: Some avoid information in order to avoid potential demotivation if the piece rate turns out to be low; others say that learning about a high piece rate could make them feel stressed and lead to choking under pressure. Both effects have been documented in the both the psychology and economics literature.

In an otherwise standard Brunnermeier and Parker (2005) model, we incorporate heterogeneity of agents and the possibility of choking under pressure. This extended version of BP captures all key patterns of our data. A substantial part of subjects appears to have biased their beliefs about the piece rate considerably upwards in order to stay motivated in the task. Other subjects may have behaviorally distorted their beliefs in order to avoid feelings of pressure.

By and large, our study documents that giving agents room to design their beliefs may not only be beneficial in contexts such as health (as has been documented by Oster et al., 2013 and Ganguly and Tasoff, 2015), but also in economic settings of paid effort exertion and performance in the workplace. While a direct randomization over piece rates may be unpopular with workers and unions, there are other more subtle and perfectly accepted ways of introducing uncertainty about effective pay in a firm. Any type of incentive scheme that introduces interdependencies between workers' payments generates scope for beneficial belief distortion and rampant overprovision of effort in contests may have one of its root causes in anticipatory utility.[18] Additionally, in many, typically large, firms workers are often paid in company shares as part of their salary. Since the individual worker probably has a negligible effect on the evolution of the share's value, spiecifically when he or she works on a lower hierarchy level within the company, this will introduce exogenous uncertainty into the worker's compensation. Our results suggest that this might be an effective way for a firm to increase workers' efforts. Employees may distort their

[18]For a detailed survey of the experimental literature on contests, see Dechenaux et al. (2015) who document evidence for overprovision of effort in a large number of different settings.

beliefs about the firms' future performance according to their needs, which increases their effort.

The possibility of inducing more effort with less pay is tantalizing. We believe that it offers much scope for further research.

Appendix A

Appendix for Chapter 2

A.1 Proof of Proposition 2.2

The notation is as follows: denote by $LM(1,1)$ the condition that it is a PE for the consumer to buy at p_L and to buy at p_M. Analogously, denote by $LM_u(1,0)$ and $LM_l(1,0)$ the conditions that it is a PE to only buy at p_L but not at p_M. Here u denotes the upper bound, that is the condition that it is not profitable to deviate by buying at p_M, and l denotes the lower bound, ruling out deviations to not buying at p_L. Finally denote by $LM(0,0)$ the condition that it is a PE not to buy at either price. Accordingly, I define $MH(1,1)$, $MH_u(1,0)$, $MH_l(1,0)$, and $MH(0,0)$. Using equation (2.3), these conditions can be written as:

$$\begin{aligned}
LM(1,1):\ & u - p_M - \mu_g(-u) + \frac{1}{2}\mu_m(p_L - p_M) - \frac{1}{2}\mu_m(p_L) - \frac{1}{2}\mu_m(p_M) \geq 0 \\
LM_u(1,0):\ & u - p_M + \frac{1}{2}\mu_g(u) - \frac{1}{2}\mu_g(-u) + \frac{1}{2}\mu_m(-p_M) - \frac{1}{2}\mu_m(p_L) \\
& \qquad + \frac{1}{2}\mu_m(p_L - p_M) \leq 0
\end{aligned}$$

$$\begin{aligned}
LM_l(1,0): &\ u - p_L + \frac{1}{2}\mu_g(u) - \frac{1}{2}\mu_g(-u) + \mu_m(-p_L) - \frac{1}{2}\mu_m(p_L) \geq 0 \\
LM(0,0): &\ u - p_L + \mu_g(u) + \mu_m(-p_L) \leq 0 \\
MH(1,1): &\ u - p_H - \mu_g(-u) + \frac{1}{2}\mu_m(p_M - p_H) - \frac{1}{2}\mu_m(p_M) - \frac{1}{2}\mu_m(p_H) \geq 0 \\
MH_u(1,0): &\ u - p_H + \frac{1}{2}\mu_g(u) - \frac{1}{2}\mu_g(-u) + \frac{1}{2}\mu_m(-p_H) - \frac{1}{2}\mu_m(p_M) \\
&\qquad + \frac{1}{2}\mu_m(p_M - p_H) \leq 0 \\
MH_l(1,0): &\ u - p_M + \frac{1}{2}\mu_g(u) - \frac{1}{2}\mu_g(-u) + \mu_m(-p_M) - \frac{1}{2}\mu_m(p_M) \geq 0 \\
MH(0,0): &\ u - p_M + \mu_g(u) + \mu_m(-p_M) \leq 0
\end{aligned}$$

When using the concept of preferred personal equilibrium (PPE) the notation will be as follows. $LM(1,1 \succ 1,0)$ describes the condition that (in case LM) $LM(1,1)$ is preferred over $LM(1,0)$ based on ex-ante utilities. The rest follows analogously, and the conditions are as follows (only stating the ones needed):

$$\begin{aligned}
LM(1,0 \succ 1,1): &\ \frac{1}{2}u - \frac{1}{2}p_M - \frac{1}{4}\mu_g(-u) - \frac{1}{4}\mu_g(u) - \frac{1}{4}\mu_m(-p_L) \\
&\qquad - \frac{1}{4}\mu_m(p_L) + \frac{1}{4}\mu_m(p_L - p_M) + \frac{1}{4}\mu_m(p_M - p_L) \leq 0 \\
LM(0,0 \succ 1,1): &\ u - \frac{1}{2}p_M - \frac{1}{2}p_L + \frac{1}{4}\mu_m(p_L - p_M) + \frac{1}{4}\mu_m(p_M - p_L) \leq 0 \\
MH(1,1 \succ 0,0): &\ u - \frac{1}{2}p_H - \frac{1}{2}p_M + \frac{1}{4}\mu_m(p_M - p_H) + \frac{1}{4}\mu_m(p_H - p_M) \geq 0 \\
MH(1,0 \succ 0,0): &\ \frac{1}{2}u - \frac{1}{2}p_M + \frac{1}{4}\mu_m(-p_M) + \frac{1}{4}\mu_m(p_M) + \frac{1}{4}\mu_g(u) \\
&\qquad + \frac{1}{4}\mu_g(-u) \geq 0
\end{aligned}$$

Proposition 2.2 is proven in two parts. First, I will concentrate on establishing the result by only looking at PE. What we are interested in is always a combination of PE. To make statements about how behaviour differs between cases LM and MH, one has to state the PE in case LM and the corresponding PE in the case MH. The idea behind Lemma A.1-A.5 is that the existence of a specific PE in case LM rules out the existence of some PE in case MH and vice versa. By doing so, I will be able to establish that the combination of PE according to Proposition 2.2 exists. There is a personal equilibrium where, in case LM, the agent buys at both prices p_L and p_M, but the same agent does not buy at p_M (neither at p_H) in case MH. Furthermore, I show that the conditions on the prices stated above ensure that in cases where at least one of the PE is not unique the result survives under PPE. Finally, one can then establish that there never exists a consumer who will behave as in the good deal model, because - keeping all other parameters fixed - as we increase u starting from $u = 0$, the following cases of buying behaviour at p_M will always (and never any other) exist. An agent either never buys at p_M in any of the two cases, or he only buys at p_M in case LM, that is if he expected the prices to be p_L and p_M, or he buys at p_M in both LM and MH. The first and third case then imply the same buying behaviour across LM and MH whereas the second predicts the behaviour stated in Proposition 2.2.

As described in the text, the conditions for the functions $\mu_g(\cdot)$ and $\mu_m(\cdot)$ are as follows (see Bowman et al. (1999) or Kőszegi and Rabin (2006)):

A0. $\mu(x)$ is continuous for all x, twice differentiable for $x \neq 0$, and $\mu(0) = 0$.

A1. $\mu(x)$ is strictly increasing.

A2. If $y > x > 0$, then $\mu(y) + \mu(-y) < \mu(x) + \mu(-x)$.

A3. $\mu''(x) \leq 0$ for $x > 0$, and $\mu''(x) > 0$ for $x < 0$.

A4. $\mu'_-(x)/\mu'_+(x) \equiv \lambda > 1$,

where $\mu'_+(x) \equiv \lim_{x \to 0} \mu'(|x|)$ and $\mu'_-(x) \equiv \lim_{x \to 0} \mu'(-|x|)$

I furthermore make the following assumption—as discussed in the main text—that:

$$\mu_m(-p_M) < -\mu_m(p_L) + \mu_m(p_L - p_M) \tag{C1}$$

Note that when using the linear specification, this can be written as

$$-\lambda_m \eta_m p_M < -\eta_m p_L + \lambda_m \eta_m (p_L - p_M)$$

which can be simplified to $0 < \eta_m(\lambda_m - 1)p_L$ to see that this always holds.

In terms of PE one can then establish that:

Lemma A.1. *There exists a combination of PE such that $LM(1,1)$ and $MH(0,0)$ are satisfied*

Proof. Rewriting $LM(1,1)$ and $MH(0,0)$ yields:

$$u - p_M \geq \mu_g(-u) - \frac{1}{2}\mu_m(p_L - p_M) + \frac{1}{2}\mu_m(p_L) + \frac{1}{2}\mu_m(p_M)$$
$$u - p_M \leq -\mu_g(u) - \mu_m(-p_M) \leq 0$$

It can be seen that the RHS of the second equation is strictly larger than the RHS of the first equation whenever $-\mu_g(-u) > \mu_g(u)$ and $-\mu_m(-p_M) - \mu_m(p_M) > -\mu_m(p_L - p_M) + \mu_m(p_L) + \mu_m(-p_M)$. According to A2, the first inequality is satisfied, and the LHS of the second inequality is strictly positive. It remains to show that the RHS of the second inequality is non-positive. This directly follows from (C1). □

For fixed η, λ, all u satisfying this are within $[\alpha_1, \alpha_2]$, that is $u \equiv \alpha_1$ solves $LM(1,1)$ with equality, and $u \equiv \alpha_2$ solves $MH(0,0)$ with equality.

Lemma A.2. *$MH_l(1,0)$ is satisfied only if $LM(1,1)$ is, and there are values for u where $LM(1,1)$ is a PE but $MH(1,0)$ is not.*

Proof. Using the expressions for $LM(1,1)$ and $MH_l(1,0)$ from above,

$$\begin{aligned} u - p_M - \mu_g(-u) + \frac{1}{2}\mu_m(p_L - p_M) - \frac{1}{2}\mu_m(p_L) - \frac{1}{2}\mu_m(p_M) &\geq 0 \\ u - p_M + \frac{1}{2}\mu_g(u) - \frac{1}{2}\mu_g(-u) + \mu_m(-p_M) - \frac{1}{2}\mu_m(p_M) &\geq 0 \end{aligned}$$

implies that we need to show that the LHS of the first equation is larger than the LHS of the second. (This is the more general version of the comparison of (2.4') and (2.5') in the main text.) This is the case whenever $-\mu_g(-u) > \mu_g(u)$ and $-\mu_m(-p_M) - (\mu_m(-p_M) - \mu_m(p_L - p_M) + \mu_m(p_L)) > 0$. According to A2, the first inequality holds. Also, the first term on the LHS of the second inequality is positive, and the term in brackets is non-positive (C1). □

Lemma A.3. *If $p_H - p_M \geq p_M - p_L$, $MH(1,1)$ is satisfied only if $LM(1,1)$ is, and there are values for u where $LM(1,1)$ is a PE but $MH(1,1)$ is not.*

Proof. Using the expressions for $LM(1,1)$ and $MH(1,1)$ from above,

$$\begin{aligned} u - p_M - \mu_g(-u) + \frac{1}{2}\mu_m(p_L - p_M) - \frac{1}{2}\mu_m(p_L) - \frac{1}{2}\mu_m(p_M) &\geq 0 \\ u - p_H - \mu_g(-u) + \frac{1}{2}\mu_m(p_M - p_H) - \frac{1}{2}\mu_m(p_M) - \frac{1}{2}\mu_m(p_H) &\geq 0 \end{aligned}$$

implies that we need to show that the LHS of the first equation is larger than the LHS of the second. Since $\mu_m(\cdot)$ is strictly increasing (A1), $-\mu_m(p_L) > -\mu_m(p_H)$ and $\mu_m(p_L - p_M) \geq \mu_m(p_M - p_H)$, provided that $p_H - p_M \geq p_M - p_L$. □

Denote by β_1 the smallest value of u that satisfies both the conditions $MH_l(1,0)$ and $MH(1,1)$, hence it will solve (at least) one of the two with equality. According to Lemma A.2 and A.3, $\beta_1 > \alpha_1$.

Lemma A.4. *$LM_u(1,0)$ is satisfied only if $MH(0,0)$ is, and there are values for u where $MH(0,0)$ is a PE but $LM(1,0)$ is not.*

Proof. Using the expressions for $LM_u(1,0)$ and $MH(0,0)$ from above,

$$u - p_M + \frac{1}{2}\mu_g(u) - \frac{1}{2}\mu_g(-u) + \frac{1}{2}\mu_m(-p_M) - \frac{1}{2}\mu_m(p_L) + \frac{1}{2}\mu_m(p_L - p_M) \leq 0$$
$$u - p_M + \mu_g(u) + \mu_m(-p_M) \leq 0$$

implies that we need to show that the LHS of the first equation is larger than the LHS of the second. This is the case because $-\mu_g(-u) > \mu_g(u)$ (following from A2) and $-\mu_m(-p_M) > \mu_m(p_L) - \mu_m(p_L - p_M)$, which is exactly (C1). □

Lemma A.5. *$LM(0,0)$ is satisfied only if $MH(0,0)$ is, and there are values for u where $MH(0,0)$ is a PE but $LM(0,0)$ is not.*

Proof. Using the expressions for $LM(0,0)$ and $MH(0,0)$ from above,

$$u - p_L + \mu_g(u) + \mu_m(-p_L) \leq 0$$
$$u - p_M + \mu_g(u) + \mu_m(-p_M) \leq 0$$

implies that we need to show that the LHS of the first equation is larger than the LHS of the second, which is immediate from $p_M > p_L$ and $\mu_m(\cdot)$ increasing (assumption A1). □

Denote by β_2 the largest value of u that satisfies both the conditions $LM_u(1,0)$ and $LM(0,0)$, hence it will solve (at least) one of the two with equality. According to Lemma A.4 and A.5, $\beta_2 < \alpha_2$.

If $\beta_1 > \beta_2$, for $u \in [\beta_2, \beta_1]$, combining the statements from above, the unique PE in case LM is $LM(1,1)$ and in case MH it is $MH(0,0)$. For any

$u < \beta_2$, $MH(0,0)$ is the unique PE in case MH, and for any $u > \beta_1$, $LM(1,1)$ is the unique PE in case LM.

If $\beta_1 \leq \beta_2$, for $u > \beta_2$, $LM(1,1)$ is the unique PE in case LM. For $u < \beta_1$, $MH(0,0)$ is the unique PE in case MH. For $u \in [\beta_1, \beta_2]$ and $\beta_1 < \beta_2$ I need to rely on the PPE as the selection criterion. Outside this interval, one can see that for $u < \beta_1$ the combination of PE that can exist will either involve not buying at p_M in both cases LM and MH, or not buying at p_M in case MH but buying at p_M in case LM. Also, for $u > \beta_2$, in case LM, the agent will buy at p_M and either buy at p_M or not buy at p_M in case MH. In short, outside $[\beta_1, \beta_2]$ buying behaviour at p_M will either be the same in both cases, or the agent will only buy at p_M in case LM but not in case MH. What is important here for now is not which exact combination results (which will depend on what the PPE is) but that there can be no behaviour in accordance with the effect predicted by the good deal model. Inside $[\beta_1, \beta_2]$ there will be multiple PE in both cases hence I need to ensure that there can be no behaviour in accordance with the good deal model when looking at ex-ante utilities. This is what the following Lemmas establish:

Lemma A.6. *If $p_H - p_M \geq p_M - p_L$ and $2p_L \geq p_M$, then $LM(1, 0 \succ 1, 1)$ and $MH(1, 1 \succ 0, 0)$ cannot hold jointly.*

Proof. The conditions $LM(1, 0 \succ 1, 1)$ and $MH(1, 1 \succ 0, 0)$ can be written as:

$$\begin{aligned}\frac{1}{2}u - \frac{1}{2}p_M - \frac{1}{4}\mu_g(-u) - \frac{1}{4}\mu_g(u) - \frac{1}{4}\mu_m(-p_L) - \frac{1}{4}\mu_m(p_L) \\ + \frac{1}{4}\mu_m(p_L - p_M) + \frac{1}{4}\mu_m(p_M - p_L) \leq 0 \\ \frac{1}{2}u - \frac{1}{4}p_H - \frac{1}{4}p_M + \frac{1}{8}\mu_m(p_M - p_H) + \frac{1}{8}\mu_m(p_H - p_M) \geq 0\end{aligned}$$

To see that the LHS of the first equation is always larger than the LHS of the second note first that $\frac{1}{2}p_M < \frac{1}{4}(p_M + p_H)$. By A2, $-\frac{1}{4}\mu_g(-u) - \frac{1}{4}\mu_g(u) > 0$, and it remains to show that $-\frac{1}{4}\mu_m(-p_L) - \frac{1}{4}\mu_m(p_L) + \frac{1}{4}\mu_m(p_L - p_M) + \frac{1}{4}\mu_m(p_M - p_L) \geq \frac{1}{8}\mu_m(p_M - p_H) + \frac{1}{8}\mu_m(p_H - p_M)$. A sufficient condition (using A2) for this is $p_H - p_M \geq p_M - p_L$ and $\mu_m(p_L - p_M) + \mu_m(p_M - p_L) \geq 2\mu_m(-p_L) + 2\mu_m(p_L)$. Again using A2, we note that if $2p_L \geq p_M$, the latter inequality is satisfied. If $\mu_m(\cdot)$ is linear as in the main text, this condition can be relaxed to $3p_L \geq p_M$. □

Lemma A.7. *If $p_H - p_M \geq p_M - p_L$, $LM(0,0 \succ 1,1)$ and $MH(1,1 \succ 0,0)$ cannot hold jointly.*

Proof. Compare $LM(0,0 \succ 1,1)$ and $MH(1,1 \succ 0,0)$:

$$u - \frac{1}{2}p_M - \frac{1}{2}p_L + \frac{1}{4}\mu_m(p_L - p_M) + \frac{1}{4}\mu_m(p_M - p_L) \leq 0$$
$$u - \frac{1}{2}p_H - \frac{1}{2}p_M + \frac{1}{4}\mu_m(p_M - p_H) + \frac{1}{4}\mu_m(p_H - p_M) \geq 0$$

The fact that the LHS of the first equation is larger than the LHS of the second is immediate from A2 since the condition $p_H - p_M \geq p_M - p_L$ ensures that $\frac{1}{4}\mu_m(p_L - p_M) + \frac{1}{4}\mu_m(p_M - p_L) > \frac{1}{4}\mu_m(p_M - p_H) + \frac{1}{4}\mu_m(p_H - p_M)$. □

Lemma A.8. *$LM(1,0 \succ 1,1)$ and $MH(1,0 \succ 0,0)$ cannot hold jointly.*

Proof. Compare $LM(1,0 \succ 1,1)$ and $MH(1,0 \succ 0,0)$:

$$\frac{1}{2}u - \frac{1}{2}p_M - \frac{1}{4}\mu_g(-u) - \frac{1}{4}\mu_g(u) - \frac{1}{4}\mu_m(-p_L) - \frac{1}{4}\mu_m(p_L)$$
$$+\frac{1}{4}\mu_m(p_L - p_M) + \frac{1}{4}\mu_m(p_M - p_L) \leq 0$$
$$\frac{1}{2}u - \frac{1}{2}p_M + \frac{1}{4}\mu_m(-p_M) + \frac{1}{4}\mu_m(p_M) + \frac{1}{4}\mu_g(u) + \frac{1}{4}\mu_g(-u) \geq 0$$

In order to show that the LHS of the first equation is larger than the LHS of the second, we note first that $-\frac{1}{4}\mu_g(-u) - \frac{1}{4}\mu_g(u) > 0$ by A2. It then remains to show that $-\mu_m(-p_L) - \mu_m(p_L) + \mu_m(p_L - p_M) + \mu_m(p_M - p_L) > \mu_m(-p_M) +$

$\mu_m(p_M)$. Again from A2, it follows that $-\mu_m(-p_L) - \mu_m(p_L) > 0$ and, since $p_M > p_M - p_L$, that $\mu_m(p_L - p_M) + \mu_m(p_M - p_L) > \mu_m(-p_M) + \mu_m(p_M)$. □

Lemma A.9. *$LM(0,0 \succ 1,1)$ and $MH(1,0 \succ 0,0)$ cannot hold jointly.*

Proof. The conditions $LM(0,0 \succ 1,1)$ and $MH(1,0 \succ 0,0)$ can be written as:

$$\frac{1}{2}u - \frac{1}{4}p_M - \frac{1}{4}p_L + \frac{1}{8}\mu_m(p_L - p_M) + \frac{1}{8}\mu_m(p_M - p_L) \leq 0$$
$$\frac{1}{2}u - \frac{1}{2}p_M + \frac{1}{4}\mu_m(-p_M) + \frac{1}{4}\mu_m(p_M) + \frac{1}{4}\mu_g(u) + \frac{1}{4}\mu_g(-u) \geq 0$$

To show that the LHS of the first equation is larger than the LHS of the second, we first note that $\frac{1}{2}p_M > \frac{1}{4}(p_M+p_L)$ and, by A2, $\mu_g(u)+\mu_g(-u) < 0$. It remains to show that $\frac{1}{8}\mu_m(p_L - p_M) + \frac{1}{8}\mu_m(p_M - p_L) > \frac{1}{4}\mu_m(-p_M) + \frac{1}{4}\mu_m(p_M)$. Again by A2, $\mu_m(p_L - p_M) + \mu_m(p_M - p_L) > \mu_m(-p_M) + \mu_m(p_M)$ and $\mu_m(-p_M) + \mu_m(p_M) < 0$ which establishes the claim. □

Any combination that would predict the effect as in the good deal model has to have either $LM(1,0 \succ 1,1)$ or $LM(0,0 \succ 1,1)$ in case LM, and either $MH(1,1 \succ 0,0)$ or $MH(1,0 \succ 0,0)$ in case MH. Lemmas A.6 to A.9 show that any of the four resulting combinations are infeasible.

Then, only by looking at ex-ante utilities, there will be a range of u where $MH(0,0)$ and either $LM(0,0)$ or $LM(1,0)$ (not buying at p_M in both cases) is a combination of plans with the highest ex-ante utility. For u outside this range (larger) there exist values of u with $MH(0,0)$ and $LM(1,1)$ as a combination (buy at p_M only in case LM). Finally, for u even larger, $LM(1,1)$ will coexist with either $MH(1,0)$ or $LM(1,1)$ (buy at p_M in both cases). Define γ_1 and γ_2 with $\gamma_2 > \gamma_1$ as the two cutoff levels, that is for $u \in [\gamma_1, \gamma_2]$ the combination of $MH(0,0)$ and $LM(1,1)$ exists.

Remember that from above, I identified the region $[\beta_1, \beta_2]$ as the one where multiple PE exist in both cases simultaneously. By looking at ex-ante utilities I showed that in this region one can exclude all combinations that would yield the effect as in the good deal model. Now the final step to prove the Proposition is to show that there will always exist a region of u's where one gets the effect described in Proposition 2.2, that is $LM(1,1)$ and $MH(0,0)$ as together as (P)PE. Step by step denoting the region where it exists:

- $\gamma_2 > \gamma_1 \geq \beta_2 \geq \beta_1$: desired combination in $[\beta_2, \min\{\alpha_2, \gamma_2\}]$
- $\gamma_2 \geq \beta_2 \geq \gamma_1 \geq \beta_1$: desired combination in $[\gamma_1, \min\{\alpha_2, \gamma_2\}]$
- $\gamma_2 \geq \beta_2 > \beta_1 \geq \gamma_1$: desired combination in $[\max\{\alpha_1, \gamma_1\}, \min\{\alpha_2, \gamma_2\}]$
- $\beta_2 \geq \gamma_2 \geq \beta_1 \geq \gamma_1$: desired combination in $[\max\{\alpha_1, \gamma_1\}, \gamma_2]$
- $\beta_2 \geq \gamma_2 > \gamma_1 \geq \beta_1$: desired combination in $[\gamma_1, \gamma_2]$
- $\beta_2 \geq \beta_1 \geq \gamma_2 > \gamma_1$: desired combination in $[\max\{\alpha_1, \gamma_1\}, \beta_1]$

Note that these results also account for the case where $\beta_1 = \beta_2$.

A.2 Proposition 2.2 for probabilities different from one-half

This section shows how the results presented above can be generalised for probabilities different from one-half. For this, define $q_H \equiv \Pr(p = p_H)$ and $q_L \equiv \Pr(p = p_L)$. For Proposition 2.2 to hold for general probabilities, all Lemmas from above still need to hold. Therefore, the procedure is straightforward, simply establishing conditions on q_L, q_H, p_L, p_M, p_H such that all Lemmas continue to hold. For simplicity of exposition, I concentrate on the case where $\mu_k(\cdot)$ is piecewise linear, as in the main text. I will omit the calculations and state the conditions needed, Lemma by Lemma.

- Lemma A.2: $q_H p_M \geq q_L(p_M - p_L)$
- Lemma A.3: $(1 - q_H)(p_H - p_M) \geq q_L(p_M - p_L)$
- Lemma A.6: $(1 - q_H)q_H(p_H - p_M) \geq q_L(p_M - 2p_L)$
- Lemma A.7: $(1 - q_H)q_H(p_H - p_M) \geq (1 - q_L)q_L(p_M - p_L)$
- Lemma A.8: $q_H p_M \geq q_L(p_M - 2p_L)$
- Lemma A.9: $q_H p_M \geq (1 - q_L)q_L(p_M - p_L)$

A.3 Proofs for claims in section 2.2.3

If the distribution of possible prices is given by $p \sim U[a, b]$, with $b > a \geq 0$, the utility from buying at a price p, given a PE strategy to buy at every price $p \leq \hat{p}$ is given by:

$$u - p + \eta_g u(1 - F(\hat{p})) + \eta_m \lambda_m \int_a^p (r - p)dF(r) + \eta_m \int_p^{\hat{p}} (r - p)dF(r) - \eta_m \lambda_m (1 - F(\hat{p}))p$$

Deviating from this strategy and not buying at price p yields utility:

$$\eta_m \int_a^{\hat{p}} rdF(r) - \eta_g \lambda_g u F(\hat{p})$$

Hence, the PE "buy at all prices $p \leq \hat{p}$" is given by the $p = \hat{p}$ that equates these two expressions:

$$Z(\hat{p}) = u(1+\eta_g) - \hat{p}(1+\eta_m\lambda_m) + \eta_m(\lambda_m - 1)\int_a^{\hat{p}} rdF(r) + \eta_g(\lambda_g - 1)F(\hat{p})u = 0$$

As $Z(\hat{p})$ is positive for $\hat{p} \to 0$ and negative for $\hat{p} \to \infty$ such a $\hat{p}$ will always exist. Furthermore, let us assume that, as also assumed in Kőszegi and Rabin (2004), $(\eta_g(\lambda_g - 1)u + \eta_m(\lambda_m - 1)b) < (1 + \eta_m\lambda_m)(b - a)$, which implies that $Z(\hat{p})$ is decreasing everywhere on $[a, b]$ and therefore there exists a unique PE.

Now consider two distributions $F_1(p) = \frac{p-a}{b_1-a}$ and $F_2(p) = \frac{p-a}{b_2-a}$, with $b_2 > b_1$ and let $\hat{p}_1$ be the PE corresponding to $F_1(p)$ and $\hat{p}_2$ the PE corresponding to $F_2(p)$. We want to show that $\hat{p}_1 > \hat{p}_2$. Define $Z_i(\hat{p})$ as the function $Z(\hat{p})$ when the distribution of prices is F_i. The uniform distribution of F_1 and F_2 then implies that $\int_a^{\hat{p}_2} r dF_1(r) > \int_a^{\hat{p}_2} r dF_2(r)$ and $F_1(\hat{p}_2) > F_2(\hat{p}_2)$. From this it follows that $Z_1(\hat{p_2}) > Z_2(\hat{p_2}) = 0$. Since Z_i is decreasing it follows immediately that $\hat{p}_2 < \hat{p}_1$.

When proving the claim made about reservation prices in the Mazar et al. (2014) setting, we focus on reservation prices that are strictly interior, $\hat{p} \in (a, b)$. This makes the analysis more tractable and allows us, together with the assumption of a unique PE in this interval, to focus on PE behaviour. Denote by $Z_a(\hat{p})$ and $Z_b(\hat{p})$ the function $Z(\hat{p})$ as defined above when $F_a(p)$ (right-skewed distribution) and $F_b(p)$ (left-skewed distribution) are defined as follows:

$$F_a(p) = \begin{cases} \frac{1}{2}, & \text{if } p = a \\ \frac{1}{2} + \frac{1}{2}\frac{p-a}{b-a} = \frac{1}{2} + \frac{1}{2}F(p), & \text{if } a < p \leq b \end{cases}$$

$$F_b(p) = \begin{cases} \frac{1}{2}\frac{p-a}{b-a} = \frac{1}{2}F(p), & \text{if } a \leq p < b \\ 1, & \text{if } p = b \end{cases}$$

We obtain the following expressions for $Z_a(\hat{p})$ and $Z_b(\hat{p})$:

$$Z_a(\hat{p}) = u - \hat{p}(1+\eta_m\lambda_m) + \frac{1}{2}\eta_m(\lambda_m - 1)\int_a^{\hat{p}} rdF(r) + \frac{1}{2}\eta_g(\lambda_g - 1)F(\hat{p})u + \frac{1}{2}\eta_g\lambda_g u + \frac{1}{2}\eta_m(\lambda_m - 1)a$$

$$Z_b(\hat{p}) = u - \hat{p}(1+\eta_m\lambda_m) + \frac{1}{2}\eta_m(\lambda_m - 1)\int_a^{\hat{p}} rdF(r) + \frac{1}{2}\eta_g(\lambda_g - 1)F(\hat{p})u + \frac{1}{2}\eta_g u$$

Analogously to above, we assume that $\frac{1}{2}(\eta_g(\lambda_g - 1)u + \eta_m(\lambda_m - 1)b) < (1 + \eta_m\lambda_m)(b-a)$, which guarantees that both functions are decreasing everywhere on $[a, b]$. Define $Z_a(\hat{p}_a) = 0$ and $Z_b(\hat{p}_b) = 0$, i.e. $\hat{p}_a$ and $\hat{p}_b$ are the PE prices in the respective cases. We then note that $Z_a(\hat{p}) > Z_b(\hat{p})$ and it is immediate that since $Z_a(\hat{p})$ and $Z_b(\hat{p})$ are decreasing, this implies $\hat{p}_a > \hat{p}_b$.

Appendix B

Appendix for Chapter 3

B.1 Theory

B.1.1 Add-on Price Experiment with Risk Averse Agents

Under risk aversion, the optimisation problem of the seller, when buyers are assumed to be naive about their biased beliefs, becomes:

$$\max_{p,f} \quad \pi^S u(p+f) + (1-\pi^S)u(p)$$

$$\text{s.t.} \quad \pi^B u(x-p-f) + (1-\pi^B)u(x-p) \geq u(y) \tag{B.1}$$

$$p + f \leq x \tag{B.2}$$

$$p, f \geq 0 \tag{B.3}$$

where we assume that u is increasing and concave and $u(0) = 0$. Clearly, the constraint (B.1) will be binding in the optimum. Otherwise the seller could increase either p or f or both and the offer would still be accepted. In other words, the buyer's utility from accepting the offer will be equal to the outside option, $u(y)$. Momentarily ignoring the other constraints, the first order condition for an interior solution is given by:

$$\frac{1-\pi^S}{1-\pi^B}\frac{\pi^B}{\pi^S} = \frac{u'(p+f)}{u'(p)}\frac{u'(x-p)}{u'(x-p-f)} \tag{B.4}$$

If $\pi^S = \pi^B$, the LHS in (B.4) is equal to one, and therefore the only way to satisfy the equation is by choosing $f^* = 0$. Then, with (B.1) holding with equality, we have $p^* = x - y$. This satisfies (B.2) and (B.3). It is straightforward to check that (B.2) holding with equality cannot be optimal and no other solutions to this problem exist.

Naive Buyers. If $\pi^S > \pi^B$ and the buyer is naive, the LHS in (B.4) is less than one. Hence, to satisfy the first order condition for an interior solution it has to be the case that $f^* > 0$: since u' is decreasing, this ensures that both fractions on the RHS will also be less than one. This candidate solution has $x > p^* + f^*$ and $p^* > 0$, and therefore (B.2) and (B.3) are satisfied. The first order condition together with (B.1) holding with equality will determine the exact values for p^*, f^*. In addition there exists a corner solution which is given by $f^* = x - p^*$ and $(1-\pi^B)u(x-p^*) = u(y)$. Which one of the two will be the global optimum depends on the exact functional form of $u(\cdot)$. More important than solving for the explicit solution is to note that it is guaranteed that the optimal solution has $f^* > 0$. Together with the observation that (B.1) holds with equality, this ensures that Prediction 3.2 continues to hold when agents are risk averse. Seller profits are decreasing in π^B for $\pi^B \leq \pi^S$.

Sophisticated Buyers. The case of sophistication proceeds along the same lines as described in the main text (see sections 3.2 and 3.3.3). It is straightforward to check that Prediction 3.3 continues to hold. As there are no new insights gained from reproducing it, I refrain from doing so.

B.1.2 Add-on Price Experiment with Inequity Averse Agents

The model of Fehr and Schmidt (1999) has routinely been used to explain the behaviour in ultimatum games, especially the rejection of offers with a positive material payoff to the responder. In game with two players, the overall utility of agent i with material payoff X_i and an opponent j with payoff X_j is given by:

$$U(X_i, X_j) = X_i - \alpha \max\{X_j - X_i, 0\} - \beta \max\{X_i - X_j, 0\}$$

Here, α captures the disutility from disadvantageous inequality, whereas $\beta \leq \alpha$ captures the disutility from advantageous inequality. For the add-on price experiment, this implies that the utility of a buyer with belief π^B who receives the offer p, f is given by:

$$\begin{aligned} U^B(\mathit{accept}) = &\; x - p - \pi^B f - \alpha \max\{2(p + \pi^B f) - x, 0\} \\ &\qquad - \beta \max\{x - 2(p + \pi^B f), 0\} \\ U^B(\mathit{reject}) = &\; y \end{aligned}$$

In this formulation, we assume that when calculating relative payoffs, agents compare expected payoffs, as discussed in footnote 16. Assume first that the buyer perceives the offer as disadvantageous, that is $2(p + \pi^B f) \geq x$. In this case, he will accept the offer if

$$p + \pi^B f \leq \frac{1+\alpha}{1+2\alpha} x - \frac{1}{1+2\alpha} y \tag{B.5}$$

Since $x > 2y$ in our setting, the RHS is decreasing in α and therefore the higher α, the better the offer needs to be in order for the buyer to accept.

He dislikes having much less than the seller and will therefore reject offers that are very unequal. Also observe that, no matter what value α takes, the buyer will always accept an offer that gives both parties an equal payoff. If $2(p + \pi^B f) < x$, the buyer perceives an offer as advantageous and accepts if

$$(p + \pi^B f)(1 - 2\beta) \leq (1 - \beta)x - y \tag{B.6}$$

The sellers' utility is given by

$$\begin{aligned} U^S(\textit{accept}) =\ & p + \pi^S f - \alpha \max\{x - 2(p + \pi^S f), 0\} \\ & - \beta \max\{2(p + \pi^B f) - x, 0\} \\ U^S(\textit{reject}) =\ & y \end{aligned}$$

It is clear that the seller will never offer $p + \pi^S f < \frac{1}{2}x$, that is, a contract where he obtains a smaller payoff than the buyer. This is because the offer $p = \frac{1}{2}x$ and $f = 0$ has a strictly higher payoff and will surely be accepted by the buyer as it leads to payoffs without any inequality. This also means that a seller who has $\beta > 0.5$ will always make an offer with an expected payoff of $\frac{x}{2}$. For all sellers with $\beta < 0.5$, an increase in own material payoff is always preferred to a reduction in inequality of the same size, and he will thus choose the offer that makes the buyer indifferent between accepting and rejecting. If buyers and sellers have the same belief, $\pi^B = \pi^S$, this optimal offer is given by

$$p = \frac{1 + \alpha}{1 + 2\alpha}x - \frac{1}{1 + 2\alpha}y \quad \text{and} \quad f = 0 \tag{B.7}$$

or any other offer that yields the same expected payoff.

Naive Buyers. The aim of this section is to show that if buyers are naive, there is still scope for sellers to exploit biased beliefs even if agents are inequity averse. In the case where $2(p+\pi^B f) \geq x$, equation (B.6) constrains the seller's optimal offer. A seller facing a naive buyer with $\pi^B < \pi^S$ will exploit the belief bias by setting

$$p = x - f \quad \text{and} \quad f = \frac{\alpha x + y}{(1+2\alpha)(1-\pi^B)}$$

This can be derived in the same way as before, by maximising the payment in the state where the product is faulty and then setting p and f to satisfy the participation constraint in (B.5). This offer yields the seller an expected profit of

$$\frac{1+\alpha}{1+2\alpha}x - \frac{1}{1+2\alpha}y + \frac{\pi^S - \pi^B}{1-\pi^B}\left(\frac{y}{1+2\alpha} + \frac{\alpha x}{1+2\alpha}\right)$$

This profit is larger than in the case where $\pi^B = \pi^S$. They differ by the last term in the previous equation (which is zero when buyer and seller share the same belief). Intuitively, inequity aversion for the buyers forces sellers to reduce their offers because unequal offers will be rejected. But since offers with $f > 0$ will continue to be evaluated wrongly by buyers with biased beliefs, sellers benefit from facing naive buyers. This can also be seen from equation (B.5): an increase in the degree of aversion to inequality (α) decreases the *subjective* expected price that buyers are willing to pay, no matter whether beliefs are biased or unbiased.

What is left to analyse is the case where $2(p + \pi^B f) < x$. From above it follows that a seller would never find it optimal to make an offer that has $2(p + \pi^S f) < x$, because such an offer is dominated by the equal split: $p = \frac{1}{2}x, f = 0$ and all other offers that give the seller the same expected payoff. Unless the buyer is extremely averse to disadvantageous inequality, the seller can always do better than this equal split. But since $\pi^B < \pi^S$, there even exist cases where the seller can do better than the equal split by offering him a contract that he subjectively perceives as having the same payoff as the equal split. Such an offer satisfies $p + \pi^B f = \frac{1}{2}x$ and $p + f = x$. It is straightforward to check that the seller earns more than $\frac{1}{2}x$, hence even for extreme values of α, exploitation of naive buyers is possible.

Sophisticated Buyers. Here it should be noted that the reasoning outlined in sections 3.2 and 3.3.3 continues to hold when allowing for agents being inequity averse. To see this, note that buyers who take into account the possibility that sellers have a different (and correct) belief will make the same inferences as before. They reason about the sellers' incentives taking advantage of the belief differences and would continue to reject all offers with $f > 0$ (unless they can be uniquely attributed to the highest type). The resulting equilibrium is therefore given by equation (B.7).

B.1.3 Insurance Experiment with Risk Averse Agents

As before, an agent evaluates his utility according to $u(\cdot)$ which is increasing and concave. As argued in the main text, under the constraint on p, an agent who does not receive additional utility from relative payoffs will always buy the product. A buyer with belief π^B will buy insurance for a fee f if and only if

$$u(x - p - f) \geq \pi^B u(x - p - L) + (1 - \pi^B) u(x - p)$$

Denote by $\hat{f}^B$ the value of f that solves this equation with equality, i.e. the maximum fee that a buyer is willing to pay for insurance. At the same time, a seller with belief π^S will sell insurance for f if and only if

$$\pi^S u(p + f - L) + (1 - \pi^S) u(p + f) \geq u(p)$$

Analogously, denote by $\hat{f}^S$ the value of f that solves for the seller's willingness to sell insurance. Before considering the case of biased beliefs, I will consider the market for insurance when $\pi^B = \pi^S$ and show, as already stated in the main text, that insurance might be sold even when beliefs are the same.

A standard result in the insurance literature (see, for example, Mossin, 1968) states that the demand for insurance depends on whether the utility function exhibits increasing or decreasing risk aversion. In the latter case, the willingness to pay for insurance decreases in wealth. If we, as before, for ease of notation ignore any wealth outside of the experiment, it will still be the case that buyer and seller have different endowments, depending on the price at which the product is sold. Importantly, for any p, f satisfying $p + f > x - p$ the buyer will have less experimental wealth and it can then be the case that if $u(\cdot)$ exhibits decreasing absolute risk aversion (DARA) there is a non-degenerate interval of insurance fees at which buyer and seller are willing to trade. Mossin (1968) contains the proof of f decreasing in wealth and a similar result (without proof) for trading environments appears in Camerer and Kunreuther (1989). In the special case of constant absolute risk aversion, there will be one value of f where there can be trade, just like in the risk-neutral case.

Naive Buyers. Despite the fact that insurance could be sold in the control condition, it will still be the case that under risk aversion sellers will still be able to exploit biased beliefs of buyers: From above $\hat{f}^B$ is increasing in π^B for any increasing $u(\cdot)$ which immediately implies that naive buyers will be overpaying for insurance. Hence, Prediction 3.5 continues to hold.

Sophisticated Buyers. Similarly, Prediction 3.6 is unaffected by introducing risk aversion. Sophisticated buyers will correctly infer that sellers would not sell insurance at prices that would lead them to be worse off which, by similar reasoning as in the main text, prevents them from overpaying for insurance.

B.1.4 Insurance Experiment with Inequity Averse Agents

As before, using the model of Fehr and Schmidt (1999), the buyer's utility for an offer p, f is given by:

$$\begin{aligned} U^B(\textit{product + insurance}) =\ & x - p - f - \alpha \max\{2(p+f) - x - \pi^B L, 0\} \\ & - \beta \max\{x + \pi^B L - 2(p+f), 0\} \\ U^B(\textit{product only}) =\ & x - p - \pi^B L - \alpha \max\{2p - x + \pi^B L, 0\} \\ & - \beta \max\{x - 2p - \pi^B L, 0\} \\ U^B(\textit{reject}) =\ & 0 \end{aligned}$$

and for the seller we have:

$$\begin{aligned} U^S(\textit{product + insurance}) =\ & p + f - \pi^B L \alpha \max\{x + \pi^B L - 2(p+f), 0\} \\ & - \beta \max\{2(p+f) - x - \pi^B L, 0\} \\ U^S(\textit{product only}) =\ & p - \alpha \max\{x - 2p - \pi^B L, 0\} \\ & - \beta \max\{2p - x + \pi^B L, 0\} \\ U^S(\textit{reject}) =\ & 0 \end{aligned}$$

As discussed in section B.1.2 a seller will never find it optimal to make an offer that would leave him worse off than a buyer. We can therefore focus on the cases where $2(p+f) \geq x + \pi^B L$ and $2p \geq x - \pi^B L$. It is then straightforward to show that the buyer's optimal decision is to buy the product with insurance if

$$p + f \leq \frac{1+\alpha}{1+2\alpha}x + \frac{\alpha}{1+2\alpha}\pi^B L \quad \text{and} \quad f \leq \pi^B L$$

Analogously, a buyer will buy the product only if

$$p \leq \frac{1+\alpha}{1+2\alpha}\left(x - \pi^B L\right) \quad \text{and} \quad f \geq \pi^B L$$

If $\pi^B = \pi^S = \pi$, for a seller with $\beta < 0.5$, an optimal offer will be given by $f = \pi L$ and $p = \min\{p^{max}, \frac{1+\alpha}{1+2\alpha}(x - \pi L)\}$. While there may be other offers that achieve the same profit, no offer can do better. With this offer, the buyer is indifferent between buying insurance or not and the seller does not earn any extra profit from it. A seller with $\beta \geq 0.5$ would always want to implement equal payoffs, which can be done by setting $p = \frac{1}{2}(x - \pi L)$ and $f = \pi L$.

Naive Buyers. To show that sellers can earn higher profits from naive buyers with biased beliefs, note that the offer

$$f = \pi^B L \quad \text{and} \quad p = \min\left\{p^{max}, \frac{1+\alpha}{1+2\alpha}\left(x - \pi^B L\right)\right\}$$

will lead to the buyer buying insurance (breaking the indifference in favour of insurance) and the seller to achieve a higher profit than before since he sell insurance for an "exploitative" fee. If p^{max} is binding, the seller benefits from the higher fee for insurance. Otherwise, total profits are given by

$$\frac{1+\alpha}{1+2\alpha}(x - \pi^S L) + \frac{\alpha}{1+2\alpha}(\pi^B L - \pi^S L)$$

The second term in the equation is positive and represents the extra profit brought about by the belief difference.

For completeness, I consider the (arguably extreme) case where α tends to infinity and the buyer therefore would only accept (perceived) equal payoffs. Even in this case the seller benefits from biased beliefs: offering $f = \pi^B L$ and $p + f = \frac{1}{2}(x + \pi^B L)$ would lead the buyer to accept, but *true* expected payoffs (evaluated at π^S) would still favour the seller.

Sophisticated Buyers. As described in section 3.5.3, if buyers are sufficiently inequity averse, there exists a PBE where insurance is sold in equilibrium even though buyers are sophisticated. For a given buyer belief π^B a buyer buys insurance if

$$f \leq \pi^B \quad \text{and} \quad p + f \leq \frac{1+\alpha}{1+2\alpha}x + \frac{\alpha}{1+2\alpha}\pi^B L,$$

buys the product only if

$$f \geq \pi^B \quad \text{and} \quad p \leq \frac{1+\alpha}{1+2\alpha}(x - \pi^B L),$$

and rejects the offer as a whole otherwise.

Using the notation from section 3, a seller's type is given by π (dropping the sub- and superscripts for convenience) and the commonly known prior beliefs of the buyer about the seller's type are given by $\mu(\pi)$. Denote by π^l the lowest, and by π^h the highest seller type. For the buyer, the strategy $b(p, f) = (a(p, f), \sigma(p, f))$ describes whether he accepts or rejects, $a \in \{0, 1\}$, and with which probability he buys insurance, $\sigma \in [0, 1]$, for any offer p, f. That is, I allow mixed strategies over the decision whether to buy insurance. For the seller, I focus on pure strategies of the form $s(\pi) = (p(\pi), f(\pi))$ which specify, for each type π the price p and the fee f that such a seller sets. To ease exposition, I present the results by (i) focusing on the case where only disadvantageous inequality on the buyer's side is relevant, implicitly assuming that all offers will leave the seller better off than the buyer, and (ii) assuming that for the seller $\beta < 0.5$ such that (as argued before) we can treat him as only caring about material payoff provided he is better off than the buyer. (ii) is justified by the interest in the case where the seller is actually willing to exploit because he cares more about his own payoff than about reducing inequality (the validity of this can also be checked in the data). I show below that the restriction (i) does not affect the equilibrium properties. Alternatively, one can think of a model of rivalistic preferences where (in Fehr and Schmidt's notation) the buyer has $\alpha = -\beta$, in which case the proposition covers all cases. A PBE is then defined as:

Definition B.1. *A perfect Bayesian equilibrium (PBE) is a strategy profile b^*, s^* and posterior beliefs for the buyer $\mu(\pi|p, f)$ such that:*

1. $\forall \pi \quad (p^*(\pi), f^*(\pi)) \in \arg\max_{p,f} a^*(p + \sigma^*(f - \pi L))$
2. $\forall p, f \quad (a^*(p, f), \sigma^*(p, f)) \in$

   $\arg\max_{a,\sigma} \sum_\pi a\left[(1+\alpha)(x - \pi L) - (1+2\alpha)(p + \sigma(f - \pi L))\right]\mu(\pi|p, f)$
3. $\mu(\pi|p, f) = \frac{\mu(\pi)}{\sum_{\{\pi'|p^*(\pi')=p, f^*(\pi')=f\}} \mu(\pi')}$ *if* $\sum_{\{\pi'|p^*(\pi)=p, f^*(\pi)=f\}} \mu(\pi') > 0$

Proposition B.1. *Define $\eta = \frac{1+\alpha}{1+2\alpha}$ and $\hat{\pi} = \sum_\pi \mu(\pi)\pi$. Note that since $\alpha > 0, \frac{1}{2} < \eta < 1$. Also, assume that $p^{max} > \eta(x - \hat{\pi}L)$. There exists a pooling equilibrium in which the seller's strategy is given by*

$$s^*(\pi) = (p^*(\pi), f^*(\pi)) = (\eta(x - \hat{\pi}L), \hat{\pi}L)$$

and if $p^{max} > \eta(x - \pi^l L)$ there also exists a separating equilibrium in which the seller's strategy is given by

$$s^*(\pi) = (p^*(\pi), f^*(\pi)) = (\eta(x - \pi L), \pi L)\,.$$

In either equilibrium the buyer's strategy is given by:

$$b^*(p,f) = (a^*(p,f), \sigma^*(p,f)) = \begin{cases} (1,1) & \textit{if } f < \pi^l L \textit{ and } p + f \leq \eta x + (1-\eta)\pi^l L \\ (1,\eta) & \textit{if } \pi^h L \geq f \geq \pi^l L \textit{ and } p \leq \eta(x - f) \\ (1,0) & \textit{if } f > \pi^h L \textit{ and } p \leq \eta(x - \pi^h L) \\ (0,0) & \textit{otherwise} \end{cases}$$

Therefore, in either of these equilibria, the buyer will always buy the product and buy insurance with probability η.

Proof. I start by proving the pooling equilibrium. On the equilibrium path, the buyer's belief will be given by $\pi^B = \hat{\pi}$. Hence, for the buyer, we need to check that when being offered p^*, f^*, there is no profitable deviation from $a^* = 1, \sigma^* = \eta$. Using the condition from above, we need to ensure that the buyer is indifferent between buying insurance and buying the product only. Since $f^* = \pi^B L$ and $p^* = \eta(x - \pi^B L)$, this is indeed the case.

Off the equilibrium path, buying insurance is clearly optimal when $f < \pi^l L$ because in this case the buyer is guaranteed to pay less than the expected loss, no matter what the objective probability of a loss. If $p + f \leq \eta x + (1-\eta)\pi^l L$ is satisfied there is no belief that would support a strategy of not buying at all. Analogously, if $f > \pi^h L$ the buyer cannot find it optimal to buy insurance but if $p \leq \eta(x - \pi^h L)$ the price for the product alone is sufficiently low that there is no belief that would justify rejecting the offer as a whole. For the cases where f is above $\pi^h L$ or below $\pi^l L$, but p does not satisfy the corresponding condition on p, we can always find a belief such that it is optimal for the buyer to reject. In all other cases, let the buyer choose the belief that $\pi^B = \frac{f}{L}$. In this case, he is indifferent between buying insurance or not, provided that $p \leq \eta(x - \pi^B L)$. Hence, given this belief it is optimal for him to mix with probability η if p

satisfies this inequality, and to reject otherwise.

Next, consider the case of the seller. In the pooling equilibrium, the profits of seller type π are given by

$$\eta(x - \hat{\pi}L) + \eta(\hat{\pi} - \pi)L = \eta(x - \pi L)$$

Deviating and making an offer that leads the buyer to always buy insurance yields profits of at most

$$\eta x - (1-\eta)\pi^l L + \pi L \leq \eta(x - \pi L),$$

with a strict inequality for $\pi > \pi^l$, and is therefore not profitable. Similarly, offering p, f such that the buyer will always buy the product without insurance yields profits of at most

$$\eta(x - \pi^h L) \leq \eta(x - \pi L)$$

with a strict inequality for $\pi < \pi^h$, and is therefore not profitable. Making an offer that will be rejected is clearly suboptimal, so the only other deviations to consider are $f' \neq f^*$ with $\pi^h L \geq f' \geq \pi^l L$ and $p^* \leq \eta(x - f')$. Clearly, the highest profit can be made when p' satisfies the inequality without slack, in which case the seller's profits are given by

$$\eta(x - f') + \eta(f' - \pi L) = \eta(x - \pi L)$$

This is equal to the profit in equilibrium and therefore does not constitute a profitable deviation either.

In the separating equilibrium, the buyer's belief will be given by $\pi^B = \frac{f^*}{L}$.

As for the pooling equilibrium, $p^* = \eta(x - \pi^B L)$ and the buyer is indifferent between buying insurance or not. Off the equilibrium path, all arguments from before go through in the same way, but for the case where the seller deviates to f' with $p' \leq \eta(x - f')$. Only if this f' does not correspond directly to a type πL, the buyer selects $\pi^B = \frac{f'}{L}$ as before and continues to mix with probability η, which is optimal given this belief.

For the seller, his profits in the separating equilibrium will also be given by $\eta(x - \pi L)$. All arguments from before go through given that the buyer's strategy is the same. Deviating to the equilibrium strategy of any other type does not increase seller's profits.

As noted above, I still need to check that the proposition also applies to cases where we allow for offers that yield a higher expected payoff for the buyer than the seller. I analyse the three cases of the buyer's equilibrium strategy in turn.

First, if $f < \pi^l L$, for a given buyer belief π^B, the buyer perceives that buying product and insurance together yields an expected payoff at least as high as the seller's if $x - p - f > p + f - \pi^B L$. As before, specifying the off-equilibrium belief as $\pi^B = \pi^l$, this case arises if $p + f \leq \frac{1}{2}x + \frac{1}{2}\pi^l L$. Straightforward calculations using the utility function at the beginning of the section and assuming that $\beta < 0.5$ reveal that such an offer will always be accepted. But since $\eta > \frac{1}{2}$ and $x > \pi^l L$, this does not alter the buyer's equilibrium strategy.

Similarly, any f with $\pi^h L \geq f \geq \pi^l L$ and the belief that $\pi^B = \frac{f}{L}$ means that the offer involves no disadvantageous inequality for the buyer if $p \leq \frac{1}{2}(x - f)$. Such an offer will always be accepted, the buyer is indifferent between buying insurance or not, and the equilibrium strategy captures this case as well.

Finally, for $f > \pi^h L$, and choosing the appropriate off-equilibrium belief, any $p \leq \frac{1}{2}(x - \pi^h L)$ will lead the buyer to buy the product only which is already implied by the condition $p \leq \eta(x - \pi^h L)$, as stated in the equilibrium strategy. □

For completeness, note that in the case where $\eta(x - \pi^l L) > p^{max} > \eta(x - \hat{\pi} L)$, there exists a partially separating equilibrium where a subset of types $\pi \in \{\pi^l, ..., \tilde{\pi}\}$ will set $p^* = p^{max}$ and $f^* = \sum_{\pi \in \{\pi^l,...,\tilde{\pi}\}} \mu(\pi)\pi$ and all types $\pi > \tilde{\pi}$ set $(p^*, f^*) = (\eta(x - \pi L), \pi L)$. The proof follows the same logic as above and is therefore omitted. Intuitively, as the fully separating equilibrium is no longer feasible because the lowest type would have to price above p^{max}, there will be pooling among low types such that all types price at p^{max}.

The following proposition shows that in all PBE of the game the buyer will mix over the insurance decision using the mixing probability η, as characterised in Proposition B.1.

Proposition B.2. *Assume $p^{max} > \eta(x - \hat{\pi} L)$. Also restrict attention to pure strategies for the seller. In any equilibrium, the buyer always buys the product and buys insurance with probability $\eta = \frac{1+\alpha}{1+2\alpha}$.*

The proof for this proposition proceeds in four lemmas. First, I rule out that there can be an equilibrium in which the buyer plays a pure strategy over the insurance decision. Then, I show that no other probability than η can be part of an equilibrium strategy profile.

Lemma B.1. *There is no equilibrium in which the buyer always buys the product, but never buys insurance.*

Proof. I will prove this by focusing on the lowest seller type π^l. Assume, to the contrary, that $p^*(\pi^l), f^*(\pi^l)$ is an equilibrium.

1. if, in equilibrium, $p^*(\pi^l) > \eta(x - \pi^l L)$, no buyer would ever buy from this seller, because there exists no belief that would make buying optimal.

2. if, in equilibrium, $f^*(\pi^l) < \pi^l L$, a buyer would always want to buy insurance (or not buy the product at all), this cannot be part of the proposed equilibrium.

3. if, in equilibrium, $p^*(\pi^l) < \min\{p^{max}, \eta(x - \pi^l L)\}$ and $f(\pi^l) \geq \pi^l L$, then there exists the following profitable deviation: The seller sets $p' = \min\{p^{max}, \eta(x - \pi^l L)\}$ and $f' = \pi^l L - \varepsilon$. At these prices the buyer must buy insurance for any belief because f' surely is a beneficial deal and $p' + f' < \eta x + (1 - \eta)\pi^B L$ for any possible belief. This is profitable for the seller because when the buyer buys insurance, he earns $p' + f' = \eta(x - \pi^l L) - \varepsilon$, which for small enough ε will earn a higher profit than $p^*(\pi^l)$.

4. if, in equilibrium, $p^*(\pi^l) = \eta(x - \pi^l L)$ and $f^*(\pi^l) \geq \pi^l L$, and the buyer buys the product but not insurance, then this equilibrium must be (partially) separating such that the lowest type is the only type setting this price-fee combination. Otherwise, the buyer's belief would be strictly higher than π^l and he would never buy the product. But if the only type setting this price is the lowest type, all other types could earn strictly more by mimicking this type. This rules out such equilibria.

5. if, in equilibrium, $p^*(\pi^l) = p^{max} < \eta(x - \pi^l L)$ and $f^*(\pi^l) \geq \pi^l L$, then by similar argument as in 4., not all sellers could set such a price in equilibrium, because then $\pi^B = \hat{\pi}$ and the buyer would not buy. But then, if only sellers with sufficiently low π set $p^* = p^{max}$, the other sellers have an incentive to deviate to p^{max}.

These five cases cover all possible combinations of p^* and f^* and the proof is complete. □

Lemma B.2. *There is no equilibrium in which the buyer always buys the product, and always buys insurance.*

Proof. I will prove this by focusing on the highest seller type π^h. Assume, to the contrary, that $p^*(\pi^h), f^*(\pi^h)$ is an equilibrium.

1. if, in equilibrium, $p^*(\pi^h) + f^*(\pi^h) > \eta x + (1-\eta)\pi^h L$, no buyer would ever buy from this seller, because there exists no belief that would make buying optimal.

2. if, in equilibrium, $f^*(\pi^h) > \pi^h L$, a buyer would never want to buy insurance (or not buy the product at all), this cannot be part of the proposed equilibrium.

3. if, in equilibrium, $p^*(\pi^h) + f^*(\pi^h) < \eta x + (1-\eta)\pi^h L$ and $f(\pi^h) \leq \pi^l L$, then there exists the following profitable deviation: The seller sets $p' = \eta(x - \pi^h L) - \varepsilon$ and $f' > \pi^h L$. At these prices the buyer cannot buy insurance for any belief because f' surely is too high but has to buy the product since $p' < \eta x + (1-\eta)\pi^B L$ for any possible belief. This is profitable for the seller because he earns $p' = \eta(x - \pi^h L) - \varepsilon$, which for small enough ε will earn a higher profit than $p^*(\pi^h) + f^*(\pi^h) - \pi^h L < \eta(x - \pi^h L)$.

4. if, in equilibrium, $p^*(\pi^h) + f^*(\pi^h) = \eta x + (1-\eta)\pi^h L$ and $f^*(\pi^l) \leq \pi^h L$, and the buyer buys the product and insurance, then this equilibrium must be (partially) separating such that the highest type is the only type setting this price-fee combination. Otherwise, the buyer's belief would be strictly lower than π^h and he would never buy the product. But if the only type setting this price is the highest type, all other types could earn strictly more by mimicking this type. This rules out such equilibria.

These four cases cover all possible combinations of p^* and f^* and the proof is complete. □

Lemma B.3. *There is no equilibrium in which the buyer plays a mixed strategy such that he buys insurance with probability* $0 < \phi < \eta$.

Proof. Assume to the contrary that an equilibrium exists where the buyer mixes with ϕ. In order for the buyer to mix over the insurance decision with probability ϕ, he needs to be indifferent between buying insurance or not. This means that, in equilibrium, his belief when observing (p^*, f^*) needs to be given by $\pi^B = \frac{f^*}{L}$. Then, in order to (weakly) prefer buying the product, it needs to be the case that $p^* \leq \eta(x - f^*)$. Consider the lowest seller type π^l. The most he can earn in this equilibrium is $\eta(x - f^*) + \phi(f^* - \pi^l L)$. But, similar to the case above, consider the deviation $f' = \pi^l L - \varepsilon$ and $p' = \eta(x - \pi^l L)$. For any belief, the buyer has to accept this offer and buy insurance with probability one. To show that this deviation is profitable, note that:

$$\eta(x-\pi^l L)+\pi^l L-\varepsilon-\pi^L > \eta(x-f^*)+\phi(f^*-\pi^l L) \quad \Leftrightarrow \quad (\eta-\phi)(f^*-\pi^l L) > \varepsilon$$

The left hand side of the last expression will be positive whenever f^* is set by types other than π^l (the buyer's beliefs in equilibrium need to be correct) in which case for small enough ε the deviation is profitable. Whenever $f^*(\pi^l) = \pi^l L$, then $p^*(\pi^l) = \eta(x - \pi^l L)$ in equilibrium, the lowest type π^l must be the only one to set this combination. But then, all other sellers can earn strictly more by mimicking this type, which rules out this equilibrium. □

Lemma B.4. *There is no equilibrium in which the buyer plays a mixed strategy such that he buys insurance with probability* $1 > \phi > \eta$.

Proof. As before, it needs to be the case that $\pi^B = \frac{f^*}{L}$ and $p^* \leq \eta(x - f^*)$. Now, focus on the highest type π^h. To show that there can be no equilibrium where the buyer mixes with probability ϕ, observe that the most the highest type could earn in this case would be $\eta(x - f^*) + \phi(f - \pi^h L)$. By exactly the same logic as before, $f' > \pi^h L$ and $p' = \eta(x - \pi^h L) - \varepsilon$ is a profitable deviation for a seller with type π^h, because the buyer will not buy insurance and the seller earns

$$\eta(x - \pi^h L) - \varepsilon > \eta(x - f^*) + \phi(f^* - \pi^h L) \quad \Leftrightarrow \quad (\phi - \eta)(\pi^h L - f^*) > \varepsilon$$

By the same logic as above, the left hand side of the will be positive unless in equilibrium $f^*(\pi^h) = \pi^h L$. But then, there would again have to be separation at the top, allowing all other types to profitably mimic the highest type and benefiting from the high probability with which the buyer would buy in this equilibrium. □

B.2 Additional Results

B.2.1 Pilot Experiment

Table B.1 provides an overview over the key results of the pilot experiments. It is reassuring that these additional data is largely consistent with the data from the main experiment. Given the small sample, I do not conduct a detailed formal analysis here, and instead highlight the noteworthy aspects. First, what can be seen is that acceptance rates differ significantly between treatment and control ($p<0.05$). Buyers in the treatment condition only accept 69.4% of all offers, compared to 81.3% in the control condition. Interestingly, the difference in fees set by the sellers seems more pronounced than in the main experiment. While average prices set by sellers are almost identical in the two conditions, sellers in the treatment condition set fees that are £1.35 higher. A simple t-test on the whole sample indicates the difference as significant, however, this effect disappears once we account for the panel-nature of the data. Hence, the picture that emerges is consistent with the interpretation of buyers preventing exploitation.

	S, C	B, C	S, T	B, T
Initial belief	0.45	0.2963	0.3779	0.40
First-order belief	0.5652	0.5262	0.3411	0.3573
Price	14.83		14.58	
Price (accepted only)	14.41		14.18	
Fee	3.15		4.50	
Fee (accepted only)	2.88		4.42	
Exp. value of offer (true probability)	16.76	8.24	17.34	7.66
Exp. Profit (true probability)	14.08	8.11	13.25	7.16
Exp. Profit (true probability, accepted only)	16.18	8.82	16.89	8.11
Acceptance Frequency	0.813		0.694	

S=Seller, B=Buyer, T=Treatment, C=Control

Table B.1: Pilot Experiment: Summary Table of Group Averages

B.2.2 Additional Tables

Panel A All Offers

	Price (1)	Price (2)	Price (3)	Fee (4)	Fee (5)	Fee (6)
Treatment	-0.674	-0.674	-0.738	0.595	0.595	1.537***
	(0.729)	(0.736)	(0.553)	(0.466)	(0.470)	(0.432)
Constant	14.34***	14.31***	14.71***	3.094***	3.998***	1.453
	(0.602)	(0.689)	(1.353)	(0.348)	(0.483)	(1.478)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

Panel B Accepted Offers Only

	Price (1)	Price (2)	Price (3)	Fee (4)	Fee (5)	Fee (6)
Treatment	-0.413	-0.435	-0.645	0.528	0.538	1.639***
	(0.645)	(0.653)	(0.538)	(0.469)	(0.467)	(0.443)
Constant	13.60***	12.76***	13.33***	3.212***	4.190***	1.224
	(0.443)	(0.441)	(1.366)	(0.309)	(0.555)	(1.670)
Observations	688	688	688	688	688	688
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with the price or fee, respectively, set by the sellers as the dependent variable. Panel A includes all offers, whereas Panel B looks at accepted offers only. The period and individual controls (seller characteristics) are described in table 3.3.

Table B.2: Add-On Price Experiment: Prices and Fees

Panel A All Offers

	Offer (stated belief, seller)			Offer (stated belief, buyer)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.418	-0.418	0.180	0.833	0.833	0.773
	(0.656)	(0.662)	(0.550)	(0.687)	(0.694)	(0.515)
Constant	16.03***	16.47***	14.81***	9.016***	8.640***	8.281***
	(0.470)	(0.501)	(1.237)	(0.607)	(0.667)	(0.836)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

Panel B Accepted Offers Only

	Offer (stated belief, seller)			Offer (stated belief, buyer)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.254	-0.267	0.315	0.861	0.865	0.871**
	(0.615)	(0.624)	(0.561)	(0.540)	(0.552)	(0.386)
Constant	15.40***	15.09***	13.33***	9.859***	10.06***	8.961***
	(0.375)	(0.382)	(0.901)	(0.436)	(0.478)	(0.858)
Observations	688	688	688	688	688	688
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with the expected value of the offer as the dependent variable. Calculating the value of the offer is done use each individual's average stated belief. Panel A includes all offers, whereas Panel B looks at accepted offers only. In each panel, columns 1-3 estimate (3.3) for sellers, whereas columns 4-6 estimate (3.3) for buyers. The period and individual controls are described in table 3.3.

Table B.3: Add-On Price Experiment: Offers (calculated using stated belief)

	Buyer Profits (true probability)			Buyer Profits (stated belief)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.00146	-0.00146	-0.117	0.274	0.274	0.183
	(0.581)	(0.587)	(0.447)	(0.581)	(0.587)	(0.461)
Constant	8.393***	8.065***	8.303***	8.617***	8.399***	8.869***
	(0.462)	(0.512)	(0.938)	(0.500)	(0.592)	(1.024)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with buyer earnings as the dependent variable. Columns 1-3 calculate the expected value of buyer earnings using the true probability, whereas columns 4-6 use each buyer's average stated belief. The period and individual controls are described in table 3.3.

Table B.4: Add-On Price Experiment: Buyer Earnings

	accept	accept	accept $f \leq 1$	accept $f \leq 1$	accept $f \geq 5$	accept $f \geq 5$
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficients						
Treatment	-0.577***	-0.520***	-0.955	-0.358	-1.265***	-1.521***
	(0.211)	(0.137)	(1.042)	(0.260)	(0.398)	(0.330)
Offer	0.388***	0.392***	0.950***	1.094***	0.477***	0.446***
	(0.0719)	(0.0711)	(0.204)	(0.214)	(0.125)	(0.0574)
Riskaversion						
5		-1.949**		-6.736***		-2.708**
		(0.769)		(1.172)		(1.110)
7		-0.635		-2.860***		-0.658
		(0.480)		(0.891)		(0.973)
8		-0.601		-1.622*		-1.056
		(0.541)		(0.949)		(0.813)
9		-1.084**		-4.061***		-0.810
		(0.546)		(0.945)		(1.160)
10		-0.600		-2.119**		-0.810
		(0.620)		(0.839)		(1.038)
Constant	-2.496***	-1.210	-7.284***	-1.914	-2.277***	0.467
	(0.636)	(1.236)	(1.887)	(1.518)	(0.883)	(1.215)
Marginal Effects						
Treatment	-0.126***	-0.111***	-0.0715	-0.0334	-0.245***	-0.277***
	(0.0430)	(0.0290)	(0.0788)	(0.0239)	(0.0480)	(0.0443)
Offer	0.0849***	0.0835***	0.0711***	0.102***	0.0925***	0.0814***
	(0.0111)	(0.00993)	(0.0245)	(0.0106)	(0.0104)	(0.00757)
Riskaversion						
5		-0.435***		-0.578***		-0.487***
		(0.152)		(0.0516)		(0.170)
7		-0.106		-0.169***		-0.0976
		(0.0728)		(0.0437)		(0.137)
8		-0.0992		-0.0647		-0.167
		(0.0842)		(0.0411)		(0.125)
9		-0.206**		-0.301***		-0.123
		(0.0962)		(0.0723)		(0.178)
10		-0.0988		-0.100***		-0.123
		(0.0981)		(0.0362)		(0.152)
Observations	960	960	299	299	245	245
Period Controls	YES	YES	YES	YES	YES	YES
Indiv. Controls	NO	YES	NO	YES	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows the results of the regression in (3.4), similarly to table 3.4, but with the additional controls as specified in table 3.3. The dummy variables for risk aversion are listed separately and the numerical values refer to the number of safe choices made in the elicitation task (see section B.2.3). The categories 5 and 10 include individuals with less than five and more than ten safe choices, respectively. Choosing the safe option in the first six rows corresponds to risk neutrality, which is chosen as the omitted category.

Table B.5: Add-On Price Experiment - Buyer Acceptance Decision

Panel A All Offers

	Price	Price	Price	Fee	Fee	Fee
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.593	-0.593	-0.962***	0.569	0.569	0.546
	(0.378)	(0.382)	(0.294)	(0.390)	(0.394)	(0.475)
Constant	17.79***	16.51***	17.94***	3.481***	4.052***	2.865**
	(0.268)	(0.409)	(1.149)	(0.318)	(0.443)	(1.284)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

Panel B Accepted Offers Only

	Price	Price	Price	Fee	Fee	Fee
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.607	-0.606	-0.938***	0.815**	0.813**	0.987**
	(0.380)	(0.384)	(0.289)	(0.334)	(0.343)	(0.425)
Constant	17.73***	16.31***	17.82***	2.891***	3.507***	2.495
	(0.274)	(0.369)	(1.107)	(0.242)	(0.525)	(1.533)
Observations	885	885	885	423	423	423
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with the price or fee, respectively, set by the sellers as the dependent variable. Panel A includes all offers, whereas Panel B looks at accepted offers only. For f_{it} as the dependent variable in Panel B, this includes only offers where the buyer bought insurance at this fee. The period and individual controls (seller characteristics) are described in table 3.3.

Table B.6: Insurance Experiment: Prices and Fees

Panel A All Offers

	Buyer Earnings (true probability)			Buyer Earnings (stated belief)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.882*	-0.882*	-0.872*	-1.252**	-1.252**	-1.052*
	(0.523)	(0.529)	(0.526)	(0.627)	(0.633)	(0.545)
Constant	14.96***	15.89***	15.67***	14.11***	15.15***	13.94***
	(0.234)	(0.856)	(1.393)	(0.316)	(0.936)	(1.343)
Observations	960	960	960	960	960	960
Indiv. Controls	NO	NO	YES	NO	NO	YES

Panel B Accepted Offers Only

	Buyer Earnings (true probability)			Buyer Earnings (stated belief)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.390	-0.383	-0.356	-0.837*	-0.835*	-0.591**
	(0.373)	(0.377)	(0.328)	(0.458)	(0.460)	(0.238)
Constant	15.96***	17.08***	16.66***	15.04***	16.20***	14.64***
	(0.273)	(0.381)	(0.887)	(0.306)	(0.560)	(0.949)
Observations	885	885	885	885	885	885
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with the expected earnings for buyers as the dependent variable. Panel A includes all offers, whereas Panel B looks at accepted offers only. In each panel, columns 1-3 calculate the expected earnings using the true probability, and columns 4-6 use each individual buyer's average stated belief. The period and individual controls are described in table 3.3.

Table B.7: Insurance Experiment: Buyer Earnings

	Seller Profits (Insurance only, stated belief)			Seller Profits (total, stated belief)		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.587**	0.587**	0.839***	-0.527	-0.527	-0.235
	(0.237)	(0.240)	(0.240)	(0.706)	(0.713)	(0.773)
Constant	0.302***	0.602*	-0.828	16.93***	15.82***	14.73***
	(0.0388)	(0.315)	(0.706)	(0.383)	(0.474)	(1.055)
Observations	960	960	960	960	960	960
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows regression results from estimating equation (3.3) with seller profits as the dependent variable. Using each seller's average stated belief of the product being faulty, columns 1-3 look at the expected seller profits made from selling insurance separately, whereas columns 4-6 analyse potential treatment differences in total expected seller profits. The period and individual controls are described in table 3.3.

Table B.8: Insurance Experiment: Seller Profits (calculated using stated belief)

	Price (control condition)			Price (treatment condition)		
	(1)	(2)	(3)	(4)	(5)	(6)
Insurance	0.137	0.191	0.0630	-0.479**	-0.507**	-0.380**
	(0.320)	(0.320)	(0.354)	(0.192)	(0.201)	(0.171)
Constant	17.68***	15.68***	15.82***	17.37***	16.46***	16.51***
	(0.347)	(0.262)	(2.139)	(0.249)	(0.479)	(1.092)
Observations	450	450	450	435	435	435
Period Controls	NO	YES	YES	NO	YES	YES
Indiv. Controls	NO	NO	YES	NO	NO	YES

*** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered at the matching group level. The table shows the regression results from regressing p on the dummy variable $\boldsymbol{I}_{it}$, which is equal to one if individual i bought insurance in period t, as discussed in section 3.6.3. The sample is constrained to only include accepted offers. Columns 1-3 consider the control condition, and columns 4-6 the treatment condition. The period and individual controls are described in table 3.3.

Table B.9: Insurance Experiment: Buyer Insurance Decision

B.2.3 Individual Characteristics

Risk Aversion. Risk aversion of subjects was elicited as follows. In the table reproduced below, subjects had to indicate for each line whether they prefer option A or option B. This multiple-price list is an adapted version of the one used by Holt and Laury (2002). The main difference is that option A always is a safe option, whereas in their version both options contain risks (in decisions 2-10). The reason for this alteration was that in both experiments, buyers choose between a safe and risky payoff and it thus makes the choices more comparable. For each subject, one line would be randomly selected and implemented (the computer would simulate the lottery if option B was chosen). All payoffs are directly equal to pounds.

	Option A	Option B
1	**2.00** with prob. 100%	**0.50** with prob. 100% and **4.00** with prob. 0%
2	**2.00** with prob. 100%	**0.50** with prob. 90% and **4.00** with prob. 10%
3	**2.00** with prob. 100%	**0.50** with prob. 80% and **4.00** with prob. 20%
4	**2.00** with prob. 100%	**0.50** with prob. 70% and **4.00** with prob. 30%
5	**2.00** with prob. 100%	**0.50** with prob. 60% and **4.00** with prob. 40%
6	**2.00** with prob. 100%	**0.50** with prob. 50% and **4.00** with prob. 50%
7	**2.00** with prob. 100%	**0.50** with prob. 40% and **4.00** with prob. 60%
8	**2.00** with prob. 100%	**0.50** with prob. 30% and **4.00** with prob. 70%
9	**2.00** with prob. 100%	**0.50** with prob. 20% and **4.00** with prob. 80%
10	**2.00** with prob. 100%	**0.50** with prob. 10% and **4.00** with prob. 90%
11	**2.00** with prob. 100%	**0.50** with prob. 0% and **4.00** with prob. 100%

From this data, I construct the a measure of risk aversion as follows: for each subject, I count the number of safe choices (option A). Unless a subject displays a form of inconsistency, there will be a unique switching point that characterises the choice. The more often a subject chooses option A, the more risk averse he is. For subjects who switch back and forth between options A and B, the alternative would be to use the first instance of option B as the measure of risk aversion. Especially for subjects who choose option B in line 2 or 3, followed by a number of choices of option A, the latter option does not seem to provide an appropriate measure. In the regressions, when using dummy variables for the different categories, I re-classify all subjects with less than 6 choices of option A as one category, and all subjects with more than 9 choices of option A, as another, in order to avoid effects driven by one or two outliers.

Social Preferences. I use a version of Bartling et al. (2009) with scaled-down payoffs to measure agents' social preferences on four (coarsely defined) dimensions. Subjects again had to choose between option A and option B from the following table below. For each decision, one decision would be randomly selected, and the choice would be implemented. For each subject, another subject would be chosen as the "OTHER" subject, but in a way that for each subject, the person that he allocates money is different from the person that allocates money to him.

	Option A	Option B
1	**2.00** for YOU and **2.00** for OTHER	**2.00** for YOU and **1.00** for OTHER
2	**2.00** for YOU and **2.00** for OTHER	**3.00** for YOU and **1.00** for OTHER
3	**2.00** for YOU and **2.00** for OTHER	**2.00** for YOU and **3.50** for OTHER
4	**2.00** for YOU and **2.00** for OTHER	**2.50** for YOU and **4.00** for OTHER

Based on the labelling in Bartling et al. (2009), I define the variable *prosocial=1* if the subject chooses option A in line 1. Furthermore, I define *costlyprosocial=1* if the subject chooses option A in line 2, *envy=1* if option A is chosen in line 3 and *costlyenvy=1* if option A is chosen in line 4.

The following tables present a summary of the individual characteristics for the two experiments, separately for buyers and sellers and treatment and control conditions. The variable *CRT* indicates how many items of the three item cognitive reflection test (CRT) a subject answered correctly. The questions are taken directly from Frederick (2005). *female=1* if the subject is female, *mathematics=1* and *economics=1* if the subject had taken an economics or mathematics module during his/her course.

	S, C	B, C	S, T	B, T
Risk aversion (as used in regressions)	7.83	$7.29^{a,b}$	8.33^{a}	8.21^{b}
Risk aversion (without adjusting extreme cases)	7.96	$7.08^{c,d}$	8.46^{c}	8.29^{d}
Envy	0.38	0.38	0.33	0.42
Costly Envy	0.08	0.13	0.21	0.21
Prosocial	1^{e}	0.75^{e}	0.88	0.92
Costly Prosocial	0.13	0.13	0.21	0.08
Female	0.63	0.63	0.71	0.63
Economics	0.25	0.21	0.17	0.33
Mathematics	0.46	0.54	0.29^{f}	0.63^{f}
CRT	1.38	1.38	1	1.21

Letters behind numbers indicate a significant difference ($p<0.05$) between entries that share the same letter. For the risk aversion entries, the (two-sided) p-value is calculated using a Wilocxon Rank Sum test, in all other cases a Fisher exact test is used.

Table B.10: Add-On Price Experiment: Summary of Individual Controls

	S, C	B, C	S, T	B, T
Risk aversion (as used in regressions)	7.92	8.13	8.08	7.42
Risk aversion (without adjusting extreme cases)	7.92	8.29	8.21	7.33
Envy	0.33	0.42	0.29	0.33
Costly Envy	0.17	0.21	0.17	0.29
Prosocial	0.83	0.96	1	0.88
Costly Prosocial	0.25	0.17	0.25	0.29
Female	0.63	0.75	0.71	0.63
Economics	0.13	0.17	0.29	0.13
Mathematics	0.25	0.21	0.29	0.42
CRT	1	0.75	1	0.71

There are no pairwise comparison that are significantly ($p<0.05$) different. For the risk aversion entries, the (two-sided) p-value is calculated using a Wilocxon Rank Sum test, in all other cases a Fisher exact test is used.

Table B.11: Insurance Experiment: Summary of Individual Controls

Appendix C

Sample Experimental Instructions

C.1 Experimental Instructions for Chapter 2

Welcome To The Experiment

In this experiment you will have the opportunity to earn money and to purchase certain goods.

The amount of money you can earn and the goods you obtain depend on your understanding of these instructions, so please read carefully. If you have any questions, please raise your hand and an experimenter will come and answer it privately.

Please do not talk during the experiment or attempt to look at the screens of other participants. Eating, drinking and use of mobile phones is also not permitted.

Anyone violating these rules may be excluded from the experiment. In this case he/she will forfeit any earnings.

The experiment will consist of three parts.

In the first part, you will have to fill in a "personality traits" questionnaire consisting of 48 short yes/no questions. For answering these questions we will pay you 9 pounds in total.

There is no "correct" answer to each question, and your answers do not affect the rest of the experiment. Also, the answers will be kept strictly confidential and anonymous as indeed all data from this experiment.

The 9 pounds are yours once you completed the questionnaire. You can either spend it in the second or third part of the experiment or keep it.

In the second part, you will be offered a good that you can buy. You can already see it in front of you on your table; it is a chocolate bar. If you decide to buy it, it will be yours and you can take it with you at the end of the experiment.

We will provide with you with further details about the procedures and the price of the chocolate bar once we start part two.

Out of the 9 pounds that you earned in part one, 3 pounds will be allocated for this decision. This is your budget for this part. Any money that you decide not to spend will remain yours and will be paid to you at the end of the experiment, but cannot be used in part three.

We will also present you with a list of simple choice problems where for each problem you can choose between a lottery and a fixed outcome. The decisions in the choice problems are only hypothetical, they do not affect your earnings, neither do they influence the prices or the good that you are offered.

The third part will be similar to the buying decision in the second part.

The two goods that you can buy then will be shown to you at the beginning of the third part.

Your budget for the third part will be the remaining 6 pounds from answering the questionnaire. Any money that you decide not to spend will remain yours and will be paid to you at the end of the experiment.

The show-up fee of 5 pounds will be paid on top of any remaining earnings.

If you have any questions, please raise your hand. If not, you may start filling out the questionnaire once it appears on your screen.

Instructions for Part Three

In the following part of the experiment we offer you another two goods for purchase.

In front of you, you can find a pen which features the university's logo. It writes in black ink. Also, you can see a notepad. Feel free to examine both closely.

Again, the price at which each of the two goods is sold will be determined by chance. This time, however, you will be asked to make your choice before the price is drawn.

That means that we will again present you - for each good - with two different prices.

You will also be told how likely it is that you draw each price.

Now, different from before, the actual price of the good will not be revealed to you before the decision. We ask you to make a decision at both prices. Depending on the actual draw, one of your decisions will then be implemented.

The budget that you have for this choice is the remaining 6 pounds that you earned for filling out the questionnaire.

This will be split between the two decisions. That is, you have 3 pounds to spend on the pen and 3 pounds to spend on the notepad. Whatever proportion you decide not to spend, you will keep and take with you.

The determination at which of the two prices the good is actually being sold will be done as follows:

For each good the computer will choose randomly an integer number between 1 and 10.

In each decision, each price is realised if the number randomly chosen by the computer falls into a certain range. For each good, there will be a new draw. As an example, consider the following:

The two possible prices are either 4 pounds or 5 pounds.

If the number randomly chosen by the computer is less or equal than 8, the actual selling price is 4 pounds.

If it is larger than 8, the selling price is 5 pounds.

Note: The numbers given here are for illustration only. The actual values will be different.

If you have any questions, please raise your hand. If not, please follow the instructions on your screen.

C.2 Experimental Instructions for Chapter 3

Welcome To This Experiment

You are about to participate in an experiment in economic decision making. Research organisations have provided the funds for this experiment. Please read these instructions carefully. If you have any questions, please raise your hand and we will answer the questions in private.

During the experiment, you are not allowed to talk to any other participant. You are also not allowed to use any electronic devices, such as laptops or mobile phones. Anybody who violates these rules may be excluded from the experiment and will forfeit all earnings.

Preliminaries

In this experiment you will either be a buyer or a seller of a virtual commodity. Whether you are a buyer or a seller will be determined randomly by the computer at the beginning of the experiment. There are 24 people in the room and 12 of you will be buyers and the other 12 will be sellers.

You will keep the role of buyer or seller for the whole experiment.

Furthermore, there will be three groups, Group A, Group B, and Group C. Each group consists of 4 buyers and 4 sellers each. The experiment is identical for all three groups.

At the beginning of each round, the computer will, for each group separately, form 4 pairs, each consisting of one buyer and one seller. Each possible combination is equally likely. For each round, a new pair will be formed, again such that all combinations are equally likely. More specifically, if you are a buyer, in each round the probability of being paired with any of the 4 sellers from your group is 25 percent, and if you are a seller, in each round the probability of being paired with any of the 4 buyers from your group is 25 percent.

Note that you can only be paired with participants from your group. You will not interact with participants from the other groups.

The basic structure of the experiment:

The experiment consists of 20 identical rounds. Each round consists of 3 stages:

1. Each seller chooses their offer (consisting of a price and a repair fee)
2. Each buyer chooses whether to accept the seller's offer or not
3. The results of the round are summarised

At the end of the experiment, for each participant, one round out of the 20 rounds will be randomly selected for payment. Each of the 20 rounds is equally likely to be selected; you should therefore treat each round as if it were relevant for your earnings. All monetary values in the following are equal to pounds. In each round, the seller offers a virtual commodity to the buyer. The commodity has a value of 0 to the seller, but it has a value of 25 to the buyer, provided it is not faulty. If the buyer accepts the offer of the seller, she/he will obtain the commodity but has to pay the price to the seller. If the buyer obtains a commodity that is faulty, she/he will have to pay an additional repair fee to the seller. If the buyer does not accept the offer, both seller and buyer earn 5. Whether the commodity is faulty or not will be determined via a random process that we will explain below.

A more detailed description of each stage:

The Offer Stage

In this stage, only sellers have to make a decision. If you happen to be in the role of a seller (you will learn your role once we start the experiment by activating the computer program) you will have to choose a price and a repair fee. On the computer screen for this stage, you will see two boxes. Please enter a number (with at most two decimal places) into each box, one for the price, and the other for the repair fee. Together, these two values are the seller's offer to the buyer. The price and the repair fee have to satisfy the following requirements: Both have to be bigger or equal to zero and the sum of price and repair fee cannot be larger than 25, the value of the commodity to the buyer.

The Acceptance Stage In this stage, only buyers have to make a decision. If you happen to be in the role of a buyer, you will see on your screen for this stage the seller's offer. That is, each buyer sees the price and the repair fee that the seller that she/he is paired with chose in the offer stage. The buyer will then have to decide whether she/he accepts this offer or not, by clicking the corresponding button.

How are the payoffs determined, and what role do the price and the repair fee play?

If the buyer rejects the seller's offer: buyer and seller both earn **5**.

If the buyer accepts the seller's offer:

- The buyer obtains the virtual commodity, and the computer determines whether the commodity is faulty or not.
- If the commodity is not faulty, the buyer earns **25** minus the price that the seller offered, and the seller earns the price that she/he chose.
- If the commodity is faulty, the buyer has to pay, **in addition**, the repair fee to the seller. Then, the buyer earns **25** minus the price and minus the repair fee that the seller offered, and the seller earns the price plus the repair fee that she/he chose.

When is the commodity faulty?

The commodity that the seller sells to the buyer has a counter that determines its functionality. Initially, the counter is set to 10000 and the following process determines its final value for the round.

The computer determines, 12 times in a row, whether the counter increases or decreases: Randomly, and with equal probability, the computer either chooses "UP" or "DOWN". If the randomly chosen action is "UP", the counter **increases by 70%** relative to its previous value. If the randomly chosen action is "DOWN", the counter decreases by **60% relative** to its previous value.

For example, after the first draw by the computer, if the randomly chosen action is "UP", the counter is at 17000, and if the randomly chosen action is

“DOWN”, the counter is at 4000. The computer then again draws either “UP” or “DOWN” (again with equal probability) and the counter either increases by 70% or decreases by 60% relative to its previous value.

This is done 12 times in total, and then the final value of the counter determines whether the commodity is faulty or not:

If the final value is less than **1000**, the commodity is faulty and if the final value is higher than **1000**, the commodity is not faulty.

If the commodity is faulty and the buyer accepted the seller’s offer, the buyer pays the price and the repair fee to the seller. If the commodity is not faulty and the buyer accepted the seller’s offer, the buyer pays only the price to the seller.

If the buyer rejected the seller’s offer, whether the commodity is faulty or not does not affect payoffs. In this case, both buyer and seller earn 5.

To clarify, the only thing that matters about the counter is whether its final value is above or below 1000. This determines whether the buyer (if she/he decided to buy) has to pay the repair fee to the seller or not. Apart from this, the counter does not affect payoffs.

In each round and for each pair of a buyer and a seller, the computer determines the functionality anew, always following the same procedure as above. Thus, the final value of the counter can differ from round to round and from seller to seller.

Two additional decisions in each round

In addition to the decisions described above, at the beginning of each round, and also at the beginning of the experiment, we ask you to make two guesses. Both decisions involve guessing a probability and you can earn some money if these guesses are accurate.

For guess 1, we ask you to state how likely you think it is that the commodity is faulty. Remember that this is determined by the random process described above. The probability that the commodity is faulty can be accurately calculated. Please choose a percentage (i.e. a number between 0 and 100, with up to two decimal places) and enter it into the corresponding box on the screen. For guess 2, if you are a seller, we ask you to make a guess about the average number that all buyers in your group chose as the answer to guess 1 in the current round. If you are a buyer, we ask you to make a guess about the average number that all sellers in your group chose as the answer to guess 1 in the current round. Please choose a percentage (i.e. a number between 0 and 100, with up to two decimal places) and enter it into the corresponding box on the screen.

At the end of the experiment, one of your 21 guesses for guess 1 and one of your 21 guesses for guess 2 is selected for payment. In both cases, all 21 guesses are equally likely to be selected. For the two guesses that are selected for payment, if the difference between the number that you stated and the correct number is no more than 5, you will be paid 2.00, otherwise nothing.

For example, if the first number that you stated was 23%, you would earn 2.00 only if the true probability is between 18% and 28%.

Similarly, if the second number that you stated was 79%, you would earn 2.00 only if the average of guess 1 for all buyers (or sellers) from your group is between 74% and 84%.

The Results Stage

In this stage, there are no decisions to be made.

As a buyer, in this stage, you will see a summary of the round, but you will not learn whether the commodity was faulty in this round. Thus, you will not learn whether you (in case you accepted the seller's offer) have to pay the repair fee in addition to the price or not.

As a seller, you will be told whether the buyer accepted your offer and you will learn whether the commodity was faulty in this round. Thus, provided the buyer accepted your offer, you will learn whether you earned only the price, or the price plus the repair fee.

Note that every seller will learn whether the commodity was faulty in this round or not, no matter whether the buyer accepted the offer or not.

Then, everybody proceeds to the next round which is identical to the previous one. As described above, there will be new buyer-seller pairs formed in each round.

Before the first round

At the beginning of the experiment, after every participant has stated his/her answer for guess 1 and guess 2, each seller will be shown 20 simulations of the procedure that determines the functionality of the commodity. That is, each seller will see 20 example performances of the counter. These simulations follow the exact same rules as described above; hence they will be subject to the same randomness. They will be carried out independently for each seller in the room, and for each simulation. If you happen to be in the role of a buyer, please be patient until we start with the first round.

After the last round At the end of the experiment, for each participant, one of the 20 rounds is randomly selected for payment. You will see the all the details of the selected round on the screen, namely, the offer of the seller, whether the buyer accepted the offer or not, whether the commodity was faulty, and your final payoff. You will also see which of your 21 guesses for guess 1 and which of your 21 guesses for guess 2 has been selected for payment and you will be informed about the outcomes. After that, the first part of the experiment is over. There will then be another, shorter, second part. Your decisions in the first part do not affect the second part in any way. We will explain the second part in more detail after part one.

C.3 Experimental Instructions for Chapter 4

Welcome To Our Experiment!

During the experiment, you are not allowed to use mobile phones or communicate with other participants. Please use only the programs and functions intended for this experiment. Please do not talk to other participants. Should you have a question, please raise your hand. We will then come to your desk and answer your question in private. Please do not ask your question loudly. If the question is relevant to all participants, we will repeat the question and answer it. Anybody violating these rules will be excluded from the experiment and the payment.

In addition to the 5 EUR, which we will pay you simply for your participation, you can earn a not inconsiderable amount of money - how much exactly will depend on your effort and also on chance. We will explain the details to you soon.

At the entrance, you just received a RED or a BLACK chip. The color of the chip will soon determine, how much you are able to earn in our experiment. No matter which chip you got, you have the possibility to earn more money by solving tasks. These tasks are identical for each participant and consist of typewriting longer lines of strings, containing letters and numbers, and you need to type them from right to left into the computer interface. We will describe this task more precisely in a moment. The more strings you type correctly from right to left, the more you earn. For one chip color you get 1 Euro for each correctly typed string. For the other color you get 0.10 Euro, or 10 Cents for each correctly typed string.

Chance will determine if RED or BLACK leads to 1 Euro or 10 Cents for each correctly typed string. The instructor has two concealed cards, one red, and one black. One of the two cards will be chosen randomly. If the color of your chip coincides with the color on the card, you will get 1 Euro for each correctly typed string, otherwise 10 Cents.

[FULL INFO: *We will immediately reveal the card, which determines how much you can earn for each string, so that you can see the color.*]

[INFO CHOICE: *You can decide whether you want to be shown the card, which determines how much you can earn for entering a line of strings consisting of letters and numbers, or not. You do not need to look at the card before solving the task. You can choose to look at the card at the end of the experiment if you prefer. As soon as the experiment has started, please enter your decision into the computer by pressing the corresponding button. When all participants have taken their decision, we will pass through the room and privately show the card to those who decided to look at it. The others will get to know the color of the card at the end of the experiment after solving the tasks, directly before the payment.*]

[NO INFO: *We will reveal the card which determines how much you earn for entering a line of strings consisting of letters and numbers at the end of the experiment, after solving the tasks, directly before the payment.*]

No matter what the color of the card: For each string that is typed correctly you will receive additional money. The more tasks you solve, the more money you will earn.

Description of the task: You will now have one hour to type lines of strings consisting of numbers and letters into the computer interface. You need to consider upper and lower cases! You have to type each string in reversed order from right to left (i.e. backwards). Please note: Only a string which is typed completely and correctly will lead to your task being considered solved! If you have an error in your solution, you will not receive money for this task. Then you will continue with the next task. In total, you can solve as many tasks as you like. There is a maximal number of tasks which are programmed in the computer, however, when seriously solving the individual tasks, you will, in all likelihood, not reach this limit. Whenever, during the experiment, you would like to take a break or pass time, you are invited to make use of the reading material (Der Spiegel) which can be found on your desk.

Please note: The amount of money that you will earn in the course of the experiment will be paid to you only at the end of the experiment, when the working time is up.

Bibliography

Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011): "The rich domain of uncertainty: Source functions and their experimental implementation," *The American Economic Review*, 101, 695–723.

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

Akerlof, G. A. (1970): "The market for "lemons": Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, 488–500.

Allcott, H. (2013): "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market," *American Economic Journal: Economic Policy*, 5, 30–66.

Angrisani, M., A. Guarino, S. Huck, and N. C. Larson (2008): "No-trade in the laboratory," *The BE Journal of Theoretical Economics*, 11.

Apesteguia, J. and I. Palacios-Huerta (2010): "Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment," *American Economic Review*, 100, 2548–64.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar (2009): "Large Stakes and Big Mistakes," *The Review of Economic Studies*, 76, 451–469.

Ariely, D., G. Loewenstein, and D. Prelec (2003): ""Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences," *The Quarterly Journal of Economics*, 118, 73–106.

Armstrong, M. (2015): "Search and Ripoff Externalities," *Review of Industrial Organization*, 47, 273–302.

ARMSTRONG, M. AND J. VICKERS (2012): "Consumer protection and contingent charges," *Journal of Economic Literature*, 50, 477–493.

AUSUBEL, L. M. (1991): "The Failure of Competition in the Credit Card Market," *The American Economic Review*, 81, 50–81.

BARTLING, B., E. FEHR, M. A. MARECHAL, AND D. SCHUNK (2009): "Egalitarianism and Competitiveness," *American Economic Review*, 99, 93–98.

BAUMEISTER, R. F. (1984): "Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance," *Journal of Personality and Social Psychology*, 46, 610–620.

BECKER, G. M., M. H. DEGROOT, AND J. MARSCHAK (1964): "Measuring utility by a single-response sequential method," *Behavioral Science*, 9, 226–232.

BELL, D. E. (1985): "Disappointment in Decision Making under Uncertainty," *Operations Research*, 33, 1–27.

BÉNABOU, R. AND J. TIROLE (2002): "Self-Confidence and Personal Motivation," *The Quarterly Journal of Economics*, 117, 871–915.

BOWMAN, D., D. MINEHART, AND M. RABIN (1999): "Loss aversion in a consumption-savings model," *Journal of Economic Behavior & Organization*, 38, 155 – 178.

BROCAS, I., J. D. CARRILLO, S. W. WANG, AND C. F. CAMERER (2014): "Imperfect choice or imperfect attention? Understanding strategic thinking in private information games," *The Review of Economic Studies*, 81, 944–970.

BROCK, J. M., A. LANGE, AND E. Y. OZBAY (2013): "Dictating the Risk: Experimental Evidence on Giving in Risky Environments," *American Economic Review*, 103, 415–37.

BRUNNERMEIER, M. K. AND J. A. PARKER (2005): "Optimal Expectations," *American Economic Review*, 95, 1092–1118.

Camerer, C. and H. Kunreuther (1989): "Experimental markets for insurance," *Journal of Risk and Uncertainty*, 2, 265–299.

Camerer, C., G. Loewenstein, and M. Weber (1989): "The curse of knowledge in economic settings: An experimental analysis," *The Journal of Political Economy*, 1232–1254.

Camerer, C. and M. Weber (1992): "Recent developments in modeling preferences: Uncertainty and ambiguity," *Journal of Risk and Uncertainty*, 5.

Caplin, A. and J. Leahy (2001): "Psychological Expected Utility Theory and Anticipatory Feelings," *The Quarterly Journal of Economics*, 116, 55–79.

——— (2004): "The supply of information by a concerned expert," *The Economic Journal*, 114, 487–505.

Carrillo, J. D. and T. R. Palfrey (2011): "No trade," *Games and Economic Behavior*, 71, 66–87.

Charness, G. and M. Dufwenberg (2006): "Promises and partnership," *Econometrica*, 74, 1579–1601.

Charness, G. and D. Levin (2009): "The origin of the winner's curse: a laboratory study," *American Economic Journal: Microeconomics*, 207–236.

Crawford, V. P. and J. Meng (2011): "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101, 1912–32.

Danz, D. N., K. Madarasz, and S. W. Wang (2015): "Anticipating Information Projection: An Experimental Investigation," mimeo.

De la Rosa, L. E. (2011): "Overconfidence and moral hazard," *Games and Economic Behavior*, 73, 429–451.

Dechenaux, E., D. Kovenock, and R. M. Sheremeta (2015): "A survey of experimental research on contests, all-pay auctions and tournaments," *Experimental Economics*, 18, 609–669.

DellaVigna, S. and U. Malmendier (2004): "Contract design and self-control: Theory and evidence," *The Quarterly Journal of Economics*, 119, 353–402.

——— (2006): "Paying not to go to the gym," *The American Economic Review*, 96, 694–719.

Dohmen, T. J. (2008): "Do professionals choke under pressure?" *Journal of Economic Behavior & Organization*, 65, 636 – 653.

Eil, D. and J. M. Rao (2011): "The good news-bad news effect: asymmetric processing of objective information about yourself," *American Economic Journal: Microeconomics*, 3, 114–138.

Eliaz, K. and A. Schotter (2010): "Paying for confidence: An experimental study of the demand for non-instrumental information," *Games and Economic Behavior*, 70, 304–324.

Eliaz, K. and R. Spiegler (2006): "Contracting with diversely naive agents," *The Review of Economic Studies*, 73, 689–714.

——— (2007): "A Mechanism-Design Approach to Speculative Trade," *Econometrica*, 75, 875–884.

——— (2009): "Bargaining over bets," *Games and Economic Behavior*, 66, 78–97.

Ensthaler, L., O. Nottmeyer, G. Weizsäcker, and C. Zankiewicz (2015): "Hidden skewness: On the difficulty of multiplicative compounding under random shocks," mimeo.

Ericson, K. M. and A. Fuster (2011): "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments," *The Quarterly Journal of Economics*, 126, 1879–1907.

Eysenck, S., H. Eysenck, and P. Barrett (1985): "A revised version of the psychoticism scale," *Personality and Individual Differences*, 6, 21 – 29.

Eyster, E. and M. Rabin (2005): "Cursed equilibrium," *Econometrica*, 73, 1623–1672.

Falk, A. and F. Zimmermann (2014): "Beliefs and Utility: Experimental Evidence on Preferences for Information," Working paper.

Fehr, E. and L. Goette (2007): "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 97, 298–317.

Fehr, E. and K. M. Schmidt (1999): "A Theory Of Fairness, Competition, And Cooperation," *The Quarterly Journal of Economics*, 114, 817–868.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

Fischhoff, B. (1975): "Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty." *Journal of Experimental Psychology: Human perception and performance*, 1, 288–299.

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19, 25–42.

Fudenberg, D. and D. K. Levine (2012): "Fairness, risk preferences and independence: Impossibility theorems," *Journal of Economic Behavior & Organization*, 81, 606–612.

Gabaix, X. and D. Laibson (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121, 505–540.

Ganguly, A. and J. Tasoff (2015): "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future," Working paper.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): "Differentiating ambiguity and ambiguity attitude," *Journal of Economic Theory*, 118, 133 – 173.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of mathematical economics*, 18, 141–153.

Gill, D. and V. Prowse (2012): "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, 102, 469–503.

Golman, R., D. Hagmann, and G. Loewenstein (2016): "Information Avoidance," mimeo.

Greiner, B. (2004): "An online recruitment system for economic experiments," in *Forschung und Wissenschaftliches Rechnen*, ed. by K. Kremer and V. Macho, Gesellschaft für wissenschaftliche Datenverarbeitung, vol. 63, 79–93.

Grubb, M. D. (2009): "Selling to overconfident consumers," *The American Economic Review*, 99, 1770–1807.

Grubb, M. D. and M. Osborne (2015): "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock," *American Economic Review*, 105, 234–71.

Gul, F. (1991): "A theory of disappointment aversion," *Econometrica: Journal of the Econometric Society*, 667–686.

Güth, W., R. Schmittberger, and B. Schwarze (1982): "An experimental analysis of ultimatum bargaining," *Journal of economic behavior & organization*, 3, 367–388.

Halevy, Y. (2007): "Ellsberg Revisited: An Experimental Study," *Econometrica*, 75, 503–536.

Handel, B. R. (2013): "Adverse selection and inertia in health insurance markets: When nudging hurts," *The American Economic Review*, 103, 2643–2682.

Heffetz, O. and J. A. List (2013): "Is the Endowment Effect an Expectations Effect?" *Journal of the European Economic Association, forthcoming.*

Heidhues, P. and B. Kőszegi (2008): "Competition and Price Variation when Consumers Are Loss Averse," *American Economic Review*, 98, 1245–1268.

——— (2014): "Regular prices and sales," *Theoretical Economics*, 9, 217–251.

Heidhues, P. and B. Kőszegi (2010): "Exploiting naivete about self-control in the credit market," *The American Economic Review*, 2279–2303.

Hoffman, E., K. McCabe, K. Shachat, and V. Smith (1994): "Preferences, Property Rights, and Anonymity in Bargaining Games," *Games and Economic Behavior*, 7, 346 – 380.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Holt, C. A. and R. Sherman (1994): "The loser's curse," *The American Economic Review*, 642–652.

Hossain, T. and J. Morgan (2006): "... plus shipping and handling: Revenue (non) equivalence in field experiments on ebay," *Advances in Economic Analysis & Policy*, 6, 1–27.

Isoni, A. (2011): "The willingness-to-accept/willingness-to-pay disparity in repeated markets: loss aversion or "bad-deal" aversion?" *Theory and Decision*, 71, 409–430.

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

Kalayci, K. and J. Potters (2011): "Buyer confusion and market prices," *International Journal of Industrial Organization*, 29, 14 – 22, special Issue: Experiments in Industrial Organization.

Kamenica, E., S. Mullainathan, and R. Thaler (2011): "Helping consumers know themselves," *The American Economic Review*, 101, 417–422.

Karle, H., G. Kirchsteiger, and M. Peitz (2015): "Loss Aversion and Consumption Choice: Theory and Experimental Evidence," *American Economic Journal: Microeconomics*, 7, 101–120.

Kőszegi, B. (2003): "Health anxiety and patient behavior," *Journal of Health Economics*, 22, 1073 – 1084.

Kőszegi, B. and M. Rabin (2004): "A Model of Reference-Dependent Preferences," Department of Economics, UC Berkeley, Working Paper Series.

——— (2006): "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 121, 1133–1165.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99, 909–36.

Knetsch, J. L. (1989): "The endowment effect and evidence of nonreversible indifference curves," *The American Economic Review*, 79, 1277–1284.

Kőszegi, B. (2014): "Behavioral contract theory," *Journal of Economic Literature*, 52, 1075–1118.

Loomes, G. and R. Sugden (1986): "Disappointment and Dynamic Consistency in Choice under Uncertainty," *The Review of Economic Studies*, 53, 271–282.

Mazar, N., B. Kőszegi, and D. Ariely (2014): "True Context-dependent Preferences? The Causes of Market-dependent Valuations," *Journal of Behavioral Decision Making*, 27, 200–208.

Mazumdar, T., S. P. Raj, and I. Sinha (2005): "Reference Price Research: Review and Propositions," *Journal of Marketing*, 69, 84–102.

Michel, C. (2014): "Contractual Structures and Consumer Misperceptions - The Case of Product Warranties," mimeo.

Milgrom, P. and N. Stokey (1982): "Information, trade and common knowledge," *Journal of Economic Theory*, 26, 17 – 27.

Mossin, J. (1968): "Aspects of Rational Insurance Purchasing," *Journal of Political Economy*, 76, 553–568.

Ngangoue, K. and G. Weizsäcker (2015): "Learning from unrealized versus realized prices," mimeo.

Oster, E., I. Shoulson, and E. R. Dorsey (2013): "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease," *American Economic Review*, 103, 804–30.

Plott, C. R. and K. Zeiler (2007): "Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?" *The American Economic Review*, 97, 1449–1466.

Rogers, B. W., T. R. Palfrey, and C. F. Camerer (2009): "Heterogeneous quantal response equilibrium and cognitive hierarchies," *Journal of Economic Theory*, 144, 1440–1467.

Samuelson, W. F. and M. H. Bazerman (1985): "The Winner's Curse in Bilateral Negotiations," in *Research in Experimental Economics*, ed. by V. L. Smith, Greenwich, CT: JAI Press, vol. 3, 105–137.

Santos-Pinto, L. (2008): "Positive Self-image and Incentives in Organisations," *The Economic Journal*, 118, 1315–1332.

Sautmann, A. (2013): "Contracts for Agents with Biased Beliefs: Some Theory and an Experiment," *American Economic Journal: Microeconomics*, 5, 124–56.

Schweizer, N. and N. Szech (2014): "Optimal Revelation of Life-Changing Information," Working paper.

Shen, L., A. Fishbach, and C. K. Hsee (2015): "The Motivating-Uncertainty Effect: Uncertainty Increases Resource Investment in the Process of Reward Pursuit," *Journal of Consumer Research*, 41, 1301–1315.

Sitzia, S. and D. J. Zizzo (2011): "Does product complexity matter for competition in experimental retail markets?" *Theory and Decision*, 70, 65–82.

Sonsino, D., I. Erev, and S. Gilat (2002): "On rationality, learning and zero-sum betting — an experimental study of the no-betting conjecture," mimeo.

Spiegler, R. (2011): *Bounded rationality and industrial organization*, Oxford University Press.

——— (2012): "Monopoly pricing when consumers are antagonized by unexpected price increases: a "cover version" of the Heidhues–Kőszegi–Rabin model," *Economic Theory*, 51, 695–711.

Sprenger, C. (2015): "An endowment effect for risk: Experimental tests of stochastic reference points," *Journal of Political Economy*, 123, 1456–1499.

Stango, V. and J. Zinman (2009): "Exponential growth bias and household finance," *The Journal of Finance*, 64, 2807–2849.

——— (2014): "Limited and varying consumer attention: evidence from shocks to the salience of bank overdraft fees," *Review of Financial Studies*, 27, 990–1030.

Sugden, R., J. Zheng, and D. J. Zizzo (2013): "Not all anchors are created equal," *Journal of Economic Psychology*, 39, 21 – 31.

Thaler, R. (1985): "Mental Accounting and Consumer Choice," *Marketing Science*, 4, pp. 199–214.

Tirole, J. (1982): "On the possibility of speculation under rational expectations," *Econometrica: Journal of the Econometric Society*, 50, 1163–1181.

Tversky, A. and D. Kahneman (1974): "Judgment under Uncertainty: Heuristics and Biases," *Science*, 185, 1124–1131.

——— (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics*, 106, 1039 –1061.

Urbancic, M. B. (2011): "Testing Distributional Dependence in the Becker-DeGroot-Marschak Mechanism," mimeo.

Weaver, R. and S. Frederick (2012): "A Reference Price Theory of the Endowment Effect," *Journal of Marketing Research*, 696–707.

Wenner, L. M. (2011): "An Experimental Test for Expectation-Based Reference Points in Consumer Behaviour," MRes Dissertation, University College London.

——— (2015): "Expected prices as reference points - Theory and experiments," *European Economic Review*, 75, 60 – 79.