Essays in Industrial Organization and Behavioral Economics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Fabian Herweg
aus Köln

Bonn 2009

Dekan:	Prof. Dr. Christian Hillgruber
Erstreferent:	Prof. Paul Heidhues, Ph.D.
Zweitreferent:	Prof. Dr. Urs Schweizer

Tag der mündlichen Prüfung: 04.11.2009

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn (http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

för de Mamm, de Papp un de Groß....

Acknowledgements

In preparing this thesis I received support from many people to whom I am grateful. First and foremost, I want to thank Paul Heidhues who has been a great supervisor. He provided very detailed and constructive comments to all four chapters of my thesis. Secondly, I want to thank Urs Schweizer for acting as a referee on my thesis committee and for many helpful comments.

Moreover, I am particularly grateful to Daniel Müller. He has not only been a good friend, our joint research has been a fruitful, usually fun, and inspiring activity. Chapters I and II of this thesis are achievements of this cooperation. Besides our joint projects, the discussions with and the comments made by Daniel also improved the quality of the remaining chapters.

The countless discussions with Philipp Weinschenk, who also contributed to Chapter I, about the correct intuitions for our/my findings significantly improved the illustration of the main results of this thesis.

I am grateful to Jörg Budde who taught me a lot about the theory of incentives. The insights obtained in this interaction facilitated me exploring the economics of contract design under moral hazard, which is the topic of Chapter I.

Chapter III benefited a lot from comments and suggestions made by Heidrun C. Hoppe. She has been extremely supportive before moving to Hanover.

The Bonn Graduate School of Economics (BGSE) provided not only financial support but also, and more importantly, a stimulating environment with many young researchers. I would like to thank all people who have contributed to this organization, especially Urs Schweizer. Moreover, I am indebted to Benny Moldovanu for motivating me to apply for the BGSE.

The BGSE topics courses improved my understanding of several fields of economics and provided useful hints for my own research. Therefore, I am grateful to the lecturers of the courses I attended: Georg Nöldeke, Benny Moldovanu, Paul Heidhues, Jörg Budde, Anja Schöttner, and Matthias Kräkel.

I have benefited from discussions with and comments made by my office mates: Konrad Mierendorff, J. Marcelo Cadena, and Sebastian Kranz. Moreover, I owe thanks to Botond Kőszegi, Daniel Krämer, Christina Matzke, Nora Szech, Thomas Rieck, Andreas Roider, Philipp Wichardt, and many other people.

Finally, I owe much to Kerstin, my family and friends, without whom I would not have gone so far, and who put up with all the negative sides of the graduate student's life that I levied on them.

Contents

List of Figures

Introduction

The analysis of contractual arrangements between two or more parties has a long tradition in economics, since the "game-theoretic" revolution. Questions among this line are: In which way should the wage of an employee depend on his own or the firm's performance? [1] What is the optimal selling contract for a given good or service? [2] For instance, a telecommunication service provide may ask how should the invoice of a customer vary with his consumption patterns within a month? This dissertation aims at contributing to the strands of theoretical literature in economics investigating the above mentioned questions.

Next to the orthodox theories on contract design and optimal selling strategies, there is a recent and growing literature investigating how rational firms respond to consumer or employee biases. During the last decade the field of economics that modifies the "standard preferences" by incorporating findings from psychological research has made a significant progress. Now, there exist reasonable and tractable frameworks for modeling agents with reference-dependent preferences, time-inconsistent discounting, fairness concerns and many other behavioral motives. These "workhorse" models not only allow to investigate the "old" questions for a richer class of the agent's preferences, they also raise new questions. For instance, if an employee's behavior is time inconsistent, this raises the question how often should there be meetings between supervisors and subordinates where the subordinates report on the progress they made on given projects. An interim deadline, i.e. a meeting with the supervisor, would impose an undesirable restriction on the choice set of a time-consistent employee. If the employee has time-inconsistent preferences, however, such an interim deadline may help to increase the employee's work input.

In three of the four main chapters of this dissertation, I follow this new "behavioral" approach and investigate contractual situations by allowing for biased consumers or employees to provide possible explanations for observed contractual arrangements that are puzzling from the perspective of standard orthodox economics.

[1] Classic contributions who investigate this question are Holmström (1979, 1982), Shavell (1979), Gjesdal (1982), and Grossman and Hart (1983).

[2] One of the first papers that formally analyzes more sophisticated pricing schemes is Oi (1972), who derives the optimal two-part tariff for a monopolistic seller. A fully nonlinear tariff schedule is derived by Mussa and Rosen (1978).

Chapter I, which is based on a joint project with Daniel Müller and Philipp Weinschenk, analyzes the well-known principal-agent problem with hidden actions. In these models the principal (firm) seeks to contract with an agent (potential employee). If the agent accepts the contract, he can undertake an action (provide effort) which is costly to him. The chosen action is unobservable for the principal but stochastically influences her profit. The principal can make the compensation payment (wage) of the agent depending on a performance measure which is imperfectly correlated with the agent's action choice (for instance the firm's profit). By making the compensation payment depending on performance, the principal can motivate the agent to provide a desired level of effort. By doing so, however, the principal imposes some risk on the agent since the performance measure is only a noisy signal of the agent's effort choice. Since human beings typically dislike to be confronted with risky situations, the principal faces a tradeoff between providing incentives and optimal risk sharing. With the tradeoff between incentive provision and risk sharing being at the heart of moral hazard, allowing for a richer description of the agent's risk preferences seems a natural starting point to obtain deeper insights into contracting.

Orthodox economic theory models risk aversion by assuming a strictly concave utility function. This approach, however, can only inadequately describe the risk preferences of human beings, see for instance Rabin (2000). In their seminal contribution, Kahneman and Tversky (1979) expose the prospect theory which is an alternative to describe risk preferences based on psychological findings. A main module of this theory is the concept of loss aversion. This concept suggests that a person's preferences are determined by changes in outcomes relative to his reference level, and not merely by absolute levels of outcomes. Moreover, this concept posits that losses loom larger than gains of equal size. In particular, prospect theory assumes loss aversion even for small stakes. A "standard" agent is locally risk neutral, i.e., his degree of risk aversion is only of second order. The drawback of the original prospect theory is that it does not provide a satisfying approach how to determine a decision maker's reference point. In two recent articles, Kőszegi and Rabin (2006, 2007) provide a loss-aversion theory where the reference point is fully determined by the decision maker's rational expectations about outcomes. We consider a principal-agent model with moral hazard, framed as an employer-employee relationship, which is completely standard but for one twist: the agent is assumed to be loss averse in the sense of Kőszegi and Rabin. The main finding is that a simple (lump-sum) bonus contract is optimal when loss aversion is the predominant determinant of the agent's risk preferences. If the employee's performance is below a certain threshold he receives only his base wage, whereas for performances above this threshold he receives the base wage plus a fixed performance independent bonus payment. This is in stark contrast to the finding for a “purely” risk averse agent who is not loss averse and thus exhibits local risk neutrality, which implies that paying slightly different wages for different signals improves incentives at negligible costs (see Holmström, 1979). Thus, standard theory predicts fairly complex performance-based wage schemes which are rarely observed in practice. Simple bonus contracts, however, are commonly utilized in practice.[3]

Chapter II, which is based on a joint paper with Daniel Müller, investigates the behavior and in particular the performance of a decision maker who faces self-control problems. What are self-control problems? Self-control problems in our sense arise from time-inconsistent preferences. A person’s preferences are time consistent if he feels the same about a given intertemporal tradeoff no matter when he is asked. A person is time consistent if and only if he discounts streams of utility over time exponentially, which is traditionally assumed in economics since this concept was introduced by Samuelson (1937). The concept of exponential discounting was originally proposed merely for purposes of formal tractability since it compresses all behavioral concerns into one single parameter. Psychological researches as well as casual observations suggest that the assumption of exponential discounting and thus time-consistent preferences is problematically.[4] As is familiar probably to everyone, the short term tastes for immediate gratification (eating a chocolate bar) are inconsistent with our long-run preferences (subsisting healthier).

To capture these preferences one needs a discount function with declining discount rates, such preferences are called “hyperbolic discounting”. The first formal model of such time-variant preferences is Strotz (1955). Now, established in economics is a tractable two-parameter model that slightly modifies Samuelson’s discounting model. This model, which was originally introduced by Phelps and Pollak (1968) to study intergenerational altruism, is often referred to as quasi-hyperbolic discounting and captures the essence of hyperbolic discounting. Seminal contributions who apply this concept to model decision makers with self-control problems are Laibson (1997, 1998) and O’Donoghue and Rabin (1999b). We also apply this concept of quasi-hyperbolic discounting to analyze the performance of agents with self-control problems, and in particular we investigate how interim deadlines can help to improve performance.

Former research has shown that procrastination can be explained by quasi-hyperbolic discounting, see for instance O’Donoghue and Rabin (1999b). We develop a model of continuous effort choice over time that shifts the focus from completion of to per-

[3]Bonus contracts are widely used for instance in sales force compensation plans, see Joseph and Kalwani (1998).

[4]For a review of the psychological literature on time-inconsistent discounting see Frederick et al. (2002).

formance on a single task. Besides showing that procrastination induced by time-inconsistent preferences in general hampers performance, we mainly ask two questions: First, does sophistication about the own self-control problems increase performance and overall well-being? Secondly, do interim deadlines enhance performance? In contrast to the existing literature on procrastination, we find that being aware of the own self-control problems can reduce a person's performance as well as his overall well-being. With regard to the second question, we show that being exposed to an interim deadline increases the performance as well as the overall well-being of a hyperbolic discounter, irrespectively of his awareness of own self-control problems. These findings suggest that there is also scope for the employer of such an agent to benefit from imposing interim deadlines. Therefore, Chapter II provides a theoretical underpinning for the frequent observation of interim deadlines in working environments.

In Chapter III, which presents a non-behavioral model, I analyze the effect of a larger contract space on firms' decisions to produce differentiated products. Since Bertrand's (1883) article, it is well-known that if firms produce homogeneous products price competition leads to prices equal to marginal costs. Thus, if firms produce with the same constant returns to scale technology, in the Bertrand equilibrium all firms make a zero profit. This finding, which is also known as the Bertrand-Paradox, does not seem to fit to the vast majority of industries. Since typically firms engage in price and not in quantity competition à la Cournot (1838) many economists augmented the Bertrand model to allow for product differentiation.[5]

If firms produce differentiated products, price competition is softened which in turn increases industry profits. The theories of product differentiation can roughly be subdivided into two categories: horizontal and vertical differentiation. For horizontally differentiated products, consumers have different ordinal rankings over the products. Thus, if there are two products, say A and B, which are priced equally, then some consumers purchase product A while others purchase product B.[6] If products are vertically differentiated, all consumers have the same ordinal ranking over these products. The best example for a vertical product feature is quality. If all products have the same price, each consumer purchases the good with the highest quality.

In a classic contribution on vertical differentiation, Gabszewicz and Thisse (1979) show that price competition is relaxed if firms produce different qualities. The qualities the firms produce are fixed exogenously in their model. Shaked and Sutton (1982) extend the Gabszewicz-Thisse model by endogenizing quality levels. The main result

[5]Kreps and Scheinkman (1983) show that if firms first choose capacities and than compete in prices, under mild assumptions about demand, the unique equilibrium outcome is the Cournot outcome.

[6]The linear city model of Hotelling (1929) is probably the most frequently applied model for horizontally differentiated products in industrial organization.

is that in equilibrium of a two-stage game, where firms first choose quality levels and afterwards compete in prices, firms produce distinct qualities and thereby relax price competition. The fact that firms can relax competition via quality differentiation is possible due to heterogeneity of consumers with regard to their tastes for quality. All consumers prefer a higher quality, however, the willingness to pay for an improved quality differs across consumers. Thus, by selecting different qualities at the first stage the firms attract different types of consumers at the second stage which in turn relaxes competition.

In contrast to these earlier findings, Chapter III shows that firms may have incentives for quality differentiation even when consumers do not differ in their tastes for quality but differ in their preferences for quantity. It is shown that quality differentiation can relax competition when firms can offer two-part tariffs. If firms are restricted to linear pricing, however, quality differentiation does not relax competition. The intuition is that selecting different qualities may facilitate market segmentation in which different classes of consumers are also served different quantities. Without quality differentiation the firms produce perfect substitutes and thus compete in a Bertrand fashion. Thus, the novel contribution of this chapter to the theory of product differentiation is to identify differences in consumers' preferences for quantity as a reason for strategic quality differentiation by firms.

Chapter IV investigates the widespread use of flat-rate tariffs. With a flat-rate tariff a consumer pays a fixed amount which is independent of his usage, the basic fee, to obtain unlimited access to a good or service. Nowadays, flat-rate contracts are common in many industries, e.g., telephone services, Internet access, car rental, car leasing, DVD rental, amusement parks, health clubs, and so on. Besides these examples that come immediately in one's mind, all-you-can-eat buffets, flat-rate parties, and a season ticket for a sports team are examples for flat-rate contracts.

The fact that flat-rate tariffs are such favorable pricing schemes is hard to reconcile with orthodox economic theory, in particular for industries where marginal costs are non-negligible. If marginal costs are positive, a marginal payment of zero leads to an inefficiently high level of consumption which hardly can be optimal. Usage-based pricing, however, may cause positive transaction costs for measuring the actual usage of a consumer. In many of the examples provided above marginal costs of production or service provision are positive but transaction costs for measuring usage are close to zero. For instance, the price for a rental car typically is fixed per day and does not depend on the mileage. The costs for the rental car company are clearly higher if the car is used more heavily, due to for instance a higher wear of the tires. To ascertain how many miles a customer drove with the car is relatively easy and not very costly for the company. Is there any other reason, maybe outside standard economics, that can explain the prevalence of flat-rate contracts?

There is evidence that people like the idea of paying one fee for unlimited access.[7] A common phenomenon among those consumers who prefer unlimited access is the tendency to pay a fixed fee that costs more than measured service would have cost, given their demonstrated demand. Put differently, consumers facing the choice between several tariffs often do not select the optimal one given their consumption patterns. In particular, consumers often prefer a flat-rate tariff even though they would save money with a measured tariff. Train (1991) referred to this phenomenon as “flat-rate bias”. Evidence for tariff-choice biases and in particular for the flat-rate bias was first documented for U.S. households among telephone service options. Notable works on this topic are Train et al. (1987), Hobson and Spady (1988), and Train et al. (1989). Already these papers point out, that consumers are uncertain about their demand when selecting a tariff and that they choose tariffs on the basis of the “insurance” provided by these contracts. Thus, the flat-rate bias is a consequence of customers’ risk preferences in the face of uncertain consumption patterns. Since variations in the monthly billing rates are small compared to a consumer’s income, standard risk aversion cannot capture this insurance motive.

Therefore, to capture first-order risk aversion I posit that consumers are loss averse. A loss averse consumer dislikes even small deviations from his reference point. In my model, the consumer’s demand is uncertain when he selects a tariff. I assume that a consumer forms rational expectations about his invoice, which determine his reference point. The consumer feels a loss if his actual invoice amount is above his reference point, and he feels a gain if it is below his reference point. I follow Kőszegi and Rabin (2006, 2007) in assuming that the reference point is a full distribution of the possible billing rates. I show that a consumer with these preferences is biased in favor of flat-rate tariffs, since flat rates insure against the risk of losses in periods of greater than average usage.

Because observed tariffs are the result of strategic interactions of firms and consumers, I set an oligopoly model up where firms compete for loss averse consumers. Moreover, I allow for consumer heterogeneity with respect to their degree of loss aversion. The timing is as follows: (i) Firms offer a menu of two-part tariffs to consumers; (ii) Each consumer selects either one tariff or none; (iii) The consumer observes the realization of the state of the world which determines his preferences for the good and chooses a demanded quantity. I analyze the symmetric information case in which firms observe a consumer’s degree of loss aversion, as well as the asymmetric information

[7] Cf. Lambrecht and Skiera 2006.

case in which the degree of loss aversion is private information. In the benchmark case of symmetric information, in equilibrium firms offer a flat-rate tariff to those consumers whose degree of loss aversion compared to marginal costs exceeds a certain threshold. Consumers with a lower degree of loss aversion are assigned to a metered tariff, i.e., a two-part tariff with a strictly positive unit price. These findings turn out to carry over likewise to the asymmetric information case.

Before moving to the main part of this dissertation, a few words to the use of the first person plural in the following chapters may be in order. Chapter I and Chapter II have been developed in collaboration with other PhD students. The use of the plural in Chapter IV is due to the advice by Thomson (1999) to use this form even for single-authored articles in economics. Unfortunately, while I was writing my first paper, which is contained in Chapter III, I was unaware of Thomson's guidelines.

The next four chapters are each devised as independent self-contained units. Moreover, for notational convenience, the number of the chapter is often suppressed when referring to a proposition, lemma, or corollary. For instance, if it is referred to Proposition 2 in Chapter III, then it is meant Proposition III.2.

I. The Optimality of Simple Contracts: Moral Hazard and Loss Aversion

This chapter extends the standard principal-agent model with moral hazard to allow for agents having reference-dependent preferences according to Kőszegi and Rabin (2006, 2007). When loss aversion is the predominant determinant of the agent's risk preferences, the principal optimally offers a simple bonus contract, i.e., when the agent's performance exceeds a certain threshold, he receives a fixed bonus payment. Also when risk aversion becomes more important, the optimal contract displays less complexity than predicted by orthodox theory. Thus, loss aversion introduces an endogenous complexity cost into contracting.

1. Introduction

The recent literature provides very strong evidence that contractual forms have large effects on behavior. As the notion that "incentive matters" is one of the central tenets of economists of every persuasion, this should be comforting to the community. On the other hand, it raises an old puzzle: if contractual form matters so much, why do we observe such a prevalence of fairly simple contracts?

— *Bernard Salanié (2003)*

A lump-sum bonus contract, with the bonus being a payment for achieving a certain level of performance, is probably the most simple incentive scheme for employees one can think of. According to Steenburgh (2008), salesforce compensation plans provide incentives mainly via a lump-sum bonus for meeting or exceeding the annual sales quota.[1] Simple contracts are commonly found not only in labor contexts but also in insurance markets. A prevalent form of insurance contracts is a straight-deductible contract, widely used, for example, in automobile insurance. The observed plainness of contractual arrangements, however, is at odds with predictions made by economic theory, as nicely stated in the above quote by Salanié. While Prendergast (1999) already referred to the discrepancy between theoretically predicted and actually observed contractual form, over time this question was raised again and again, recently by Lazear and Oyer (2007), and the answer still is not fully understood.[2]

Beside this gap between theoretical prediction and observed practice, both theoretical as well as empirical studies demonstrate that these simple contractual arrangements create incentives for misbehavior of the agent that is outside the scope of most standard models. As Oyer (1998) points out, facing an annual sales quota provides incentives for salespeople to manipulate prices and timing of business to maximize their own income rather than firms' profits. For insurance markets, Dionne and Gagné (2001) show that "deductible contracts can introduce perverse effects when falsification behavior is

[1] Incentives for salespeople in the food manufacturing industry often are solely created by a lump-sum bonus, see Oyer (2000). Moreover, in his book about designing effective sales compensation plans, Moynahan (1980) argues that for a wide range of industries lump-sum bonus contracts are optimal. For a survey on salesforce compensation plans see Joseph and Kalwani (1998).

[2] For evidence on deductibles in the automobile insurance, see Puelz and Snow (1994) or Chiappori et al. (2006). As was shown by Rothschild and Stiglitz (1976), the use of deductibles can theoretically be explained if the insurance market is subject to adverse selection. Besides adverse selection, however, moral hazard plays an important role in automobile insurance. Deductibles were found to be optimal under moral hazard by Holmström (1979) if the insured person's action influences only the probability of an accident but not its severity. As pointed out by Winter (2000), however, "[d]riving a car more slowly and carefully reduces both the probability of an accident and the likely costs of an accident should it occur." Thus, existing theories cannot explain the prevalence of deductibles in these markets.

potentially present".[3] These observations raise "the interesting question of why these [...] contracts are so prevalent. [...] It appears that there must be some benefit of these contracts that outweighs these apparent costs" (Lazear and Oyer, 2007).

To give one possible explanation for the widespread use of these contractual arrangements, we consider a principal-agent model with moral hazard, framed as an employer-employee relationship, which is completely standard but for one twist: the agent is assumed to be loss averse in the sense of Kőszegi and Rabin (2006, 2007). With the tradeoff between incentive provision and risk sharing being at the heart of moral hazard, allowing for a richer description of the agent's risk preferences that goes beyond standard risk aversion seems a natural starting point to gain deeper insights into contract design. Our main finding is that a simple (lump-sum) bonus scheme is optimal when loss aversion is the driving force of the agent's risk preferences.[4] This is in stark contrast to the findings for a standard risk-averse agent. An agent who is risk but not loss averse exhibits local risk neutrality, which implies that paying slightly different wages for different signals improves incentives at negligible cost. A loss-averse agent, on the other hand, is first-order risk averse. Since losses loom larger than equally-sized gains, in expectations the agent suffers from deviations from his reference point. With the reference point being multidimensional under the concept of Kőszegi and Rabin, the agent is first-order risk averse at all possible wage levels. In consequence, paying even slightly different wages reduces the agent's expected utility, for which in turn he demands to be compensated. Thus, by offering a simple contract, that specifies only few different wage levels, the principal can lower the expected payment necessary to compensate the agent for the induced losses.

We present our model of a principal-agency that is subject to moral hazard in Section 2. The principal, who is both risk and loss neutral, does not observe the agent's effort directly. Instead, she observes a measure of performance that is correlated with the agent's effort decision. Following Kőszegi and Rabin, we posit that a decision maker – next to intrinsic consumption utility from an outcome – also derives gain-loss utility from comparing the actual outcome with his rational expectations about outcomes. More precisely, the sensation of gains and losses is derived by comparing a given outcome to all possible outcomes. To illustrate this point, consider an employee who receives a wage of \$5000 for good performance, a wage of \$4400 for mediocre performance, and a wage of \$4000 for bad performance. If the employee's performance

[3] For evidence on fraudulent claims being a major problem in the car insurance market see Caron and Dionne (1997), who estimated the cost of fraud in the Québec automobile insurance market in 1994 at \$100 million, just under 10% of total claims. For an estimation of the costs of fraudulent claims in the United States, see Foppert (1994).

[4] In the following, we will use the terms bonus contract and bonus scheme interchangeably to refer to a contract that specifies exactly two distinct wage payments, a base wage and a lump-sum bonus.

is mediocre, this generates mixed feelings, a loss of \$600 and a gain of \$400.[5] The key feature of the Kőszegi-Rabin model is that expectations matter in determining the reference point.[6] While mainly based on findings in the psychological literature,[7] evidence for this assumption is provided also by two recent contributions to the economic literature. In a real-effort experiment, Abeler et al. (2009) find strong evidence for individuals taking their expectations as a reference point, rather than the status quo.[8] Similarly, analyzing decision making in a large-stake game show, Post et al. (2008) come to the conclusion that observed behavior "is consistent with the idea that the reference point is based on expectations." Regarding applications, the Kőszegi-Rabin concept is used by Heidhues and Kőszegi (2005, 2008a) to introduce consumer loss aversion into otherwise standard models of industrial organization. While the former paper explains why monopoly prices react less sensitive to cost shocks than predicted by orthodox theory, the latter provides an answer to the question why non-identical competitors charge identical prices for differentiated products.

As a benchmark, in Section 3 we reconsider the case of a purely risk-averse agent: Under the optimal contract, signals that are more indicative of higher effort are rewarded strictly higher, thereby giving rise to a strictly increasing wage profile. We then turn to the analysis of a purely loss-averse agent who does not exhibit risk aversion in the usual sense. After providing sufficient conditions for the first-order approach to be valid, we establish our main result: when the agent is loss averse, it is optimal to offer a bonus contract. No matter how rich the set of possible realizations of the performance measure, the optimal contract comprises of only two different wage payments. We already touched on the intuition underlying this finding: With the agent's action being unobservable, the necessity to create incentives makes it impossible for the principal to bear the complete risk. With losses looming larger than equally sized gains, this ex ante imposes an expected net loss on the agent, which equals the sum over the ex ante expected wage differences weighted by the product of the corresponding probabilities. To illustrate, let us return to the example introduced above. Suppose the agent expects to perform well, moderately, or poorly with probability p_G, p_M and p_B, respectively. Then, ex ante, the agent expects a wage difference – or net loss – of \$600 with probability $p_M p_G$, a net loss of \$400 with probability $p_B p_M$, and a net loss of \$1000 with probability $p_B p_G$. The agent demands to be compensated for his

[5]For at least suggestive evidence on mixed feelings, see Larsen et al. (2004).

[6]The feature that the reference point is determined by the decision maker's forward-looking expectations is shared with the disappointment aversion models of Bell (1985), Loomes and Sugden (1986), and Gul (1991).

[7]For instance, Mellers et al. (1999) and Breiter et al. (2001) document that both the actual outcome and unattained possible outcomes affect subjects' satisfaction with their payoff.

[8]The status quo was most often assumed as reference point in the wake of Kahneman and Tversky's (1979) original formulation of prospect theory.

overall expected net loss, which the principal therefore seeks to minimize. Consider, for the sake of argument, a principal who wants to strengthen incentives to provide effort, starting out from a not fully differentiated wage scheme. There are two ways to do so. First, the principal can introduce a new wage spread, i.e., pay slightly different wages for two signals that were rewarded equally in the original wage scheme, while keeping the differences between all other neighboring wages constant. Secondly, the principal can increase an existing wage spread, holding constant all other spreads between neighboring wages. Both procedures increase the overall expected net loss by increasing the size of some of the expected losses without reducing others. Introducing a new wage spread, however, additionally increases the overall expected net loss by increasing the ex ante expected probability of experiencing a loss. Therefore, in order to improve incentives, it is advantageous to increase an existing wage spread without adding to the contractual complexity in the sense of increasing the number of different wages. In this sense, reference-dependent preferences according to Kőszegi and Rabin introduce an endogenous complexity cost into contracting based on psychological foundations.

Thereafter, we establish several properties displayed by the optimal contract. Let a signal that is the more likely to be observed the higher the agent's effort be referred to as a *good* signal. We find that the subset of signals that are rewarded with the high wage contains either only good signals, or all good signals and possibly a few bad signals as well.[9] When abstracting from integer-programming problems, it is optimal for the principal to order the signals according to their relative informativeness (likelihood ratio), i.e., the agent receives the high wage for all signals that are more indicative of high effort than a cutoff signal. Last, we show that an increase in the agent's degree of loss aversion may allow the principal to use a lower-powered incentive scheme in order to implement a desired level of effort. The reason is that a higher degree of loss aversion may be associated with a stronger incentive for the agent to choose a high effort in order to reduce the probability of incuring a loss. The overall cost of implementation, however, increases in the agent's degree of loss aversion.

In the last part of Section 3, we analyze the general case in which the agent is both risk and loss averse. It is shown that our results are robust towards a small degree of risk aversion. Moreover, we give a heuristic reasoning why a reduction in the complexity of the contract is also to be expected to be optimal for a non-negligible degree of risk aversion, and confirm our conjecture by means of a numerical example.[10]

[9] The theoretical prediction that inferior performance may also well be rewarded with a bonus is in line with both Joseph and Kalwani (1998)'s suggestion that organizations tend to view the payment of a bonus as a reward for good or even acceptable performance rather than an award for exceptional performance, and Churchill et al. (1993)'s prescription that bonuses should be based on objectives that can be achieved with reasonable rather than Herculean efforts.

[10] This finding also relates to the observation that, within a firm, pay for individuals often seems to be less variable than productivity, as recently surveyed by Lazear and Shaw (2007). Our model

Returning to the case of a purely loss-averse agent, in Section 4 we relax the assumptions that guaranteed validity of the first-order approach. Here, to keep the analysis tractable, we focus on binary measures of performances. If the agent's degree of loss aversion is sufficiently high and if the performance measure is sufficiently informative, then only extreme actions – work as hard as possible or do not work at all – are incentive compatible. Put differently, the principal may face severe problems in fine-tuning the agent's incentives. These implementation problems, however, can be remedied if the principal can commit herself to stochastically ignoring the low realization of the performance measure, i.e., by turning a blind eye from time to time. Besides alleviating implementation problems, turning a blind eye may also lower the cost of implementing a certain action. Thus, the sufficiency part of Blackwell's theorem does not hold when the agent has reference-dependent preferences.

After briefly summarizing our main findings, Section 5 concludes by discussing robustness of our results with respect to imposed assumptions. All proofs are given in the appendix.

Related Literature Before presenting our model, we relate our paper to the small but steadily growing literature that analyzes the implications of loss aversion on incentive design.[11] With reference-dependent preferences being at the heart of loss aversion on the one hand, but with no unifying approach provided how to determine a decision maker's reference point on the other hand, it is little surprising that all contributions differ in this particular aspect. While Dittmann et al. (2007) posit that the reference income is exogenously given by the previous year's fixed wage, Iantchev (2005), who considers a market environment with multiple principals competing for the services of multiple agents, applies the concept of Rayo and Becker (2007). Here, an agent's reference point is endogenously determined by the equilibrium conditions in the market. When focusing on a particular principal-agent pair, however, both the principal and the agent take the reference point as exogenously given. An exogenous reference point does not always seem plausible. Starting out from the premise that the reference point is forward looking and depends on the distributions of outcomes, as suggested

suggests an alternative explanation for this pay compression outside the realms of inequity aversion, tournament theory, and influence activities.

[11] Beside loss aversion there are other behavioral biases that are incorporated into contracting problems with moral hazard. Non-standard risk preferences in a moral hazard framework are analyzed by Schmidt (1999), who applies Yaari's (1987) concept of dual expected utility theory. Englmaier and Wambach (2006) characterize the optimal contract for the case of an inequity-averse agent in the sense of Fehr and Schmidt (1999). A multi-agent contracting problem in which agents care about their own status is investigated by Besley and Ghatak (2008) in a static context, and by Auriol and Renault (2008) in a dynamic setting. By introducing worker overconfidence into a multi-agent moral-hazard problem, Fang and Moscarini (2005) show that it can be optimal not to screen workers according to their skills. For a review of behavioral economics of organizations see Camerer and Malmendier (2007).

by ample evidence, De Meza and Webb (2007) consider both exogenous as well as endogenous formulations of the reference point. Concluding that the disappointment concept of Gul (1991), which equates the reference point with the certainty equivalent of the income distribution, does yield some questionable implications,[12] De Meza and Webb propose that the reference income is the median income, which captures the idea that the agent incurs a loss at all incomes for which it is odds-on that a higher income would be drawn. Taking median income as reference income, however, suffers from the drawback that it is discontinuous in the underlying probability distribution.[13]

All of the aforementioned contributions explore questions of both empirical importance as well as theoretical interest: Dittmann et al. (2007) find that a loss aversion model dominates an equivalent risk aversion model in explaining observed CEO compensation contracts if the reference point is equal to the previous year's fixed wage. Iantchev (2005) finds evidence for his theoretically predicted results in panel data from Safelite Glass Corporation. Last, by explaining why bonuses are paid for good performance rather than penalties for poor performance, De Meza and Webb (2007) provide a theoretical underpinning for the frequent usage of option-like incentive schemes in CEO compensation. The contractual form predicted by these papers, however, is rather complex: while the optimal contract typically displays a range where pay is independent of performance, for performance above this range payment varies with performance in a fairly complex way, depending crucially on the underlying distribution of signals. Theoretical predictions differ in whether or not the optimal contract includes punishment for very poor performance or where in the wage schedule the optimal contract features discontinuities. Thus, none of these papers provides a rationale for the prevalence of fairly simple contracts, bonus contracts in particular.[14]

To the best of our knowledge, Daido and Itoh (2007) is the only paper that also applies the concept of reference dependence à la Kőszegi and Rabin to a principal-agent setting. The focus of Daido and Itoh, however, greatly differs from ours. Assuming that the performance measure comprises of only two signals, two types of self-fulfilling prophecy are explained, the Galatea and the Pygmalion effects.[15] While sufficient to

[12]De Meza and Webb consider two otherwise identical agents who differ only in their degree of loss aversion. They point out that with the certainty equivalent as reference point, there are situations where the less loss-averse agent experiences a loss, but the more loss-averse agent does not.

[13]For example, suppose that with a probability of .51 a manager earns \$1m and with a probability of .49 he earns \$2m. With median income as reference point the manager will never suffer a loss because his reference income is \$1m. A small shift in probabilities, however, makes the median income equal to \$2m. Now, the agent suffers a loss in almost 50% of all cases.

[14]De Meza and Webb (2007) find conditions under which a simple bonus contract is optimal. For this to be the case, however, they assume that the reference point is exogenously given and that all wage payments are in the loss region, where the agent is assumed to be risk loving.

[15]Roughly speaking, the former effect refers to empirical findings that an agent's self-expectation about his performance is an important determinant of his actual performance, whereas the latter effect refers to the phenomenon that a principal's expectation about the agent's performance has an impact on the agent's actual performance.

capture these two effects, the assumption of a binary measure of performance does not allow one to inquire into the form that contracts take under moral hazard.

2. The Model

There are two parties, a principal and an agent.[16] The principal offers a one-period employment contract to the agent, who has an outside employment opportunity (or reservation utility) yielding expected utility $\bar{u}$. If the agent accepts the contract, then he chooses an effort level $a \in \mathcal{A} \equiv [0,1]$. The agent's action a equals the probability that the principal receives a benefit $B > 0$. The principal's expected net benefit is

$$\pi = aB - E[W] \ ,$$

where W is the compensation payment the principal pays to the agent.[17] The principal is assumed to be risk and loss neutral, thus she maximizes π. We wish to inquire into the form that contracts take under moral hazard and loss aversion. Therefore, we focus on the cost minimization problem to implement a certain action $\hat{a} \in (0,1)$.[18]

The action choice $a \in \mathcal{A}$ is private information of the agent and unobservable for the principal. Furthermore, the realization of B is not directly observable. A possible interpretation is that B corresponds to a complex good whose quality cannot be determined by a court, thus a contract cannot depend on the realization of B. Instead the principal observes a contractible measure of performance, $\hat{\boldsymbol{\gamma}}$, with $s \in \mathcal{S} \equiv \{1,\ldots,S\}$ being the realization of the performance measure, also referred to as signal. Let $S \geq 2$. The probability of observing signal s conditional on B being realized is denoted by γ_s^H. Accordingly, γ_s^L is the probability of observing signal s conditional on B not being realized. Hence, the unconditional probability of observing signal s for a given action a is $\gamma_s(a) \equiv a\gamma_s^H + (1-a)\gamma_s^L$. For technical convenience, we make the following assumption.

Assumption (A1): *For all $s,\tau \in \mathcal{S}$ with $s \neq \tau$,*

(i) $\gamma_s^H/\gamma_s^L \neq 1$ *(informative signals),*

(ii) $\gamma_s^H, \gamma_s^L \in (0,1)$ *(full support),*

(iii) $\gamma_s^H/\gamma_s^L \neq \gamma_\tau^H/\gamma_\tau^L$ *(different signals).*

Assumption (i) guarantees that any signal s is either a good or a bad signal, in the

[16]The framework is based on MacLeod (2003), who analyzes subjective performance measures without considering loss-averse agents.

[17]The particular functional form of the principal's profit function is not crucial for our analysis. We assume this specific structure since it allows for a straight-forward interpretation of the performance measure.

[18]The second-best action maximizes the principal's expected benefit, aB, minus the minimum cost of implementing action a. The overall optimal contract exhibits the same characteristics as the contract that minimizes the cost of implementing an arbitrary action $\hat{a}$.

sense that the overall probability of observing that signal unambiguously increases or decreases in a. Part (ii) ensures that for all $a \in \mathcal{A}$, all signals occur with positive probability. Last, with assumption (iii) signals can unambiguously be ranked according to the relative impact of an increase in effort on the probability of observing a particular signal.[19]

The contract which the principal offers to the agent consists of a payment for each realization of the performance measure, $\{w_s\}_{s=1}^{S} \in \mathbb{R}^S$.[20]

The agent is assumed to have reference-dependent preferences in the sense of Kőszegi and Rabin (2006): Overall utility from consuming $\boldsymbol{x} = (x_1, \ldots, x_K) \in \mathbb{R}^K$ – when having reference level $\boldsymbol{r} = (r_1, \ldots, r_K) \in \mathbb{R}^K$ for each dimension of consumption – is given by

$$v(\boldsymbol{x}|\boldsymbol{r}) \equiv \sum_{k=1}^{K} m_k(x_k) + \sum_{k=1}^{K} \mu(m_k(x_k) - m_k(r_k)).$$

Put verbally, overall utility is assumed to have two components: consumption utility and gain-loss utility. Consumption utility, also called intrinsic utility, from consuming in dimension k is denoted by $m_k(x_k)$. How a person feels about gaining or losing in a dimension is assumed to depend in a universal way on the changes in consumption utility associated with such gains and losses. The universal gain-loss function $\mu(\cdot)$ satisfies the assumptions imposed by Tversky and Kahneman (1991) on their "value function".[21] In our model, the agent's consumption space comprises of two dimensions, money income ($x_1 = W$) and effort ($x_2 = a$).[22] The agent's intrinsic utility for money is assumed to be a strictly increasing, (weakly) concave, and unbounded function. Formally, $m_1(W) = u(W)$ with $u'(\cdot) > \varepsilon > 0$, $u''(\cdot) \leq 0$. The intrinsic disutility from exerting effort $a \in [0, 1]$ is a strictly increasing, strictly convex function of effort, $m_2(a) = -c(a)$ with $c'(0) = 0$, $c'(a) > 0$ for $a > 0$, $c''(\cdot) > 0$, and $\lim_{a \to 1} c(a) = \infty$. We assume that the gain-loss function is piece-wise linear,

$$\mu(m) = \begin{cases} m \;, & \text{for } m \geq 0 \\ \lambda m, & \text{for } m < 0 \end{cases} .$$

The parameter λ characterizes the weight put on losses relative to gains.[23] The weight

[19]Formally, for all $a \in [0, 1]$, $(\gamma_s^H - \gamma_s^L)/\gamma_s(a) > (\gamma_\tau^H - \gamma_\tau^L)/\gamma_\tau(a) \iff \gamma_s^H/\gamma_s^L > \gamma_\tau^H/\gamma_\tau^L$.

[20]Restricting the principal to offer non-stochastic wage payments is standard in the principal-agent literature and also in accordance with observed practice. In a later section we comment on this assumption.

[21]Roughly speaking, $\mu(z)$ is strictly increasing, continuous for all z, twice differentiable for all $z \neq 0$ with $\mu(0) = 0$, convex over the range of losses, and concave over the range of gains. For a more formal statement of these properties, see Bowman et al. (1999).

[22]We implicitly assume that the agent is a "narrow bracketer" in the sense that he ignores that the risk from the current employment relationship is incorporated with substantial other risk.

[23]Alternatively, one could assume that $\mu(m) = \eta m$ for gains and $\mu(m) = \eta\lambda m$ for losses, where $\eta \geq 0$ can be interpreted as the weight attached to gain-loss utility relative to intrinsic utility.

on gains is normalized to one. When $\lambda > 1$, the agent is loss averse in the sense that losses loom larger than equally-sized gains.[24]

Following Kőszegi and Rabin (2006, 2007), the agent's reference point is determined by his rational expectations about outcomes. A given outcome is then evaluated by comparing it to all possible outcomes, where each comparison is weighted with the ex-ante probability with which the alternative outcome occurs. With the actual outcome being itself uncertain, the agent's ex ante expected utility is obtained by averaging over all these comparisons.[25] We apply the concept of choice-acclimating personal equilibrium (CPE) as defined in Kőszegi and Rabin (2007), which assumes that a person correctly predicts his choice set, the environment he faces, in particular the set of possible outcomes and how the distribution of these outcomes depends on his decisions, and his own reaction to this environment. The eponymous feature of CPE is that the agent's reference point is affected by his choice of action. As pointed out by Kőszegi and Rabin, CPE refers to the analysis of risk preferences regarding outcomes that are resolved long after all decisions are made. This environment seems well-suited for many principal-agent relationships: Often the outcome of a project becomes observable, and thus performance-based wage compensation feasible, long after the agent finished working on that project. Under CPE, the expectations relative to which a decision's outcome is evaluated are formed at the moment the decision is made and, therefore, incorporate the implications of the decision. More precisely, suppose the agent chooses action a and that signal s is observed. The agent receives wage w_s and incurs effort cost $c(a)$. While the agent expected signal s to come up with probability $\gamma_s(a)$, with probability $\gamma_\tau(a)$ he expected signal $\tau \neq s$ to be observed. If $w_\tau > w_s$, the agent experiences a loss of $\lambda(u(w_s) - u(w_\tau))$, whereas if $w_\tau < w_s$, the agent experiences a gain of $u(w_s) - u(w_\tau)$. If $w_s = w_\tau$, there is no sensation of gaining or losing involved.

Our implicit normalization $\eta = 1$ is without loss of generality due to the applied concept of choice-acclimating personal equilibrium (CPE). Carrying η through the whole analysis would only replace $(\lambda - 1)$ by $\eta(\lambda - 1)$ in all formulas.

[24]The assumption of a piece-wise linear gain-loss function is not uncommon in the literature on incentive design with loss-averse agents, see De Meza and Webb (2007), Daido and Itoh (2007). In their work on asset pricing, Barberis et al. (2001) also apply this particular functional form, reasoning that "curvature is most relevant when choosing between prospects that involve only gains or between prospects that involve only losses. For gambles that can lead to both gains and losses, [...] loss aversion at the kink is far more important than the degree of curvature away from the kink."

[25]Suppose the actual outcome $\boldsymbol{x}$ and the vector of reference levels $\boldsymbol{r}$ are distributed according to distribution functions F and G, respectively. As introduced above, overall utility from two arbitrary vectors $\boldsymbol{x}$ and $\boldsymbol{r}$ is given by $v(\boldsymbol{x}|\boldsymbol{r})$. With the reference point being distributed according to probability measure G, the utility from a certain outcome is the average of how this outcome feels compared to all other possible outcomes, $U(\boldsymbol{x}|G) = \int v(\boldsymbol{x}|\boldsymbol{r})\ dG(\boldsymbol{r})$. Last, with $\boldsymbol{x}$ being drawn according to probability measure F, utility is given by $E[U(F|G)] = \iint v(\boldsymbol{x}|\boldsymbol{r})\ dG(\boldsymbol{r})dF(\boldsymbol{x})$. Since we use choice acclimating personal equilibrium, $F = G$.

The agent's utility from this particular outcome is given by

$$u(w_s) + \sum_{\{\tau | w_\tau < w_s\}} \gamma_\tau(a)(u(w_s) - u(w_\tau)) + \sum_{\{\tau | w_\tau \geq w_s\}} \gamma_\tau(a)\lambda(u(w_s) - u(w_\tau)) - c(a).$$

Averaging over all possible outcomes yields the agent's expected utility from choosing action a:

$$E[U(a)] = \sum_{s=1}^{S} \gamma_s(a) \Bigg\{ u(w_s) + \sum_{\{\tau | w_\tau < w_s\}} \gamma_\tau(a)(u(w_s) - u(w_\tau)) + \sum_{\{\tau | w_\tau \geq w_s\}} \gamma_\tau(a)\lambda(u(w_s) - u(w_\tau)) \Bigg\} - c(a).$$

Note that since the agent's expected and actual effort choice coincide, there is neither a gain nor a loss in the effort dimension.

We conclude this section by briefly summarizing the underlying timing.

1) The principal makes a take-it-or-leave-it offer to the agent.

2) The agent either accepts or rejects the contract. If the agent rejects the game ends and each party receives her/his reservation payoff. If the agent accepts the game moves to the next stage.

3) The agent chooses his action and forms rational expectations about the monetary outcomes. The agent's rational expectations about the realization of the performance measure determine his reference point.

4) Both parties observe the realization of the performance measure and payments are made according to the contract.

3. The Analysis

Let the inverse function of the agent's intrinsic utility of money be $h(\cdot)$, i.e., $h(\cdot) := u^{-1}(\cdot)$. Thus, the monetary cost for the principal to offer the agent utility u_s is $h(u_s) = w_s$. Due to the assumptions imposed on $u(\cdot)$, $h(\cdot)$ is a strictly increasing and weakly convex function. Following Grossman and Hart (1983), we regard $\boldsymbol{u} = \{u_1, \ldots, u_S\}$ as the principal's control variables in her cost minimization problem to implement action $\hat{a} \in (0, 1)$. The principal offers the agent a contract that specifies for each signal a monetary payment or, equivalently, an intrinsic utility level. With this notation, the agent's expected utility from exerting effort a is given by

$$E[U(a)] = \sum_{s \in \mathcal{S}} \gamma_s(a) u_s - (\lambda - 1) \sum_{s \in \mathcal{S}} \sum_{\{\tau | u_\tau > u_s\}} \gamma_\tau(a)\gamma_s(a)(u_\tau - u_s) - c(a). \tag{I.1}$$

For $\lambda = 1$ the agent's expected utility equals expected net intrinsic utility. Thus, for $\lambda = 1$ we are in the standard case without loss aversion. Moreover, from the above formulation of the agent's utility it becomes clear that λ captures not only the weight put on losses relative to gains, but $(\lambda - 1)$ also characterizes the weight put on gain-loss utility relative to intrinsic utility. Thus, for $\lambda \leq 2$, the weight attached to gain-loss utility is below the weight attached to intrinsic utility. For a given contract $\boldsymbol{u}$, the agent's marginal utility of effort

$$\begin{aligned} E[U'(a)] = & \sum_{s \in \mathcal{S}} (\gamma_s^H - \gamma_s^L) u_s \\ & - (\lambda - 1) \sum_{s \in \mathcal{S}} \sum_{\{\tau | u_\tau > u_s\}} [\gamma_\tau(a)(\gamma_s^H - \gamma_s^L) + \gamma_s(a)(\gamma_\tau^H - \gamma_\tau^L)](u_\tau - u_s) - c'(a). \quad \text{(I.2)} \end{aligned}$$

Suppose the principal wants to implement action $\hat{a} \in (0,1)$. The optimal contract minimizes the expected wage payment to the agent subject to the usual incentive compatibility and individual rationality constraints:

$$\min_{u_1, \ldots, u_S} \sum_{s \in \mathcal{S}} \gamma_s(\hat{a}) h(u_s)$$

$$\text{subject to} \quad E[U(\hat{a})] \geq \bar{u} \, , \qquad \text{(IR)}$$

$$\hat{a} \in \arg\max_{a \in \mathcal{A}} E[U(a)] \, . \qquad \text{(IC)}$$

As a first benchmark, consider the case where the agent's action choice is observable and contractible, i.e., the incentive constraint (IC) is absent. In order to implement action $\hat{a}$ in this first-best situation, the principal pays the agent $u^{FB} = \bar{u} + c(\hat{a})$ irrespective of the realization of the performance measure if the agent chooses the desired action, thereby compensating him for his outside option and his effort cost. In the presence of moral hazard, on the other hand, the principal faces the classic tradeoff between risk sharing and providing incentives: When the agent is anything but risk and loss neutral, it is neither optimal to have the agent bear the complete risk, nor to fully insure the agent.

At this point we simplify the analysis by imposing two assumptions. These assumptions are sufficient to guarantee that the principal's cost minimization problem exhibits the following two properties: First, there are incentive-compatible wage contracts, i.e., contracts under which it is optimal for the agent to choose the desired action $\hat{a}$. Existence of such contracts is not generally satisfied with the agent being loss averse. Second, the first-order approach is valid, i.e., the incentive constraint to implement action $\hat{a}$ can equivalently be represented as $E[U'(\hat{a})] = 0$. The first assumption that we introduce requires that the weight attached to gain-loss utility does not exceed the weight put on intrinsic utility.

Assumption (A2): *No dominance of gain-loss utility,* $\lambda \leq 2$.

As carefully laid out in Kőszegi and Rabin (2007), CPE implies a strong notion of risk aversion, in the sense that a decision maker may choose stochastically dominated options when $\lambda > 2$.[26] The reason is that, with losses looming larger than gains of equal size, the person ex ante expects to experience a net loss. In consequence, if reducing the scope of possibly incuring a loss is the decision maker's primary concern, the person would rather give up the slim hope of experiencing a gain at all in order to avoid the disappointment in case of not experiencing this gain. In our model, if the agent is sufficiently loss averse, the principal may be unable to implement any action $\hat{a} \in (0, 1)$. The reason is that the agent minimizes the ex ante expected net loss by choosing one of the two extreme actions. The values of λ for which this behavior is optimal for the agent depend on the precise structure of the performance measure. Assumption ($A2$) is sufficient, but not necessary, to ensure that there is a contract such that $\hat{a} \in (0, 1)$ satisfies the necessary condition for incentive compatibility. Moreover, the tendency to choose stochastically dominated options seems counterintuitive.[27] Next to ensuring existence of an incentive compatible contract, ($A2$) rules out that our findings are driven by such counterintuitive behavior of the agent. It is worthwhile to emphasize, that our main findings (Propositions 2 and 6) still hold for $\lambda > 2$ as long as existence and validity of the first-order approach are guaranteed. In Section 4 we relax Assumption ($A2$) and discuss in detail the implications of higher degrees of loss aversion.

To keep the analysis tractable we impose the following assumption.

Assumption (A3): *Convex marginal cost function,* $\forall\, a \in [0, 1] : \; c'''(a) \geq 0$.

Given ($A2$), Assumption ($A3$) is a sufficient but not a necessary condition for the first-order approach to be applicable.[28] Alternatively, it would also suffice to have λ sufficiently small, or the slope of the marginal cost function sufficiently steep. In fact, our results only require the validity of the first-order approach, not that Assumption ($A3$) holds. In Section 4 we consider the case in which the first-order approach is invalid.

[26] Suppose a loss-averse person has to choose between two lotteries: lottery 1 pays x for sure; lottery 2 pays $x + y$ with probability p, where $y > 0$, and x otherwise. Then, for each $\lambda > 2$, the decision maker prefers the dominated lottery 1 if $p < (\lambda - 2)/(\lambda - 1)$. For further details on this point, see Kőszegi and Rabin (2007).

[27] The "uncertainty effect" identified by Gneezy et al. (2006) refers to people valuing a risky prospect less than its worst possible outcome. While this may be interpreted as experimental evidence for people having preferences for stochastically dominated options, this finding crucially relies on the lottery currency not being stated in purely monetary terms. Therefore, we believe that in the context of wage contracts most people do not choose dominated options.

[28] The validity of the first-order approach under assumptions (A1)-(A3) is rigorously proven in the appendix. The reader should be aware, however, that the proof requires some notation introduced later on. We therefore recommend to defer reading the proof until having read the preliminary considerations up to Section 3.1.

Lemma I.1: *Suppose (A1)-(A3) hold, then the constraint set of the principal's cost minimization problem is nonempty for all* $\hat{a} \in (0,1)$.

The above lemma states that there are wage contracts such that the agent is willing to accept the contract and then chooses the desired action. Moreover, we will show that a second-best optimal contract exists. This, however, is shown separately for the three cases analyzed in this section: pure risk aversion, pure loss aversion, and the intermediate case.

Sometimes it will be convenient to state the constraints in terms of increases in intrinsic utilities instead of absolute utilities. Note that whatever contract $\{\hat{u}_s\}_{s\in\mathcal{S}}$ the principal offers, we can relabel the signals such that this contract is equivalent to a contract $\{u_s\}_{s=1}^S$ with $u_{s-1} \le u_s$ for all $s \in \{2,\dots,S\}$. This, in turn, allows us to write the contract as $u_s = u_1 + \sum_{\tau=2}^{s} b_\tau$, where $b_\tau = u_\tau - u_{\tau-1} \ge 0$ is the increase in intrinsic utility for money when signal τ instead of signal $\tau-1$ is observed. Let $\boldsymbol{b} = (b_2,\dots,b_S)$. Using this notation allows us to rewrite the individual rationality constraint as follows:

$$u_1 + \sum_{s=2}^{S} b_s \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) - \rho_s(\hat{\boldsymbol{\gamma}},\lambda,\hat{a})\right] \ge \bar{u} + c(\hat{a}) \ , \tag{IR$'$}$$

where

$$\rho_s(\hat{\boldsymbol{\gamma}},\lambda,\hat{a}) := (\lambda-1)\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right]\left[\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right] .$$

Let $\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}},\lambda,\hat{a}) = (\rho_2(\hat{\boldsymbol{\gamma}},\lambda,\hat{a}),\dots,\rho_S(\hat{\boldsymbol{\gamma}},\lambda,\hat{a}))$. The first part of the agent's utility, $u_1 + \sum_{s=2}^{S} b_s(\sum_{\tau=s}^{S}\gamma_\tau(\hat{a}))$, is the expected intrinsic utility for money. Due to loss aversion, however, the agent's utility has a second negative component, the term $\boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}},\lambda,\hat{a})$. Where does this term come from? With bonus b_s being paid to the agent whenever a signal higher or equal to s is observed, the agent expects to receive b_s with probability $\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})$. With probability $\sum_{t=1}^{s-1}\gamma_t(\hat{a})$, however, a signal below s will be observed, and the agent will not be paid bonus b_s. Thus, with "probability" $[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})][\sum_{t=1}^{s-1}\gamma_t(\hat{a})]$ the agent experiences a loss of λb_s. Analogous reasoning implies that the agent will experience a gain of b_s with the same probability. With losses looming larger than gains of equal size, in expectation the agent suffers from deviations from his reference point. This ex ante expected net loss is captured by the term, $\boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}},\lambda,\hat{a})$, which we will refer to as the agent's "loss premium".[29] A crucial point is that the loss premium increases in the complexity of the contract. When there is no wage differentiation at all, i.e., $\boldsymbol{b} = \boldsymbol{0}$, then the loss premium vanishes. If, in contrast, the contract specifies many different wage payments, then the agent ex ante considers a

[29] Our notion of the agent's loss premium is highly related to the average self-distance of a lottery defined by Kőszegi and Rabin (2007). Let $D(\boldsymbol{u})$ be the average self-distance of incentive scheme $\boldsymbol{u}$, then $[(\lambda-1)/2]D(\boldsymbol{u}) = \boldsymbol{b}'\boldsymbol{\rho}(\hat{\gamma},\lambda,\hat{a})$.

deviation from his reference point very likely. Put differently, for each additional wage payment an extra negative term enters the agent's loss premium and therefore reduces his expected utility.[30]

Given the first-order approach is valid, the incentive constraint can be rewritten as

$$\sum_{s=2}^{S} b_s \beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}), \tag{IC$'$}$$

where

$$\begin{aligned}\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) :=& \left(\sum_{\tau=s}^{S}(\gamma_\tau^H - \gamma_\tau^L)\right)\\ &- (\lambda - 1)\left[\left(\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right)\left(\sum_{\tau=s}^{S}(\gamma_\tau^H - \gamma_\tau^L)\right) + \left(\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{s-1}(\gamma_t^H - \gamma_t^L)\right)\right].\end{aligned}$$

Here, $\beta_s(\cdot)$ is the marginal effect on incentives of an increase in the wage payments for signals above $s-1$. Without loss aversion, i.e., $\lambda = 1$, this expression equals the marginal probability of observing at least signal s. If the agent is loss averse, on the other hand, an increase in the action also affects the agent's loss premium. The probability of bearing a loss of size b_s is a quadratic function of the probability of observing at least signal s. The agent's action balances the tradeoff between maximizing intrinsic utility and minimizing the expected net loss. Depending on the precise signal structure and the action to be implemented, loss aversion may facilitate as well as hamper the creation of incentives. Let $\boldsymbol{\beta}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = (\beta_2(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}), \ldots, \beta_S(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}))$.

As in the standard case, incentives are created solely by increases in intrinsic utilities, $\boldsymbol{b}$. In consequence, (IR$'$) is binding in the optimum. If this was not the case, i.e., if $\boldsymbol{b}$ satisfies (IC$'$) but (IR$'$) holds with strict inequality, then the principal can lower payment u_1 up to the point where the (IR$'$) is satisfied with equality. Thus, reducing u_1 while holding $\boldsymbol{b}$ constant lowers the principal's expected wage payment while preserving incentives.

It is obvious that (IC$'$) can only be satisfied if there exists at least one $\beta_s > 0$. If, for example, signals are ordered according to their likelihood ratios, then $\beta_s(\cdot) > 0$ for all $s = 2, \ldots, S$. More precisely, for a given ordering of signals, under $(A2)$ the following equivalence follows immediately from the fact that $\sum_{t=1}^{s-1}(\gamma_t^H - \gamma_t^L) = -\sum_{\tau=s}^{S}(\gamma_\tau^H - \gamma_\tau^L)$:

$$\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) > 0 \iff \sum_{\tau=s}^{S}(\gamma_\tau^H - \gamma_\tau^L) > 0\,. \tag{I.3}$$

[30] While the exact change of the loss premium from adding more and more wage payments is hard to grasp, this point can heuristically be illustrated by considering the upper bound of the loss premium. Suppose the principal sets $n \leq S$ different wages. It is readily verified that the loss premium is bounded from above by $(\lambda-1)[(u_S - u_1)/2] \times [(n-1)/n]$, and that this upper bound increases as n increases. Note, however, that even for $n \to \infty$ the upper bound of the loss premium is finite.

3.1. Two Polar Cases: Pure Risk Aversion vs. Pure Loss Aversion

In this part of the paper we analyze the two polar cases: The standard case where the agent is only risk averse but not loss averse, on the one hand, and the case of a loss-averse agent with a risk-neutral intrinsic utility function, on the other hand.

Pure Risk Aversion

First consider an agent who is risk averse in the usual sense, i.e., $h''(\cdot) > 0$, but does not exhibit loss aversion. As discussed earlier, the latter requirement corresponds to the case where $\lambda = 1$. With the agent not being loss averse, the first-order approach is valid even without Assumption $(A3)$.

Proposition I.1 (Holmström, 1979): *Suppose $(A1)$ holds, $h''(\cdot) > 0$, and $\lambda = 1$. Then there exists a second-best optimal contract to implement $\hat{a} \in (0,1)$. The second-best contract has the property that $u_s \neq u_\tau$ $\forall s, \tau \in \mathcal{S}$ and $s \neq \tau$. Moreover, $u_s > u_\tau$ if and only if $\gamma_s^H/\gamma_s^L > \gamma_\tau^H/\gamma_\tau^L$.*

Proposition 1, restates the well-known finding by Holmström (1979) for discrete signals: Signals that are more indicative of higher effort, i.e., signals with a higher likelihood ratio γ_s^H/γ_s^L, are rewarded strictly higher. Thus, the optimal wage scheme is complex in the sense that it is fully differentiated, with each signal being rewarded differently.

Pure Loss Aversion

Having considered the polar case of pure risk aversion, we now turn to the other extreme, a purely loss-averse agent. Formally, intrinsic utility of money is a linear function, $h''(\cdot) = 0$, and the agent is loss averse, $\lambda > 1$. As we have already reasoned, whatever contract the principal offers, relabeling the signals always allows us to represent this contract as an (at least weakly) increasing intrinsic utility profile. Therefore we can decompose the principal's problem into two steps: first, for a given ordering of signals, choose a nondecreasing profile of intrinsic utility levels that implements the desired action $\hat{a}$ at minimum cost; second, choose the signal ordering with the lowest cost of implementation. As we know from the discussion at the end of the previous section, a necessary condition for an upward-sloping incentive scheme to achieve incentive compatibility is that for the underlying signal ordering at least one $\beta_s(\cdot) > 0$. In what follows we restrict attention to the set of signal orderings that are incentive feasible in the afore-mentioned sense. Nonemptiness of this set follows immediately from Lemma 1.

Consider the first step of the principal's problem, i.e., taking the ordering of signals as given, find the nondecreasing payment scheme with the lowest cost of implementation. In what follows, we write the agent's intrinsic utility in terms of additional payments, $u_s = u_1 + \sum_{\tau=2}^{S} b_\tau$. With $h(\cdot)$ being linear, the principal's objective function is $C(u_1, \boldsymbol{b}) = u_1 + \sum_{s=2}^{S} b_s(\sum_{\tau=2}^{S} \gamma_\tau(\hat{a}))$. Remember that at the optimum, (IR$'$) holds with equality. Inserting (IR$'$) into the principal's objective allows us to write the cost minimization problem for a given order of signals in the following simple way:

Program ML:

$$\min_{\boldsymbol{b} \in \mathbb{R}_+^{S-1}} \ \boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$$

$$\text{subject to} \quad \boldsymbol{b}'\boldsymbol{\beta}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \qquad \text{(IC}'\text{)}$$

The minimization problem (ML) has a simple intuition. The principal seeks to minimize the agent's expected net loss subject to the incentive compatibility constraint. Similar to the case of pure risk aversion, where the principal would like to cut back the agent's risk premium, here she is interested in minimizing the agent's loss premium. Due to the incentive constraint, however, this loss premium has to be strictly positive.

We want to emphasize that solving Program ML also yields insights for the more general case with a concave intrinsic utility function. Even though the principal's objective will not reduce to minimizing the agent's loss premium alone when intrinsic utility is non-linear, this nevertheless remains an important aspect of the principal's problem. Since the solution to Program ML tells us how to minimize the agent's loss premium irrespective of the functional form of intrinsic utility, one should expect its properties to carry over to some extent to the solution of the more general problem.

The principal's cost minimization problem for a given order of signals is a simple linear programming problem: minimize a linear objective function subject to one linear equality constraint. Since we restricted attention to orderings of signals with $\beta_s(\cdot) > 0$ for at least one signal s, a solution to (ML) exists. Due to the linear nature of problem (ML), (generically) this solution sets exactly one $b_s > 0$ and all other $b_s = 0$. Put differently, the problem is to find that b_s which creates incentives at the lowest cost.

So far we have seen that, for a given ordering of signals, the principal considers it optimal to offer the agent a bonus contract: pay a low wage for signals below some threshold, and a high wage for signals above this threshold. What remains to do for the principal, in a second step, is to find the signal ordering that leads to the lowest cost of implementation. With the number of different orders of signals being finite, this problem clearly has a solution.

Before summarizing the above discussion more concisely, we want to relate our finding to the benchmark case of pure risk neutrality. As is well-known, with both contracting parties being risk (and loss) neutral a broad range of contracts – including simple bonus schemes – is optimal. With the agent being loss averse even to a negligible degree, however, the unique optimal contractual form is a bonus scheme.[31]

Proposition I.2: *Suppose (A1)-(A3) hold, $h''(\cdot) = 0$ and $\lambda > 1$. Then there exists a second-best optimal contract to implement action $\hat{a} \in (0, 1)$. Generically, the second-best optimal incentive scheme $\{u_s^*\}_{s=1}^S$ is a bonus contract, i.e., $u_s^* = u_H^*$ for $s \in \mathcal{B}^* \subset \mathcal{S}$ and $u_s^* = u_L^*$ for $s \in \mathcal{S} \setminus \mathcal{B}^*$, where $u_H^* > u_L^*$.*

According to Proposition 2, the principal considers it optimal to offer the agent a bonus contract which entails only a minimum degree of wage differentiation in the sense that, no matter how rich the signal space, the contract specifies only two different wage payments. This endeavor to reduce the complexity of the contract is plausible, since a high degree of wage differentiation increases the agent's loss premium: With the employment contract she offers to the agent, the principal determines the dimensionality of the agent's reference point. The higher the dimensionality of the reference point is, the more likely it is that the agent incurs a loss in a particular dimension. Therefore, with the concept of reference-dependent preferences developed by Kőszegi and Rabin, it truly pains a person to be exposed to numerous potential outcomes. This disutility of the agent from facing several possible (monetary) outcomes, which he demands to be compensated for, makes it costly for the principal to offer complex contracts. In consequence, the optimal contract entails only a minimum of wage differentiation. To provide a more intuitive explanation for this finding, consider a principal who – starting out from a given wage scheme – has to improve incentives. There are basically two ways to do so. On the one hand, the principal can introduce a new wage spread, i.e., pay slightly different wages for two signals that were rewarded equally in the original wage scheme, while keeping the differences between all other neighboring wages constant. On the other hand, the principal can increase an existing wage spread, holding constant all other spreads between neighboring wages. Both procedures increase the loss premium by increasing the size of some of the the expected losses without reducing others. Introducing a new wage spread, however, additionally increases the loss premium by increasing the ex ante expected probability of experiencing a loss. Therefore, in order to improve incentives for a loss-averse agent, it is advantageous to increase a

[31]If, in addition to both the principal and the agent being risk neutral, the agent is protected by limited liability, Park (1995), Kim (1997), Oyer (2000), and Demougin and Fluet (1998) show that the optimal contract is a bonus scheme. These findings, however, immediately collapse when the agent is somewhat risk averse as is demonstrated by Jewitt et al. (2008). Our findings, on the other hand, are robust towards introducing a slightly concave intrinsic utility function, as we will illustrate in Section 3.2.

particular existing wage spread without adding to the contractual complexity in the sense of increasing the number of different wages. Under the standard notion of a risk-averse agent, however, one should not expect to encounter this tendency to reduce the complexity of contracts. The reason is that increasing incentives by introducing a small new wage spread is basically costless for the principal because locally the agent is risk neutral. Therefore, under risk aversion different signals are rewarded differently.

Up to now, however, we have not specified which signals are generally included in the set $\mathcal{B}^*$. In light of the above observation, the principal's problem boils down to choosing a binary partition of the set of signals, $\mathcal{B} \subset \mathcal{S}$, which characterizes for which signals the agent receives the high wage and for which signals he receives the low wage. The wages u_L and u_H are then uniquely determined by the corresponding individual rationality and incentive compatibility constraints. The problem of choosing the optimal partition of signals, $\mathcal{B}^*$, which minimizes the principal's expected cost of implementing action $\hat{a}$ is an integer programming problem. As is typical for this class of problems, and as is nicely illustrated by the well-known "0-1 Knapsack Problem", it is impossible to provide a general characterization of the solution.[32]

Next to these standard intricacies of integer programming, there is an additional difficulty in our model: the principal's objective behaves non-monotonically when including an additional signal into the "bonus set" $\mathcal{B}$. This is due to different – possibly conflicting – targets that the principal pursues when deciding how to partition the set $\mathcal{S}$. From Program (ML) it follows that, for a given "bonus set" $\mathcal{B}$, the minimum cost of implementing action $\hat{a}$ is given by

$$C_{\mathcal{B}} = \bar{u} + c(\hat{a}) + \frac{c'(\hat{a})(\lambda - 1)P_{\mathcal{B}}(1 - P_{\mathcal{B}})}{[\sum_{s\in\mathcal{B}} \gamma_s^H - \gamma_s^L][1 - (\lambda - 1)(1 - 2P_{\mathcal{B}})]}\,, \tag{I.4}$$

where $P_{\mathcal{B}} := \sum_{s\in\mathcal{B}} \gamma_s(\hat{a})$. The above costs can be rewritten such that the principal's problem amounts to

$$\max_{\mathcal{B}\subset\mathcal{S}} \left[\sum_{s\in\mathcal{B}} (\gamma_s^H - \gamma_s^L)\right] \left\{\frac{1}{(\lambda - 1)P_{\mathcal{B}}(1 - P_{\mathcal{B}})} - \frac{1}{P_{\mathcal{B}}} + \frac{1}{1 - P_{\mathcal{B}}}\right\}\,. \tag{I.5}$$

This objective function illustrates the tradeoff that the principal faces when deciding how to partition the signal space. The first term, $\sum_{s\in\mathcal{B}}(\gamma_s^H - \gamma_s^L)$, is the aggregate marginal impact of effort on the probability of the bonus $b := u_H - u_L$ being paid out. In order to create incentives for the agent, the principal would like to make this term

[32]The "0-1 Knapsack Problem" refers to a hiker who has to select from a group of items, all of which may be suitable for her trip, a subset that has greatest value while not exceeding the capacity of her knapsack. Suppose there are n items, each item j has a value $v_j > 0$ and a weight $w_j > 0$. Let the capacity of the knapsack be $c > 0$. The 0-1 Knapsack Problem may be formulated as the following maximization problem: max $\sum_{j=1}^n v_j x_j$ subject to $\sum_{j=1}^n w_j x_j \leq c$ and $x_j \in \{0, 1\}$ for $j = 1, \ldots, n$.

as large as possible, which in turn allows her to lower the bonus payment. This can be achieved by including only good signals in $\mathcal{B}$. The second term, on the other hand, is maximized by making the probability of paying the agent the high wage either as large as possible or as small as possible, depending on the exact signal structure and the action to be implemented. With the loss premium being given by $(\lambda - 1)P_{\mathcal{B}}(1 - P_{\mathcal{B}})b$, this is intuitive: By making the event of paying the high wage very likely or unlikely, the principal minimizes the scope for the agent to experience a loss that he demands to be compensated for. Depending on the signal structure, these two goals may conflict with each other, which makes a complete characterization of the optimal contract very intricate. Nevertheless, it can be shown that the optimal contract displays the following plausible property.

Proposition I.3: *Let $\mathcal{S}^+ \equiv \{s \in \mathcal{S} | \gamma_s^H - \gamma_s^L > 0\}$. The optimal partition of the signals for which the high wage is paid, $\mathcal{B}^*$, has the following property: either $\mathcal{B}^* \subseteq \mathcal{S}^+$ or $\mathcal{S}^+ \subseteq \mathcal{B}^*$.*

Put verbally, the optimal partition of the signal set takes one of the two possible forms: the high wage is paid out to the agent (i) either only for good signals though possibly not for all good signals, or (ii) for all good signals and possibly a few bad signals as well. Loosely speaking, if the principal considers it optimal to pay the high wage very rarely, she will reward only good signals with the extra payment b. If, on the other hand, she wants the agent to receive the high wage with high probability, then she will reward at least all good signals.

Without further assumptions, due to the discrete nature of the problem it is hard to characterize the signals that are included in $\mathcal{B}^*$. Back to the "0-1 Knapsack Problem", here it is well-established for the continuous version of the problem that the solution can easily be found by ordering the items according to their value-to-weight ratio.[33] Defining $\kappa := \max_{\{s,t\}\subseteq\mathcal{S}} |\gamma_s(\hat{a}) - \gamma_t(\hat{a})|$, we can obtain a similar result. Assuming that κ is sufficiently small, which is likely to hold if the performance measure is, for instance, sales revenues measured in cents, makes the principal's problem of choosing $\mathcal{B}^*$ similar to a continuous problem.[34] With this assumption, we can show that it is optimal to order the signals according to their likelihood ratios.

Proposition I.4: *Suppose κ is sufficiently small, then there exists a constant K such that $\mathcal{B}^* = \{s \in \mathcal{S} \mid \gamma_s^H/\gamma_s^L \geq K\}$.*

[33]In the continuous "0-1 Knapsack Problem" the constraints on the variables $x_j \in \{0, 1\}$ are relaxed to $x_j \in [0, 1]$, e.g. Dantzig (1957).

[34]Here, the probability of observing a specific signal, say, sales revenues of exactly \$13,825.32 is rather small.

Though wage payments are only weakly increasing in the likelihood ratio, this finding resembles the standard result for a risk-averse agent, where the incentive scheme is strictly increasing in the likelihood ratio.

Before moving on to the discussion of the more general case in which the agent is both risk and loss averse, we want to pause to point out the following comparative static results.

Proposition I.5: *An increase in the agent's degree of loss aversion (i) strictly increases the minimum cost of implementing action $\hat{a}$; (ii) decreases the necessary wage spread to implement action $\hat{a}$ if and only if $P_{\mathcal{B}^*} > 1/2$, given that the change in λ does not lead to a change of $\mathcal{B}^*$.*

Part (ii) of Proposition 5 relates to the reasoning by Kőszegi and Rabin (2006) that if the agent is loss averse and expectations are the driving force in the determination of the reference point, then "in principal-agent models, performance-contingent pay may not only directly motivate the agent to work harder in pursuit of higher income, but also indirectly motivate [him] by changing [his] expected income and effort." As can be seen from (1), the agent's expected utility under the second-best contract comprises of two components, the first of which is expected net intrinsic utility from choosing effort level $\hat{a}$, $u_L + b^* \sum_{s \in \mathcal{B}^*} \gamma_s(\hat{a}) - c(\hat{a})$. Due to loss aversion, however, there is a second component: With losses looming larger than equally sized gains, in expectation the agent suffers from deviations from his reference point. While the strength of this effect is determined by the degree of the agent's loss aversion, λ, his action choice – together with the signal parameters – determines the probability that such a deviation from the reference point actually occurs. We refer to this probability, which is given by $P_{\mathcal{B}^*}(1 - P_{\mathcal{B}^*})$, as loss probability. Therefore, when choosing his action, the agent has to balance off two possibly conflicting targets, maximizing expected net intrinsic utility and minimizing the loss probability. The loss probability, which is a strictly concave function of the agent's effort, is locally decreasing at $\hat{a}$ if and only if $P_{\mathcal{B}^*} > 1/2$. In this case, an increase in λ, which makes reducing the loss probability more important, may lead to the agent choosing a higher effort level, which in turn allows the principal to use lower-powered incentives. The principal, however, cannot capitalize on this since, according to part (i) of Proposition 5, the overall cost of implementation strictly increases in the agent's degree of loss aversion.

3.2. The General Case: Loss Aversion and Risk Aversion

We now turn to the intermediate case where the agent is both risk and loss averse. The agent's intrinsic utility for money is a strictly increasing and strictly concave function, which implies that $h(\cdot)$ is strictly increasing and strictly convex. Moreover, the agent is loss averse, i.e., $\lambda > 1$. From Lemma 1, we know that the constraint set of the principal's problem is nonempty. By relabeling signals, each contract can be interpreted as a contract that offers the agent a (weakly) increasing intrinsic utility profile. This allows us to assess whether the agent perceives receiving u_s instead of u_t as a gain or a loss. As in the case of pure loss aversion, we analyze the optimal contract for a given feasible ordering of signals.

The principal's problem for a given arrangement of the signals is given by

PROGRAM MG:

$$\min_{u_1,\dots,u_S} \sum_{s=1}^{S} \gamma_s(\hat{a}) h(u_s)$$

subject to

$$\sum_{s=1}^{S} \gamma_s(\hat{a}) u_s - (\lambda - 1) \sum_{s=1}^{S-1} \sum_{t=s+1}^{S} \gamma_s(\hat{a}) \gamma_t(\hat{a}) [u_t - u_s] - c(\hat{a}) = \bar{u} \ , \tag{IR$_G$}$$

$$\sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) u_s - (\lambda - 1) \sum_{s=1}^{S-1} \sum_{t=s+1}^{S} \left[\gamma_s(\hat{a})(\gamma_t^H - \gamma_t^L) + \gamma_t(\hat{a})(\gamma_s^H - \gamma_s^L) \right] [u_t - u_s] = c'(\hat{a}) \ , \tag{IC$_G$}$$

$$u_S \geq u_{S-1} \geq \ldots \geq u_1 \ . \tag{OC$_G$}$$

Since the objective function is strictly convex and the constraints are all linear in $\boldsymbol{u} = \{u_1, \dots, u_S\}$, the Kuhn-Tucker theorem yields necessary and sufficient conditions for optimality. Put differently, if there exists a solution to the problem (MG) the solution is characterized by the partial derivatives of the Lagrangian associated with (MG) set equal to zero.

Lemma I.2: *Suppose (A1)-(A3) hold and $h''(\cdot) > 0$, then there exists a second-best optimal incentive scheme for implementing action $\hat{a} \in (0,1)$, denoted $\boldsymbol{u}^* = \{u_1^*, \dots, u_S^*\}$.*

In order to interpret the first-order conditions of the Lagrangian to problem (MG) it is necessary to know whether the Lagrangian multipliers are positive or negative.

Lemma I.3: *The Lagrangian multipliers of program (MG) associated with the incentive compatibility constraint and the individual rationality constraint are both strictly positive, i.e., $\mu_{IC} > 0$ and $\mu_{IR} > 0$.*

We now give a heuristic reasoning why pooling of information may well be optimal in this more general case. For the sake of argument, suppose there is no pooling of information in the sense that it is optimal to set distinct wages for distinct signals. In this case all order constraints are slack; formally, if $u_s \neq u_{s'}$ for all $s, s' \in \mathcal{S}$ and $s \neq s'$, then $\mu_{O,s} = 0$ for all $s \in \{2, \ldots, S\}$. In this case, i.e., when none of the ordering constraints is binding, then the first-order condition of optimality with respect to u_s, $\partial \mathcal{L}(\boldsymbol{u})/\partial u_s = 0$, can be written as follows:

$$h'(u_s) = \underbrace{\left(\mu_{IR} + \mu_{IC}\frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}\right)}_{=:H_s} \underbrace{\left[1 - (\lambda - 1)\left(2\sum_{t=1}^{s-1}\gamma_t(\hat{a}) + \gamma_s(\hat{a}) - 1\right)\right]}_{=:\Gamma_s} \\ \underbrace{- \mu_{IC}(\lambda - 1)\left[2\sum_{t=1}^{s-1}(\gamma_t^H - \gamma_t^L) + (\gamma_s^H - \gamma_s^L)\right]}_{=:\Lambda_s}. \quad \text{(I.6)}$$

For $\lambda = 1$ we have $h'(u_s) = H_s$, the standard "Holmström-formula".[35] Note that $\Gamma_s > 0$ for $\lambda \leq 2$. More importantly, irrespective of the signal ordering, we have $\Gamma_s > \Gamma_{s+1}$. The third term, Λ_s, can be either positive or negative. If the compound signal of all signals below s and the signal s itself are bad signals, then $\Lambda_s < 0$.

Since the incentive scheme is nondecreasing, when the order constraints are not binding it has to hold that $h'(u_s) \geq h'(u_{s-1})$. Thus, if $\mu_{OC,s-1} = \mu_{OC,s} = \mu_{OC,s+1} = 0$ the following inequality is satisfied:

$$H_s \times \Gamma_s - \Lambda_s \geq H_{s-1} \times \Gamma_{s-1} - \Lambda_{s-1}. \quad \text{(I.7)}$$

Note that for the given ordering of signals, if there exists any pair of signals $s, s-1$ such that (I.7) is violated, then the optimal contract for this ordering involves pooling of wages. Even when $H_s > H_{s-1}$, as it is the case when signals are ordered according to their likelihood ratio, it is not clear that inequality (I.7) is satisfied. In particular, when s and $s-1$ are similarly informative it seems to be optimal to pay the same wage for these two signals as can easily be illustrated for the case of two good signals: If s and $s-1$ are similarly informative good signals then $H_s \approx H_{s-1} > 0$ but $\Gamma_s < \Gamma_{s-1}$ and $\Lambda_s > \Lambda_{s-1}$, thus condition (I.7) is violated. In summary, it may well be that for a given incentive-feasible ordering of signals, and thus overall as well, the order constraints are binding, i.e., it may be optimal to offer a contract which is less complex than the signal space allows for. We illustrate this conjecture in the following with an example.

Application with Constant Relative Risk Aversion In the general case of a risk and loss-averse agent the principal seeks to minimize the loss and the risk premium.

[35] See Holmström (1979). This formula is also referred to as the modified Borch sharing rule (Borch, 1962).

Roughly speaking, the risk premium is increasing in the curvature of the agent's intrinsic utility function. Put differently, when the agent's intrinsic utility function becomes close to linearity the risk premium goes to zero. Thus, for a slightly concave intrinsic utility function one should expect that the principal's main objective is to minimize the loss premium, which is achieved by a bonus scheme as is shown in Section 3.1. In the following we show that these reasoning is correct for the case of an intrinsic utility function that features constant relative risk aversion (CRRA).

Suppose $h(u) = u^r$, with $r \geq 0$ being a measure for the agent's risk aversion. More precisely, the Arrow-Pratt measure for relative risk aversion of the agent's intrinsic utility function is $R = 1 - \frac{1}{r}$ and therefore constant. The following result states that the optimal contract is still a bonus contract when the agent is not only loss averse, but also slightly risk averse.

Proposition I.6: *Suppose (A1)-(A3) hold, $h(u) = u^r$ with $r > 1$, and $\lambda > 1$. Generically, for r sufficiently small the optimal incentive scheme $\{u_s^*\}_{s=1}^S$ is a bonus scheme, i.e., $u_s^* = u_H^*$ for $s \in \mathcal{B}^* \subset \mathcal{S}$ and $u_s^* = u_L^*$ for $s \in \mathcal{S} \backslash \mathcal{B}^*$ where $u_L^* < u_H^*$.*

Without loss aversion, in contrast, according to Proposition 1 the optimal contract is fully differentiated even for intrinsic utility being arbitrarily close to linearity.

Next, we demonstrate by means of an example that pooling of signals may well be optimal even for a non-negligible degree of risk aversion. Suppose the agent's effort cost is $c(a) = (1/2)a^2$ and the effort level to be implemented is $\hat{a} = \frac{1}{2}$. Moreover, we assume that the reservation utility $\bar{u} = 10$, which guarantees that all utility levels are positive.[36] To keep the example as simple as possible, it is assumed that the agent's performance can take only three values, i.e., the agent's performance is either excellent (E), satisfactory (S) or inadequate (I). We consider two specifications of the performance measure. In the first specification the satisfactory signal is a good signal, whereas in the second specification it is a bad signal. In all parameter constellations we consider, it turns out that it is always (weakly) optimal to order signals according to their likelihood ratio, i.e., $u_1 = u_I$, $u_2 = u_S$ and $u_3 = u_E$. In the first specification the conditional probabilities take the following values:

$$\begin{aligned} \gamma_E^H &= 5/10 \qquad & \gamma_E^L &= 1/10 \\ \gamma_S^H &= 4/10 \qquad & \gamma_S^L &= 3/10 \\ \gamma_I^H &= 1/10 \qquad & \gamma_I^L &= 6/10\ . \end{aligned}$$

The structure of the optimal contract for this specification and various values of r and λ is presented in Table 1.

[36] Increasing $\bar{u}$ makes the agent less (absolutely) risk averse and thus is similar to a reduction in r.

r \ λ	1.0	1.1	1.3	1.5
1.5	$u_1 < u_2 < u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$
2	$u_1 < u_2 < u_3$	$u_1 < u_2 < u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$
3	$u_1 < u_2 < u_3$	$u_1 < u_2 < u_3$	$u_1 < u_2 = u_3$	$u_1 < u_2 = u_3$

Table I.1.: Structure of the optimal contract with two "good" signals.

Table 1 suggests that the optimal contract typically involves pooling of the two good signals, in particular when the agent's intrinsic utility is not too concave, i.e., if the agent is not too risk averse. Table 1 nicely illustrates the trade-off the principal faces when the agent is both, risk and loss averse: If the agent becomes more risk averse pooling is less likely to be optimal. If, on the other hand, he becomes more loss averse, pooling is more likely to be optimal.[37]

In the second specification we assume that there are two bad signals. The conditional probabilities are as follows:

$$\gamma_E^H = 6/10 \qquad \gamma_E^L = 1/10$$
$$\gamma_S^H = 2/10 \qquad \gamma_S^L = 4/10$$
$$\gamma_I^H = 2/10 \qquad \gamma_I^L = 5/10\ .$$

The results for this case are presented in Table 2.

r \ λ	1.0	1.1	1.3	1.5
1.5	$u_1 < u_2 < u_3$	$u_1 = u_2 < u_3$	$u_1 = u_2 < u_3$	$u_1 = u_2 < u_3$
2	$u_1 < u_2 < u_3$	$u_1 = u_2 < u_3$	$u_1 = u_2 < u_3$	$u_1 = u_2 < u_3$
3	$u_1 < u_2 < u_3$	$u_1 = u_2 < u_3$	$u_1 = u_2 < u_3$	$u_1 = u_2 < u_3$

Table I.2.: Structure of the optimal contract with two "bad" signals.

In this specification, a binary statistic that pools the two bad signals seems to be optimal almost always. The reason behind this observation is that the two bad signals are very similar. In consequence, paying the same wage for satisfactory as well as inadequate performance increases the risk premium only slightly. On the other hand, by pooling satisfactory and inadequate performance it becomes less likely for the agent

[37] For a given r, the degree of pooling does not monotonically increase in λ. As discussed at the end of Section 3.1, a higher degree of loss aversion of the agent may help the principal to create incentives. If this is the case, a contract that contains less pooling is preferred from an incentive point of view. If this positive effect of less pooling on incentives outweighs the negative effect on the agent's loss premium, then the optimal contract consists of more distinct wage payments when λ increases. This can happen, however, only locally: at some point the degree of pooling increases in λ.

ex ante to experience a loss, i.e., the loss premium is reduced. Therefore, it is optimal for the principal to use a bonus scheme even when the agent's degree of loss aversion is small.

4. Implementation Problems, Turning a Blind Eye, and Stochastic Contracts

In this section, we relax the assumptions that guaranteed the validity of the first-order approach. In particular, in order to explore the implications of a higher degree of loss aversion, we relax ($A2$). We restrict attention to two simplifications of the former model. First, we return to the assumption of a purely loss-averse agent. Second, only binary measures of performance are considered.

4.1. The Case of a Binary Measure of Performance

As before, the principal cannot observe the agent's action a or whether the benefit B was realized or not. Instead she observes a contractible binary measure of performance, i.e., $\mathcal{S} = \{1, 2\}$. For notational convenience , let $(1-\gamma^H)$ and γ^H denote the probabilities of observing signal $s = 1$ and $s = 2$, respectively, conditional on B being realized. Accordingly, $(1 - \gamma^L)$ and γ^L are the probabilities of observing signal $s = 1$ and $s = 2$, respectively, conditional on B not being realized.[38] Thus, the unconditional probability of observing signal $s = 2$ for a given action a is $\gamma(a) \equiv a\gamma^H + (1 - a)\gamma^L$. Let $\hat{\boldsymbol{\gamma}} = (\gamma^H, \gamma^L)$. We reformulate ($A1$) for the binary case as follows.

Assumption (A4): $1 > \gamma^H > \gamma^L > 0$.

With only two possible signals to be observed, the contract takes the form of a bonus contract: the agent is paid a base wage which yields intrinsic utility u if the bad signal is observed, and he is paid the base wage plus a bonus b resulting in intrinsic utility $u + b$ if the good signal is observed. For now assume that $b \geq 0$.[39] We assume that the agent's intrinsic disutility of effort is a quadratic function, $c(a) = (k/2)a^2$.[40] The agent's expected utility from choosing effort level a then is

$$E[U(a)] = u + \gamma(a)b - \frac{k}{2}a^2 - (\lambda - 1)\gamma(a)(1 - \gamma(a))b. \tag{I.8}$$

[38] In the notation introduced above, we have $\gamma_1^H = 1 - \gamma^H$, $\gamma_2^H = \gamma^H$, $\gamma_1^L = 1 - \gamma^L$ and $\gamma_2^L = \gamma^L$.

[39] The assumption $b \geq 0$ is made only for expositional purposes, the results hold true for $b \in \mathbb{R}$.

[40] This functional form does not fit exactly the assumptions on $c(\cdot)$ that we imposed above, but is made for expositional convenience. Allowing for more general effort cost functions does not qualitatively change the insights that are to be obtained.

As before, the first component is expected net intrinsic utility from choosing effort level a, that is, expected wage payment minus effort cost. The second component is the loss premium, with $\gamma(a)(1-\gamma(a))$ denoting the loss probability.

4.2. Invalidity of the First-Order Approach

The first derivative of expected utility with respect to effort is given by

$$E\left[U'(a)\right] = \underbrace{(\gamma^H - \gamma^L)b\left[2-\lambda+2\gamma(a)(\lambda-1)\right]}_{MB(a)} - \underbrace{ka}_{MC(a)}. \tag{I.9}$$

While the marginal cost, $MC(a)$, obviously is a straight line through the origin with slope k, the marginal benefit, $MB(a)$, also is a positively sloped, linear function of effort a. An increase in b unambiguously makes $MB(a)$ steeper. Letting a_0 denote the intercept of $MB(a)$ with the horizontal axis, we have

$$a_0 = \frac{\lambda - 2 - 2\gamma^L(\lambda-1)}{2(\gamma^H-\gamma^L)(\lambda-1)}.$$

The cases for $a_0 < 0$ and $a_0 > 0$ are depicted in Figures 1 and 2, respectively. Implementation problems in our sense refer to a situation where there are actions

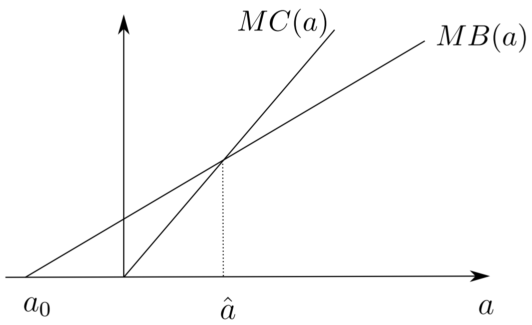

Figure I.1.: $a_0 < 0$.

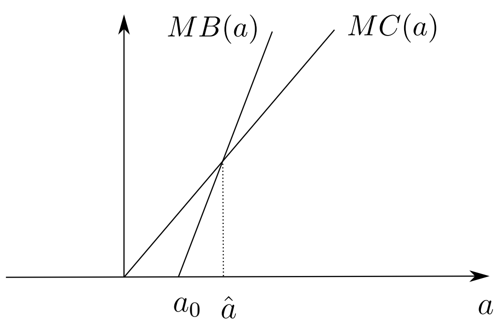

Figure I.2.: $a_0 > 0$.

$a \in (0,1)$ that are not incentive compatible for any bonus payment.

Proposition I.7: *Suppose (A4) holds, then effort level $\hat{a} \in (0,1)$ is implementable if and only if $a_0 \leq 0$.*

Implementation problems arise when $a_0 > 0$, or equivalently, when $\gamma^L < 1/2$ and $\lambda > 2(1-\gamma^L)/(1-2\gamma^L) > 2$. Somewhat surprisingly, this includes performance measures with $\gamma^L < 1/2 < \gamma^H$, which (possibly) are highly informative. Informative in this context means that it is more likely to observe the bad signal if benefit B was not realized, whereas it is more likely to observe the good signal if B was realized. So, why do these implementation problems arise in the first place? Remember that the agent has two targets: First, as in classic models, he seeks to maximize net intrinsic utility, $u + b\gamma(a) - (k/2)a^2$. When the agent cares only about this net intrinsic utility (e.g., he is loss neutral) then each action can be implemented by choosing a sufficiently high bonus. Due to loss aversion, however, the agent has a second target which is minimizing the expected loss. How can the agent pursue this goal? He can do so by choosing an action such that the loss probability, $\gamma(a)(1-\gamma(a))$, becomes small. The crucial point is that these two targets may conflict with each other in the sense that an increase in effort may increase net intrinsic utility but at the same time also increases the loss probability. First of all, note that implementation problems never arise when $\gamma^L \geq 1/2$ or $\lambda \leq 2$. For $\gamma^L \geq 1/2$, the loss probability is strictly decreasing in the agent's action. Consequently, with both targets of the agent being aligned, an increase in the bonus unambiguously leads to an increase in the agent's action. For $\lambda \leq 2$, the weight put on gain-loss utility, $\lambda - 1$, is lower than the weight put on intrinsic utility, so the agent is more interested in maximizing net intrinsic utility than in minimizing the loss probability. For $\gamma^L < 1/2$, on the other hand, implementation problems do arise when λ is sufficiently large. Roughly speaking, being sufficiently loss averse, the agent primarily cares about reducing the loss probability. With the loss probability being inverted U-shaped, the agent achieves this by choosing one of the two extreme actions $a \in \{0, 1\}$. Therefore, the principal cannot motivate the agent to choose an action $\hat{a} \in (0, 1)$ when $\gamma^L < 1/2$ and the agent's loss aversion is sufficiently severe.

4.3. Turning a Blind Eye

As we have seen in the preceding analysis, the principal faces implementation problems whenever $a_0 > 0$. One might wonder if there is a remedy for these implementation problems. The answer is "yes", there is a remedy. The principal can manipulate the signal in her favor by not paying attention to the signal from time to time but nevertheless paying the bonus in these cases. Formally, suppose the principal commits herself to stochastically ignoring the signal with probability $p \in [0, 1)$.[41] Thus, the overall probability of receiving the bonus is given by $\gamma(a; p) \equiv p + (1-p)\gamma(a)$. This strategic ignorance of information gives rise to a transformed performance measure $\hat{\boldsymbol{\gamma}}(p) = (\gamma^H(p), \gamma^L(p))$. As before, $\gamma^H(p)$ denotes the probability that the bonus is paid to the agent conditional on benefit B being realized. Given that B is realized, this happens either when the performance measure is ignored, or – if the principal pays attention to the performance measure – when the good signal is realized. Hence, $\gamma^H(p) = p + (1-p)\gamma^H$. Analogously, the probability of the bonus being paid out conditional on B not being realized is given by $\gamma^L(p) = p + (1-p)\gamma^L$. As it turns

[41] Always ignoring the signal, i.e., setting $p = 1$, would be detrimental for incentives because then the agent's monetary payoff is independent of his action. Hence, he would choose the least cost action $a = 0$. Therefore, we a priori restrict the principal to choose p from the interval $[0, 1)$.

out, ignoring the whole performance measure with probability p is formally equivalent to ignoring only the bad signal with probability p.[42] For this reason, we refer to the principal not paying attention to the performance measure as turning a blind eye on bad performance of the agent. It is readily verified that under the transformed performance measure $\hat{\boldsymbol{\gamma}}(p)$ the intercept of the $MB(a)$ function with the horizontal axis,

$$a_0(p) \equiv \frac{\lambda - 2 - 2\left[p + (1-p)\gamma^L\right](\lambda - 1)}{2(1-p)(\gamma^H - \gamma^L)(\lambda - 1)},$$

not only is decreasing in p but also can be made arbitrarily small, in particular, arbitrarily negative. Formally, $da_0(p)/dp < 0$ and $\lim_{p \to 1} a_0(p) = -\infty$. In the light of Proposition 7 this immediately implies that the principal can eliminate any implementation problems by choosing p sufficiently high, that is, by turning a blind eye sufficiently often.

Besides alleviating possible implementation problems, turning a blind eye can also benefit the principal from a cost perspective. Using the definition of $\gamma(a; p)$ it can be shown that the minimum cost of implementing action $\hat{a}$ under the transformed performance measure, $C(\hat{a}; p)$, takes the following form:

$$C(\hat{a}; p) \;=\; \overline{u} + \frac{k}{2}\hat{a}^2 + \frac{k\hat{a}(\lambda - 1)(1 - \gamma(\hat{a}))}{(\gamma^H - \gamma^L)} \frac{\gamma(\hat{a}) + p(1 - \gamma(\hat{a}))}{1 - (\lambda - 1)\left[1 - 2\gamma(\hat{a}) - 2p(1 - \gamma(\hat{a}))\right]} \tag{I.10}$$

Differentiating the principal's cost with respect to p reveals that $\text{sign}\{dC(\hat{a}; p)/dp\} = \text{sign}\{2 - \lambda\}$. Hence, an increase in the probability of ignoring the bad signal decreases the cost of implementing a certain action if and only if $\lambda > 2$. Hence, whenever the principal turns a blind eye in order to remedy implementation problems, she will do so to the largest possible extent.[43,44] We summarize the preceding analysis in the following proposition.

Proposition I.8: *Suppose the principal can commit herself to stochastic ignorance of the signal. Then each action $\hat{a} \in [0, 1]$ can be implemented. Moreover, the implementation costs are strictly decreasing in p if and only if $\lambda > 2$.*

We restricted the principal to offer non-stochastic payments conditional on which signal is observed. If the principal was able to do just that, then she could remedy

[42] In this latter case, the agent receives the bonus either when the good signal is observed, which happens with probability $\gamma(a)$, or when the bad signal is observed but is ignored, which happens with probability $(1 - \gamma(a))p$. Hence, the overall probability of the bonus being paid out is given by $\gamma(a) + (1 - \gamma(a))p$.

[43] Formally, for $\lambda > 2$, the solution to the principal's problem of choosing the optimal probability to turn a blind eye, p^*, is not well defined because $p^* \to 1$. If the agent is subject to limited liability or there is a cost of ignorance, however, the optimal probability of turning a blind eye is well defined.

[44] This is in the spirit of Becker and Stigler (1974), who show that despite a small detection probability of malfeasance, incentives can be maintained if the punishment is sufficiently severe.

implementation problems by paying the base wage plus a lottery in the case of the bad signal. For instance, when the lottery yields b with probability p and zero otherwise, this is just the same as turning a blind eye. This observation suggests that the principal may benefit from offering a contract that includes randomization, which is in contrast to the finding under conventional risk aversion, see Holmström (1979).[45]

4.4. Blackwell Revisited

We conclude this section by briefly pointing out an interesting implication of the above analysis. Suppose the principal has no access to a randomization device, i.e., turning a blind eye is not possible. Then the above considerations allow a straight-forward comparison of performance measures $\hat{\boldsymbol{\zeta}} = (\zeta^H, \zeta^L)$ and $\hat{\boldsymbol{\gamma}} = (\gamma^H, \gamma^L)$ if $\hat{\boldsymbol{\zeta}}$ is a convex combination of $\hat{\boldsymbol{\gamma}}$ and $\mathbf{1} \equiv (1,1)$.

Corollary I.1: *Let $\hat{\boldsymbol{\zeta}} = p\mathbf{1} + (1-p)\hat{\boldsymbol{\gamma}}$ with $p \in (0,1)$. Then the principal at least weakly prefers performance measure $\hat{\boldsymbol{\zeta}}$ to $\hat{\boldsymbol{\gamma}}$ if and only if $\lambda \geq 2$.*

The finding that the principal prefers the "garbled" performance measure $\hat{\boldsymbol{\zeta}}$ over performance measure $\hat{\boldsymbol{\gamma}}$ is at odds with Blackwell's theorem. To see this, let performance measures $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\zeta}}$ be characterized, respectively, by the stochastic matrices

$$\boldsymbol{P}_\gamma = \begin{pmatrix} 1-\gamma^H & \gamma^H \\ 1-\gamma^L & \gamma^L \end{pmatrix} \quad \text{and} \quad \boldsymbol{P}_\zeta = \begin{pmatrix} 1-\zeta^H & \zeta^H \\ 1-\zeta^L & \zeta^L \end{pmatrix}.$$

According to Blackwell's theorem, any decision maker prefers information system $\hat{\boldsymbol{\gamma}}$ to $\hat{\boldsymbol{\zeta}}$ if and only if there exists a non-negative stochastic matrix $\boldsymbol{M}$ with $\sum_j m_{ij} = 1$ such that $\boldsymbol{P}_\zeta = \boldsymbol{P}_\gamma \boldsymbol{M}$.[46] It is readily verified that this matrix $\boldsymbol{M}$ exists and takes the form

$$\boldsymbol{M} = \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}.$$

Thus, even though comparison of the two performance measures according to Blackwell's theorem implies that the principal should prefer $\hat{\boldsymbol{\gamma}}$ over $\hat{\boldsymbol{\zeta}}$, the principal actually prefers the "garbled" information system $\hat{\boldsymbol{\zeta}}$ over information system $\hat{\boldsymbol{\gamma}}$. While Kim (1995) has already shown that the necessary part of Blackwell's theorem does not hold in the agency model, the sufficiency part was proven to be applicable to the agency

[45] The finding that stochastic contracts may be optimal is not novel to the principal-agent literature. Haller (1985) shows that in the case of a satisficing agent, who wants to achieve certain aspiration levels of income with certain probabilities, randomization may pay for the principal. Moreover, Strausz (2006) finds that deterministic contracts may be suboptimal in a screening context.

[46] See Blackwell (1951, 1953).

framework by Gjesdal (1982).[47] Our findings, however, show that this is not the case anymore when the agent is loss averse.

5. Robustness, Extensions, and Concluding Remarks

In this paper, we explore the implications of reference-dependent preferences on contract design in an otherwise standard model of principal-agency. We find that introducing a loss-averse agent leads to a reduction in the complexity of the optimal contractual arrangement. When loss aversion is the predominant feature of the agent's risk preferences, the optimal contract takes the form of a simple bonus contract even if the agent is also risk averse: some realizations of the performance measure are rewarded with a bonus payment, while others are not. Thus, loss aversion provides a theoretical rationale for bonus contracts, the wide application of which is hard to reconcile with obvious drawbacks – as seasonality effects or insurance fraud – that come along with this particular contractual form.

In the remainder of this section, we consider the robustness of our results. After a brief and semi-formal analysis of an alternative equilibrium concept, we explore the consequences of non quadratic effort costs for implementation problems. Finally, we conclude by discussing diminishing sensitivity of the gain-loss function. Throughout the whole analysis, we adopted the concept of choice-acclimating personal equilibrium (CPE). Kőszegi and Rabin (2006, 2007) provide another concept, called unacclimating personal equilibrium (UPE). The major difference between UPE and CPE is the timing of expectation formation and actual decision making. Under UPE a decision maker first forms his expectations, which determine his reference point, and thereafter, given these expectations, chooses the optimal action. To rule out that people can systematically cheat themselves, for action $\hat{a}$ to be an UPE, it must be optimal for the decision maker to choose $\hat{a}$ given that he expected to do so. In the following, we will argue that applying UPE instead of CPE does not change our main findings. Intuitively, the optimality of simple contracts is rooted in the agent's dislike of being exposed to numerous outcomes, which is reflected in the functional form of his ex ante expected utility. With expectations being met on the equilibrium path under UPE, the expected utility takes the same form under both equilibrium concepts. Thus, one should expect simple contracts to be optimal also under UPE. For the sake of a more formal argument, consider the case of a purely loss-averse agent, i.e., suppose intrinsic utility is linear.

[47] In order to avoid confusion: The necessary part of Blackwell's theorem states that the principal being better off implies that she uses a more informative performance measure. The sufficiency part conversely states that making use of more informative performance measure implies that the principal is better off.

The agent's ex ante expected utility from choosing action a when expecting action $\hat{a}$ is

$$E[U(a|\hat{a})] = \sum_{s=1}^{S} \gamma_s(a) \left[u_s + \sum_{j=1}^{s-1} \gamma_j(\hat{a})(u_s - u_j) - \lambda \sum_{t=s+1}^{S} \gamma_t(\hat{a})(u_t - u_s) \right] - c(a) + \mu(c(\hat{a}) - c(a)) \ .$$

The agent's ex ante expected utility, and in consequence the individual rationality constraint, takes the same form under both equilibrium concepts, CPE and UPE. The incentive compatibility constraint, on the other hand, depends on the applied equilibrium concept. Given the agent expected to choose $\hat{a}$, his marginal utility from choosing a is

$$E[U'(a|\hat{a})] = \sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) u_s + \sum_{s=1}^{S} \sum_{j=1}^{s-1} \gamma_j(\hat{a})(\gamma_s^H - \gamma_s^L)(u_s - u_j) - \lambda \sum_{s=1}^{S} \sum_{j=s+1}^{S} \gamma_j(\hat{a})(\gamma_s^H - \gamma_s^L)(u_j - u_s) - c'(a) + \mu'(c(\hat{a}) - c(a)) \ .$$

Note that either $\mu'(\cdot) = 1$ or $\mu'(\cdot) = \lambda$, depending on whether $\hat{a}$ is greater or lower than a. Even though $E[U(a|\hat{a})]$ is a strictly concave function in the agent's actual action choice a for all values of $\lambda \geq 1$, under UPE there arises the problem of multiplicity of equilibria. More precisely, for a given incentive scheme $\boldsymbol{u}$, there exists a range of actions $a \in [\underline{a}(\boldsymbol{u}), \bar{a}(\boldsymbol{u})]$ all of which constitute a UPE. This problem can be circumvented by assuming that the agent chooses the highest action which constitutes a UPE. In this case there is no need to impose additional assumptions on the cost function or to assume that λ is sufficiently small.[48] By imposing this alternative assumption, the incentive compatibility constraint can be rewritten as

$$\sum_{s=2}^{S} b_s \left\{ \left(\sum_{t=s}^{S} (\gamma_t^H - \gamma_t^L) \right) \left(1 + \sum_{j=1}^{s-1} \gamma_j(\hat{a}) \right) - \lambda \left(\sum_{t=s}^{S} \gamma_t(\hat{a}) \right) \left(\sum_{j=1}^{s-1} (\gamma_j^H - \gamma_j^L) \right) \right\} = 2c'(\hat{a}) \ .$$

Clearly, the incentive compatibility constraint is linear in the additional payments $\boldsymbol{b} = (b_2, \ldots, b_S)$. Thus, our bonus contract result is robust with respect to this change of assumptions.

There is another way to resolve the multiplicity problem under UPE. Kőszegi and Rabin (2006, 2007) define a preferred personal equilibrium (PPE) as a decision maker's

[48] For given expectations $\hat{a}$, let EU_g and EU_l denote the agent's expected utility given that $\mu(x) = x$ and $\mu(x) = \lambda x$, respectively. Both EU_g and EU_l are strictly concave functions, with EU_g achieving its maximum at a strictly higher action than EU_l. EU_g and EU_l intersect at $\hat{a}$. Action $\hat{a}$ is an UPE if it lies between the maximizing actions of EU_g and EU_l. Therefore, expecting to choose the action which maximizes EU_g not only constitutes an UPE, but also is the highest possible UPE.

ex ante favorite plan among those plans he actually will follow through. Put differently, given incentive scheme $\boldsymbol{u}$, the agent chooses the action $a^{PPE} \in [\underline{a}(\boldsymbol{u}), \bar{a}(\boldsymbol{u})]$ that maximizes expected utility among those actions that constitute a UPE. If for all incentive-compatible incentive schemes we have $a^{PPE} \in (\underline{a}(\boldsymbol{u}), \bar{a}(\boldsymbol{u}))$ then PPE and CPE coincide, i.e., a^{PPE} is determined by the first-order condition that characterizes the agent's action under CPE. Thus, by imposing the PPE-analogue of ($A2$) and ($A3$) we can derive results identical to those under CPE. If $a^{PPE} \in \{\underline{a}(\boldsymbol{u}), \bar{a}(\boldsymbol{u})\}$ for all incentive-compatible incentive schemes, the optimal contract also is a bonus contract since both boundary actions are determined by functions linear in $\boldsymbol{b} = (b_2, \ldots, b_S)$.[49] In the intermediate case, however, where $a^{PPE} \in (\underline{a}(\boldsymbol{u}), \bar{a}(\boldsymbol{u}))$ for some incentive-compatible incentive schemes but $a^{PPE} \in \{\underline{a}(\boldsymbol{u}), \bar{a}(\boldsymbol{u})\}$ for others, the optimal contract is not necessarily a bonus scheme.

If the agent's action is characterized by PPE, for all actions $\hat{a} \in (0, 1)$ to be implementable we still need the assumption that λ is not too high. Put differently, similar implementation problems as discussed in Section 4 also arise under PPE. Compared to CPE, however, these implementation problems are less severe. For instance, actions close to zero are always implementable under PPE.

For the discussion of implementation problems in Section 4, we restricted attention to quadratic effort costs. The finding that implementation problems are a potential issue, however, holds true for a wide variety of cost functions. Depending on the particular functional form of the corresponding marginal costs, these implementation problems may be more or less severe. For instance, the result that there are implementation problems if $a_0 > 0$ holds true for all strictly increasing and strictly convex cost functions with $c'(0) = 0$. As for strictly concave marginal costs with $c'(0) = 0$, no action $\hat{a} \in (0, 1)$ is implementable if $a_0 \geq 0$; and even for $a_0 < 0$ there may be actions, in particular actions close to 1, that are not implementable.

Moreover, we kept the whole analysis simple by ignoring diminishing sensitivity, that is, by considering a piece-wise linear gain-loss function. A more general gain-loss function makes the analysis far more complicated: Both the incentive compatibility constraint and the individual rationality constraint are no longer linear functions in the intrinsic utility levels, and thus the Kuhn-Tucker conditions are not necessarily sufficient. Nevertheless, we expect that a reduction in the complexity of the contract will benefit the principal in this case as well. Diminishing sensitivity of the agent's utility implies that the sum of two net losses of two monetary outcomes exceeds the net loss of the sum of these two monetary outcomes. Therefore, in addition to the

[49] The case of $a^{PPE} = \bar{a}(\boldsymbol{u})$ corresponds to the alternative assumption to ($A2$) discussed above. If $a^{PPE} = \underline{a}(\boldsymbol{u})$, on the other hand, then a^{PPE} maximizes EU_l, as defined in the previous footnote.

effects discussed in the paper, under diminishing sensitivity there is another channel through which melting two bonus payments into one "big" bonus affects, and in tendency reduces, the agent's expected net loss. There is, however, an argument running counter to this intuition. As we have shown, loss aversion may help the principal to create incentives. Therefore, setting many different wage payments, and thereby – in a sense – creating many kinks, proximity to which the agent strongly dislikes under diminishing sensitivity, may have favorable incentive effects. Exploring the effects of diminishing sensitivity in a principal-agent relationship with moral hazard is therefore an open question for future research.

II. Performance of Procrastinators: On the Value of Deadlines

Earlier work has shown that procrastination can be explained by quasi-hyperbolic discounting. We present a model of effort choice over time that shifts the focus from completion of to performance on a single task. We find that being aware of the own self-control problems may reduce a person's performance as well as his overall well-being, which is in contrast to the existing literature on procrastination. Extending this framework to a multi-task model, we show that interim deadlines help a quasi-hyperbolic discounter to structure his workload more efficiently, which in turn leads to better performance. Moreover, being restricted by deadlines increases a quasi-hyperbolic discounter's well-being. Thus, we provide a theoretical underpinning for recent empirical evidence and numerous casual observations.

1. INTRODUCTION

Life is pervaded with situations in which people have a certain span of time to work on a task and the final reward depends on how much devotion they put into their work: students studying for a final or writing a thesis, employees working on a long-term project, etc. Next to the final deadline, these tasks often have additional interim deadlines: mandatory problem sets are often a prerequisite to pass a class; students meet in regular intervals with their thesis adviser to report on their progress; employees have to hold several presentations or submit memos at different stages of completion of a project. A rational decision maker with time-consistent preferences would not welcome such restrictions on his choice set. But when people impulsively procrastinate, such interim deadlines can be helpful.[1] Earlier research has shown that one possible explanation for procrastination on the completion of a task is hyperbolic discounting. This paper analyzes the behavior of hyperbolic discounters in a model of effort choice over time that shifts the focus from completion of to performance on a task. We show that interim deadlines are a useful commitment device for a hyperbolic discounter, which increases his "long-run utility". Moreover - and more interestingly - interim deadlines are also performance-enhancing. Thus, while implementing interim deadlines always is in the interest of the hyperbolic discounter himself, these findings suggest that there is also scope for the employer of such an agent to benefit from imposing such deadlines. Therefore, our paper gives a theoretical underpinning for the frequent observation of interim deadlines.

We start out from a model where an individual has a given number of periods to work on a single task. In each period, this person can invest costly effort into this task. Effort is modeled as a continuous decision variable. In the final period the individual receives a reward that depends on the total amount of effort he has invested. Since serious procrastination can hardly be explained by exponential discounting with a reasonable discount factor, we adopt the assumption that the agent discounts quasi-hyperbolically, which gives rise to time-inconsistent preferences.[2] We compare the performance of three types of agents. Next to the benchmark of a time-consistent individual without self-control problems, we consider two types of quasi-hyperbolic discounters: naive persons who are totally unaware of their self-control problems, and sophisticated persons who are fully aware of them. Besides finding that procrastination in general hampers performance, we mainly ask two questions: First, does sophistication increase an individual's performance and overall well-being? Second, do interim

[1] We do not claim that procrastination issues are the only explanation for observing interim deadlines. Other explanations may be preferences for risk diversification or motives for information acquisition.

[2] See O'Donoghue and Rabin (2005) for some illustrative numerical examples.

deadlines enhance performance, and if so, how? The answer to the first question is a novelty in the literature: Earlier work on quasi-hyperbolic discounting has shown that awareness of self-control problems will always benefit a person when costs are immediate and rewards are delayed.[3] We find, in contrast, that sophistication may actually hurt an individual – even in this environment. In order to provide an intuition for why this may be the case, we identify and discuss the effects that drive the differences in the behavior of sophisticated and naive agents. A sophisticated agent realizes that he can create incentives for his future selves to work harder by working only little today. This may lead to an extremely uneven allocation of effort over time, which is undesirable with regard to the agent's long-run preferences. In order to address the second question, we augment the basic model by introducing a second task. Two different regimes are compared: a regime with an interim deadline and a regime without an interim deadline. If no interim deadline is imposed, the agent can work on both tasks up to the final period. Under an interim deadline, on the other hand, he has only half the time to perform on the first task, and the whole span of time to work on the second task. We show that being exposed to a deadline is beneficial for time-inconsistent agents. Interim deadlines help hyperbolic discounters to structure their workload and to allocate their effort more efficiently, leading to an overall performance increase, which in turn improves long-run utility.

Our paper draws on two different strands of literature on time-inconsistent preferences. First, the literature on time-inconsistent procrastination, initiated by Akerlof (1991), and second the literature on time-inconsistent consumption-saving decisions, first studied by Strotz (1956). Earlier work on procrastination assumes that the decision that an individual has to take is *when* to do a task. In general, these papers are interested in the effects of awareness on behavior. O'Donoghue and Rabin (1999b), for example, consider a setting where a single task has to be performed exactly once over a certain span of time. Each period, a person faces the binary decision whether to complete the task or not. They find that being sophisticated with regard to self-control problems leads to an earlier completion of the task. When costs are immediate and rewards are delayed, this in turn implies that sophistication never hurts a person. In O'Donoghue and Rabin (2001b) and O'Donoghue and Rabin (2008), these results are shown to carry over to situations where an individual has to choose which task to perform from a menu of mutually exclusive tasks or where a person engages in long-term projects.[4]

[3]See, for example, O'Donoghue and Rabin (1999b, 2001b, 2008).

[4]O'Donoghue and Rabin (2008) assume that a project requires two periods to be completed, one in which it is started, and a second period in which it is finished. The decision the agent has to take each period, however, remains a binary one.

In the literature on time-inconsistent consumption-saving decisions, which was carried on by Laibson (1996, 1997, 1998), Laibson et al. (1998), Angeletos et al. (2001), and Diamond and Kőszegi (2003), an individual has to decide each period anew how much to consume and how much to save, and thus chooses a continuous decision variable. In this literature, most researchers assume sophisticated beliefs.[5] The analysis of sophisticated quasi-hyperbolic discounters and continuous action spaces is fairly complicated. All the above contributions circumvent the arising analytical problems by assuming that the agent's instantaneous utility function for consumption is of the constant-relative-risk-aversion (CRRA) type. Borrowing the essential framework from this literature, in particular the assumption of a CRRA-utility function and sophisticated beliefs, Fischer (1999) analyzes procrastination issues, showing that sophisticated persons choose a decreasing leisure profile over time. To the best of our knowledge, our paper is the first that provides a detailed comparison of the behavior of naive and sophisticated individuals in a continuous-action-space framework.[6]

Moreover, we analyze the value of interim deadlines as commitment technology. O'Donoghue and Rabin (1999c) analyze optimal incentive schemes when a principal, who faces a cost of delay, hires a time-inconsistent agent, who faces a stochastic task cost, to perform a single task once. They find that under certain circumstances it is optimal to implement a deadline scheme, that is, to fix a date beyond which procrastination is severely punished. While this kind of deadline in a sense compares to the final deadline in our model, our main interest is in the impact of interim deadlines. That interim deadlines may be a valuable commitment mechanism for hyperbolic discounters is conjectured in O'Donoghue and Rabin (2005). We show that this indeed is the case, and, moreover, we lay open the beneficial effect of interim deadlines. With respect to consumption-savings decisions, there is no natural analog to the concept of interim deadlines.[7]

The remainder of the paper is structured as follows: In Section 2, we present the basic single-task model, and briefly review the concept of quasi-hyperbolic discounting and the notions of naiveté and sophistication. This model is analyzed in Section 3. In Section 4, we identify the effects driving the differences in behavior of differently aware agents and discuss the impact of awareness on performance and overall satisfaction.

[5] Diamond and Kőszegi (2003) briefly discuss the behavior of naive agents without comparing sophisticates and naifs. Skiba and Tobacman (2008) identify partially naive hyperbolic discounting as the most consistent explanation for payday borrowing without theoretically analyzing the effects of awareness on behavior.

[6] An exception is Tobacman (2008), who, in a purely technical note, analyzes how consumption depends on the degree of sophistication. An intuitive explanation for the different behavior of differently aware agents or welfare implications, however, are not derived.

[7] In the consumption-savings context, for example, an interim deadline would compare to a Christmas club that allows to deposit money only during the first half of the year.

Section 5 extends the basic model to allow for a meaningful analysis of the effect of deadlines on performance. The final section concludes. All proofs are deferred to Appendix A.2.1. In Appendix A.2.2., we consider the case of a partially naive agent.

2. The Model

An agent has to perform a task, e.g. writing a term paper. He has two periods to work on that task in the sense that in each period $t \in \{1, 2\}$ the agent chooses an effort level $e_t \geq 0$ which he invests in the task. If the agent invests some positive effort in period t then in the same period an effort cost $c(e_t)$ arises. This cost function is assumed to be time-invariant. The agent is rewarded for the task in period 3.[8] This delayed reward, which is assumed to be a function of total effort invested, is denoted by $g(\sum_{t=1}^{2} e_t)$.

Assumption: *It is assumed that the cost function and the reward function satisfy the following properties:* $\forall x > 0$,

$$\begin{array}{llll} c'(x) > 0, & c''(x) > 0, & c(0) = 0, & c'(0) = 0 \\ g'(x) > 0, & g''(x) < 0, & g(0) = 0, & g'(0) > 0 \end{array}$$

To motivate the above functional assumptions, once again consider the example of the student who has to write a term paper. The effort is the time he spends on writing the paper. Thus, the costs of effort are the opportunity costs of not enjoying leisure time. Making the standard assumption of decreasing marginal utility of leisure time is equivalent to assuming a convex cost function. The reward function is the expected grade of the term paper. The expected grade increases when the student spends more time on writing the paper. Typically, by investing somewhat more effort the probability to receive a C instead of a D increases significantly, whereas the increase in effort necessary to receive an A instead of a B is much higher.

Within this framework, we study the behavior of individuals with time-inconsistent preferences due to hyperbolic discounting.[9] In particular, we assume that a person's intertemporal preferences from the perspective of period t are given by

$$U_t(u_t, u_{t+1}, ..., u_T) = u_t + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_\tau,$$

[8] We focus on a three-period model, the shortest possible time horizon that actually generates quasi-hyperbolic discounting effects. For longer time horizons the analysis becomes very quickly very complicated.

[9] Hyperbolic discounting refers to a person discounting events in the near future at a higher discount rate than events in the distant future. For an overview of empirical studies that provide evidence of hyperbolic discounting, see Frederick et al. (2002).

where u_t denotes that person's instantaneous utility in period t. This functional form, which often is referred to as quasi-hyperbolic discounting, captures the essence of hyperbolic discounting.[10] While $\delta \in (0,1]$ represents a time-consistent discount factor, $\beta \in (0,1]$ introduces a time-inconsistent preference for immediate gratification and represents a person's self-control problem: for $\beta < 1$, at any given moment the person has an extra bias for the present over the future.[11] In order to focus on the effects that arise from the present bias embodied in the agent's preferences, we abstract from time-consistent exponential discounting, that is, formally we set $\delta = 1$.

An individual is modeled as a composite of autonomous intertemporal selves. These selves are labeled according to their respective periods of control over the effort decision. During its period of control, self t observes all past effort choices. The current self cannot commit future selves to a particular path of effort decisions. Within this framework, we study three types of agents: time-consistent agents (TC) as a benchmark, and two types of hyperbolic discounters, naifs (N) and sophisticates (S).[12] A naif is completely unaware of future self-control problems and hence wrongly predicts his future behavior: He believes that his future self's preferences will be identical to his current self's, not realizing that as the date of action gets closer his tastes will have changed. A sophisticate, in contrast, is fully aware of his future self-control problems and therefore correctly predicts how he will behave in the future. The first-period intertemporal utility of an agent of type $i \in \{TC, N, S\}$ is given by $U_1^i = -c(e_1) - \beta c(e_2) + \beta g(e_1 + e_2)$. Accordingly, given first-period effort $\hat{e}_1$, the second-period intertemporal utility takes the form $U_2^i = -c(e_2) + \beta g(\hat{e}_1 + e_2)$. The parameter $\beta \in (0,1)$ measures the degree of present bias. For a time-consistent agent we have $\beta = 1$.

Following the literature on present-biased preferences, we assume that agents follow *perception-perfect strategies*, that is, strategies such that in all periods a person chooses the optimal action given her current preferences and her perception of future behavior. In each period, time-consistent and naive agents are just choosing an optimal effort path. While a time-consistent agent will always follow the effort path chosen in the

[10] Throughout this paper, we use the terms "present-biased preferences", "hyperbolic discounting", and "quasi-hyperbolic discounting" interchangeably.

[11] While originally introduced by Phelps and Pollak (1968) to study intergenerational altruism, these present-biased preferences have been reapplied by Laibson (1996, 1997) to study intra-personal, time-inconsistent decision problems. Besides procrastination and consumption-saving decisions, present-biased preferences have been applied to a broad range of contexts of economic interest, for example contract design (DellaVigna and Malmendier (2004, 2006)), industrial organization (Nocke and Peitz (2003), Sarafidis (2005)), bargaining (Akin (2007)), information acquisition (Carrillo and Mariotti (2000), Benabou and Tirole (2000)), and labor economics (DellaVigna and Paserman (2005)).

[12] The two extreme assumptions about awareness, naiveté and sophistication, already have been discussed by Strotz (1956) and Pollak (1968). Though we focus on these two extreme cases of awareness, in the appendix we show that all our results extend to agents who are partially naive in the sense of O'Donoghue and Rabin (2001b).

first period, a naif, in contrast, will often revise his chosen effort path as his preferences change over time. Sophisticates, on the other hand, in a sense play a game against their future selves. Their behavior therefore incorporates reactions to behavior by their future selves that they cannot directly control as well as attempts to strategically manipulate the behavior of their future selves.

3. The Analysis

In this section, we solve the model for the three types of agents: time-consistent individuals, naifs and sophisticates. Hyperbolic discounters have a preference for immediate gratification. As was shown, for instance in O'Donoghue and Rabin (1999b), due to this present bias hyperbolic discounters are prone to procrastinate working on unpleasant tasks. Therefore, in our model with continuous effort choice over several periods, one should expect both naifs and sophisticates to procrastinate in the sense of an increasing effort profile over time. Moreover, compared to a time-consistent agent, both types of hyperbolic discounters perceive immediate effort costs as higher relative to future effort costs and future rewards. Hence, one should expect both types of hyperbolic discounters to exert less effort in total than a time-consistent agent. We begin the analysis with the benchmark case of an agent without self-control problems.

The Time-Consistent Agent Since the preferences of a time-consistent agent do not change over time, his intertemporal decision problem boils down to maximizing lifetime utility, U_1^{TC}, by choosing both first- and second-period effort levels simultaneously. From the corresponding first-order conditions we immediately obtain that a TC chooses the same effort level in both periods. This optimal effort level, e^{TC}, is characterized by

$$c'(e^{TC}) = g'(e^{TC} + e^{TC}) \; . \tag{II.1}$$

Hence, a TC prefers to smooth effort in the sense that in each period he invests the same effort level in the task.[13] This is intuitively plausible: With the cost of effort being a convex function, a time-consistent agent can improve on any uneven allocation of effort over time by keeping total effort - and thus the final reward - constant, but shifting effort from the high-effort period to the low-effort period, thereby reducing total effort costs.

[13]This clearly is an artifact of our choice to abstract from time-consistent discounting. With $\delta < 1$, a time-consistent agent would choose an increasing effort path, as was shown by Fischer (2001).

The Naive Agent A naive agent is unaware that his preferences will change over time. In the first period he believes that his second-period self will have the same preferences, that is, he believes he will stick to the plan he chooses now. When the second period finally rolls around, however, a naif's preferences will have changed.

Definition II.1: *A perception-perfect strategy for a naive agent is given by* $(e_1^N, e_2^N(\hat{e}_1))$ *such that (i)* $(e_1^N, e_2^{TC}) \in \arg\max_{(e_1,e_2)} U_1^N(e_1, e_2)$*, and (ii)* $\forall\, \hat{e}_1 \geq 0$, $e_2^N(\hat{e}_1) \in \arg\max_{e_2} U_2^N(\hat{e}_1, e_2)$*. Let* $e_2^N = e_2^N(e_1^N)$*.*

In the first period a naive agent maximizes U_1^N with respect to e_1 and e_2.[14] The actual first-period effort, e_1^N, and the planned second-period effort, e_2^{TC}, are characterized by the following conditions:

$$g'(e_1^N + e_2^{TC}) = c'(e_2^{TC}) \qquad \text{(II.2)}$$

$$\beta g'(e_1^N + e_2^{TC}) = c'(e_1^N)\ . \qquad \text{(II.3)}$$

Since there is no decision to be made after period 2, beliefs about own future behavior play no further role in determining the second-period effort. Hence, in the second period a naive person maximizes U_2^N with respect to e_2. The corresponding first-order condition which characterizes the second-period effort, e_2^N, is given by

$$\beta g'(e_1^N + e_2^N) = c'(e_2^N)\ . \qquad \text{(II.4)}$$

From equations (II.1)-(II.4) the following result is readily obtained.

Proposition II.1: *(i) A naive agent invests more effort in period 2 than in period 1, i.e.,* $e_1^N < e_2^N$*. (ii) The total effort a naive agent invests is lower than the total effort of a time-consistent person, i.e.,* $e_1^N + e_2^N < 2e^{TC}$*. (iii) A naive agent is overly optimistic when predicting his future-self's willingness to work, i.e.,* $e_2^N < e_2^{TC}$*.*

Parts (i) and (ii) of Proposition 1 state that the two intuitive conjectures made above hold true for naive hyperbolic discounters. According to part (i), a naive agent procrastinates in the beginning and tries to catch up in the end. Part (ii) compares the behavior of a naif and a time-consistent agent. The present bias leads to higher perceived costs for a naif, which makes him exhibit lower overall effort than a time-consistent agent. Moreover, part (iii) says that a naive agent overestimates his own capabilities. Believing that he will behave time-consistently in the future, a naive agent makes ambitious plans today, that he does not follow through tomorrow.

[14] Equivalently, we could solve for the behavior of a time-consistent agent in period 2 for a given first-period effort, $e_2^{TC}(e_1)$. Then, wrongly believing himself to behave time-consistently in the future, in period 1 a naive agent maximizes U_1^N with respect to e_1 subject to $e_2 = e_2^{TC}(e_1)$. We will actually make use of this procedure in the appendix.

The Sophisticated Agent In contrast to a naif, a sophisticate is fully aware that his preferences will change. Therefore, correctly predicting his own future behavior, a sophisticate plays a game against his future self, which can be solved per backwards induction.

Definition II.2: *A perception-perfect strategy for a sophisticated agent is given by* $(e_1^S, e_2^S(\hat{e}_1))$ *such that (i)* $\forall\, \hat{e}_1 \geq 0$, $e_2^S(\hat{e}_1) \in \arg\max_{e_2} U_2^S(\hat{e}_1, e_2)$, *and (ii)* $e_1^S \in \arg\max_{e_1} U_1^S(e_1, e_2^S(e_1))$. *Let* $e_2^S = e_2^S(e_1^S)$.

For a given first period effort level $\hat{e}_1$, in period 2 a sophisticate maximizes U_2^S with respect to e_2. The second-period effort obviously is a function of the first-period effort, $e_2^S(\hat{e}_1)$, and satisfies the corresponding first-order condition,

$$\beta g'(\hat{e}_1 + e_2^S(\hat{e}_1)) = c'(e_2^S(\hat{e}_1)) \; . \tag{II.5}$$

Differentiating (II.5) with respect to e_1 yields

$$\frac{de_2^S(e_1)}{de_1} = -\frac{\beta g''(e_1 + e_2^S(e_1))}{\beta g''(e_1 + e_2^S(e_1)) - c''(e_2^S(e_1))} \in (-1, 0) \; .$$

The above derivative describes how a second-period sophisticate reacts to a change in the first-period effort. A higher first-period effort reduces the second-period effort. Due to the strict convexity of the cost function, however, the absolute value of this reduction is lower than the increase in effort in the first period. In the first period the sophisticate maximizes U_1^S with respect to e_1 subject to $e_2 = e_2^S(e_1)$. In the appendix we show that the effort level that globally maximizes U_1^S, e_1^S, is characterized by the corresponding first-order condition.[15] This first-order condition is given by

$$-c'(e_1^S) + \beta g'\left(e_1^S + e_2^S(e_1^S)\right) + \frac{de_2^S(e_1^S)}{de_1}\beta\left[g'\left(e_1^S + e_2^S(e_1^S)\right) - c'(e_2^S(e_1^S))\right] = 0. \tag{II.6}$$

With the behavior of a sophisticated agent being characterized by (II.5) and (II.6), the following result is obtained.

Proposition II.2: *(i) A sophisticated agent invests more effort in period 2 than in period 1, i.e.,* $e_1^S < e_2^S$. *(ii) The total effort a sophisticated agent invests is lower than the total effort of a time-consistent person, i.e.,* $e_1^S + e_2^S < 2e^{TC}$.

Except for the fact that a sophisticated agent correctly predicts his own future behavior, his behavior otherwise qualitatively parallels that of a naive agent: First, a

[15] While there is not necessarily a unique perception-perfect strategy for a sophisticated agent, all perception-perfect effort pairs are characterized by the corresponding first-order conditions. Multiple perception-perfect strategies are a well-known phenomenon for sophisticated hyperbolic discounters, see for instance O'Donoghue and Rabin (2008).

sophisticated agent procrastinates working on the task in the sense of an increasing effort profile over time.[16] Secondly, with the present bias increasing the perceived cost of effort, in total a sophisticate works less than a time-consistent agent.[17]

4. Comparison of the Naive and the Sophisticated Agent

Having compared the behavior of both types of hyperbolic discounters with the behavior of a time-consistent agent, now we are interested in how naifs and sophisticates compare to each other. Put differently, what effects does awareness of self-control problems have on performance and overall satisfaction? To answer this question a welfare criterion needs to be defined. Following O'Donoghue and Rabin (1999b, 2005) we use people's long-run preferences.

Definition II.3: *A person's long-run preferences are given by* $U_0(e_1, e_2) \equiv -c(e_1) - c(e_2) + g(e_1 + e_2)$.

Long-run preferences reflect a person's preferences when asked from a prior perspective when he has no option to indulge immediate gratification. To formalize this long-run perspective, it is assumed that there is a (fictitious) period 0 where a person has no decision to make.[18] It turns out that comparing first period efforts is sufficient to answer the question who is better off, naifs or sophisticates.

Lemma II.1: *Suppose that* $e_1^i > e_1^j$, *for* $i, j \in \{S, N\}$ *and* $i \neq j$. *Then (i)* $e_2^i < e_2^j$, *(ii)* $e_1^i + e_2^i > e_1^j + e_2^j$, *and (iii)* $U_0^i \geq U_0^j$.

The lemma has a clear intuition. Since there is no decision to be made in the future, awareness plays no role in the second period. Hence, for a given effort level from the first period, both types of hyperbolic discounters face the same problem in period 2. Consequently, the type who works more in the first period works less in the second period. Due to the convexity of the cost function, however, the difference in first-period efforts is larger than the difference in second-period efforts. Thus, the type who invests more effort in the first period, in the end also has the overall better performance. The optimal effort levels from a long-run perspective are those chosen by a TC. While for both types of hyperbolic discounters total effort is below this optimal level of total effort, the type who works more in the first period is closer to the optimal total effort.

[16] A similar result can be found in Fischer (1999) for log utility functions.

[17] Similar results can be found in the consumption-saving literature for sophisticated present-biased consumers, see for instance Laibson (1996).

[18] Another possibility would be to apply the Pareto criterion, where one outcome is deemed better than another if and only if the person views it as better at all points in time. A discussion of these two welfare criteria for hyperbolic discounters is provided in O'Donoghue and Rabin (2005).

Moreover, this total effort is more evenly - and thus, more efficiently - allocated over the two periods. Therefore, the type of hyperbolic discounter who works more in the first period is better of from a long-run perspective.

An intuitive guess would be that a sophisticate, who is aware of his self-control problems, will exhibit a higher first-period effort - and hence a higher total effort - than a naif. This would also be in line with previous research. For instance, O'Donoghue and Rabin (1999b) show that "when costs are immediate, sophisticates do at least as well as naifs (i.e. $U_0^S \geq U_0^N$)" (p.113).[19] While previous research analyzing the effects of awareness solely focuses on models with discrete action spaces, we analyze a continuous action space model. The following simple example demonstrates that the earlier result that sophisticates are always better off than naifs when costs are immediate does not hold true in general.[20]

Example: Let the cost function be $c(e) = (5/3)(1+z)(1/10)^z e^2$ for $e \leq 1/10$, $c(e) = (1/3)e^{1+z} - 1/3(1/10)^{1+z}(1-z)/2$ for $e \in (1/10, 1)$ and $c(e) = (1/6)(1+z)e^2 + 1/3[1-(1/10)^{1+z}(1-z)/2-(1+z)/2]$ for $e \geq 1$. The reward function is given by $g(e_1+e_2) = 2(e_1+e_2) - (1/2)(e_1+e_2)^2$ for $e_1+e_2 \leq 2$ and $g(e_1+e_2) = 2$ otherwise. Suppose that $z = .005$ and $\beta = 1/4$.[21] The optimal effort choices of a sophisticate in the perception-perfect equilibrium are $e_1^S = .02602$ and $e_2^S = .63700$. In contrary, a naif chooses $e_1^N = .03718$ and $e_2^N = .62595$ in the perception-perfect equilibrium. In this example, a naif invests more effort in the task than a sophisticate both in the first period and in total.[22] Hence, a naif is better of than a sophisticate from a welfare point of view, i.e., $U_0^S - U_0^N < 0$. Thus, in contrast to earlier findings, awareness of future self-control problems can hurt the agent even in a model of immediate costs and delayed rewards.[23]

As the above discussion suggests, characterizing the impact of awareness is complicated.

[19] That sophisticates are better off than naifs when costs are immediate is shown in several other papers. O'Donoghue and Rabin (2001b), extend their earlier finding to a setting where a person has to choose which task to perform from a menu of mutually exclusive tasks. Most recently, considering long-term projects, O'Donoghue and Rabin (2008) have shown that in contrast to sophisticates, naifs may start costly projects but then procrastinate finishing these projects, thus never reaping the reward.

[20] That sophistication may hurt a hyperbolic discounter is well known in the literature for models where costs are delayed and rewards are immediate like models of addiction, see O'Donoghue and Rabin (2001a).

[21] While the cost function is continuously differentiable, it is not twice continuously differentiable. Thus, the example does not fit perfectly to our Assumption 1.

[22] A similar finding is obtained by Tobacman (2008) in a consumption-saving framework with CRRA preferences. He shows that current consumption can be decreasing in the degree of naiveté. Welfare implications, however, are not drawn.

[23] While this result may be somewhat counterintuitive, there actually is empirical evidence supporting this suggestion. Wong (2008) finds that time-inconsistency is associated with lower class performance irrespective of awareness. Effects of time-inconsistency on class performance, however, are smaller in magnitude and less statistically significant under naiveté than under sophistication.

Identifying the underlying effects that drive the different behavior of naifs and sophisticates, however, allows us to derive sufficient conditions for a sophisticate exhibiting higher first-period effort than a naif.

Pessimism Effect and Incentive Effect Why does sophistication may not help to increase first-period effort and thereby long-run utility? What are the driving forces behind this observation? O'Donoghue and Rabin (1999a, 2001a) carefully identify two effects how awareness of self-control problems can influence an agent's behavior. First, as O'Donoghue and Rabin (1999a) point out, "sophistication about future self-control problems can make a person pessimistic about future behavior" (p.16). Knowing that - from today's perspective - the future self will not behave optimally, may induce a sophisticate to directly respond to his future shortcomings. Reasoning like "I know that I won't work hard tomorrow, so I'll work more today" probably is familiar to everyone. This is what O'Donoghue and Rabin (1999a, 2001a) call the *pessimism effect*. This, however, is only half the story. Sophistication about one's own self-control problems has a second, less direct effect on today's behavior. Knowing about his own future misbehavior also makes a sophisticate aware of the need and the potential to strategically influence his future behavior via his behavior today. This second channel is labeled *incentive effect* by O'Donoghue and Rabin (1999a, 2001a).[24] So the following question is immediately at hand: How are these effects operative in the model presented in this paper?

A sophisticate in period 1 realizes that he will work less in period 2 than is optimal from today's perspective. He directly responds to his future shortcomings by working more today. Thus, due to the pessimism effect a sophisticate tends to work more in period 1 than a naif.[25] The incentive effect, however, in tendency leads to a lower first-period effort. The first-period self of a sophisticate would like to see his future self invest more effort in the task than he actually does. Since the second-period self increases effort when first-period effort is reduced, the first-period self can create incentives for his future self to work more by working less today. Formally, adding and subtracting $\beta g'(e_1 + e_2^{TC}(e_1))$ from dU_1^S/de_1 yields the following formulation of the marginal utility

[24]The pessimism effect and the incentive effect represent a decomposition of the "sophistication effect" identified by O'Donoghue and Rabin (1999b).

[25]O'Donoghue and Rabin (1999a, 2001a) use the term pessimism effect in models of addictive goods and present-biased preferences. In addictive good models, where rewards are immediate and costs are delayed, the pessimism effect can hurt the agent. In our context, the pessimism effect helps the sophisticate to achieve a better performance than a naif. Thus, in the model of this paper the term pessimism effect is a little bit misleading. Here, it would be more suitable to call this effect "realism effect".

of a sophisticate in period 1:

$$\frac{dU_1^S}{de_1} = \beta g'(e_1 + e_2^{TC}(e_1)) - c'(e_1) + \underbrace{\beta \left[g'(e_1 + e_2^S(e_1)) - g'(e_1 + e_2^{TC}(e_1))\right]}_{PE} + \underbrace{(1-\beta)(de_2^S/de_1)c'\left(e_2^S(e_1)\right)}_{IE},$$

where $e_2^{TC}(e_1)$ is the effort a TC chooses in period 2 for a given first period effort. Note that the first term equals zero for $e_1 = e_1^N$. The second term, PE, is positive and reflects the pessimism effect. The agent knows that his future self chooses $e_2^S(e_1)$ instead of $e_2^{TC}(e_1)$, which would be optimal from today's perspective. The third term, IE, is negative and characterizes the impact of the incentive effect.[26] Given that U_1^S is a quasi-concave function in e_1, then a sophisticate chooses higher effort levels than a naif if the incentive effect does not outweigh the pessimism effect.

At first glance, the two effects seem to be weighted by the present bias parameter β.[27] When having a closer look at the problem, however, it turns out that things are more complicated. When the present bias is low ($\beta \to 1$) then e_2^S is close to e_2^{TC} and there is not much pessimism involved. When the present bias is extreme ($\beta \to 0$) then $de_2^S/de_1 \to 0$ and the agent cannot set incentives for his future self effectively.

With pessimism effect and incentive effect moving in opposite directions, it is complicated to obtain general results concerning the comparison of naive and sophisticated behavior. Nevertheless, using the insights gained from the above discussion we can characterize sufficient conditions for the cost and reward function such that sophisticated agents are better off than naive ones.

Lemma II.2: *Suppose that $c'''(\cdot) \leq 0$ and $g'''(\cdot) \leq 0$. Then a sophisticated agent chooses a strictly higher effort in the first period than a naive agent, i.e., $e_1^S > e_1^N$.*

In the proof of the above lemma we compile sufficient conditions such that the incentive effect never outweighs the pessimism effect. So Lemma 2 states a very intuitive result: given the pessimism effect outweighs the incentive effect, then sophisticates choose higher first-period efforts than naifs.

[26] To be precise, it is not possible to completely disentangle the two effects, because the incentive effect is only operative if the pessimism effect is operative.

[27] For a low degree of present bias the pessimism effect seems to be more important than the incentive effect. The agent cares more about a high reward than delegating work to his future self, and thus works harder today. On the other hand, for a high degree of present bias the incentive effect seems to be more important. The agent's perceived cost in the second period is remarkably lower than his cost today. Thus, the agent prefers to create incentives for his future self to work harder by working less today. And indeed, this is what happens in our example: For a high degree of present-bias, $\beta = 1/4$, sophistication hurts the agent because it makes him work less in the first period than under naiveté. For a low degree of present bias, on the other hand, for instance if $\beta = 3/4$, a sophisticate works more than a naif, and hence is better off . A similar finding is obtained by Gruber and Kőszegi (2001) who analyze the behavior of sophisticates in a model of addictive goods.

Proposition II.3: *Suppose that $c'''(\cdot) \leq 0$ and $g'''(\cdot) \leq 0$. Then the long-run utility of a sophisticated agent is at least as great as the long-run utility of a naive agent, i.e., $U_0^S \geq U_0^N$. Moreover, the performance of a sophisticated agent is strictly higher than the performance of a naive agent, i.e., $e_1^S + e_2^S > e_1^N + e_2^N$.*

5. DEADLINES

In daily life deadlines are an often encountered phenomenon. As an example consider the "good-standing rules" of the Bonn Graduate School of Economics: after a year of coursework, a first paper has to be completed at the end of the second year, a second paper at the end of the third year, and a third paper at the end of the fourth year. A rational decision-maker with time-consistent preferences would not welcome constraints on his choices. But if people impulsively procrastinate, and if they are also aware of their procrastination problems, deadlines can be strategic and reasonable. Perhaps the best empirical demonstration is the study of Ariely and Wertenbroch (2002), which we will discuss in more detail at the end of this section. In this section we ask if and how the behavior of a present-biased agent is affected by the existence of deadlines. Our main finding is that deadlines help an individual to structure his workload more efficiently, which decreases effort costs and in turn improves performance.[28]

A Multi-Task Model To tackle this question we have to modify the simple framework introduced above. While we stick to the case of two periods, we now assume that there are two independent tasks to be undertaken by the agent, task A and task B. We consider two regimes: *deadline* and *no deadline*. When the agent faces no (interim) deadline he is completely free in his decision how to divide his effort on tasks and over time. More precisely, the agent can work in both periods on both tasks. When there is an (interim) deadline, however, the agent can invest effort in task A only in period 1, whereas he can work on task B in both periods.[29] The reward for a task depends on the total effort invested in that task up to its deadline.[30] Effort costs for a particu-

[28] One caveat is in order: While we solely focus on the positive commitment effect of deadlines, flexibility has a strictly positive value if, for instance, future task costs are uncertain. In this case, a deadline is welfare enhancing only if the positive commitment effect outweighs the negative effect due to the reduction in flexibility. See Amador et al. (2006) for a detailed analysis of the tradeoff between commitment and flexibility.

[29] In order to obtain a comparison of the two regimes in terms of the effort level chosen, we introduce a second task which allows us to consider a regime-independent reward scheme. With only one task, the reward under the regime without deadlines would have to be a function of total effort only, whereas the reward under the regime of deadlines would have to be a function of both first-period effort and total effort, making a comparison infeasible.

[30] Our model also encompasses another kind of deadline where task B is handed out after the deadline for task A, as it is typically the case for students' homework assignments. Formally, $e_{B1} = 0$ a priori. Since - and now we are jumping ahead - the agent optimally chooses $e_{B1} = 0$ anyway, this does not impose any additional restrictions and results do not change.

lar period are determined by the sum of efforts invested in both tasks in that period. Formally, let e_{it} denote the effort invested in task $i \in \{A, B\}$ in period $t \in \{1, 2\}$. Moreover, let $e_t = e_{At} + e_{Bt}$ be the total effort that the agent exhibits in period t, and $e_i = e_{i1} + e_{i2}$ be the total effort invested in task i. The reward for task $i \in \{A, B\}$ then is given by $g_i(e_{i1} + e_{i2})$, and the total effort cost in period $t \in \{1, 2\}$ is $c(e_{At} + e_{Bt})$. We assume that the grade function is the same for both tasks, that is, $g_A(\cdot) = g_B(\cdot) = g(\cdot)$. Moreover, we keep the functional assumptions imposed in Section 3. In all that follows, the double-superscript refers to the regime that the agent faces: D for a situation with a deadline, and ND for a situation without a deadline.

The Time-Consistent Agent As a benchmark, consider a time-consistent agent who faces no deadline. In the above language, the intertemporal utility of this agent in period 1 is given by

$$U_1^{TC^{ND}} = -c(e_{A1} + e_{B1}) - c(e_{A2} + e_{B2}) + g(e_{A1} + e_{A2}) + g(e_{B1} + e_{B2}).$$

Choosing $e_{A1}, e_{A2}, e_{B1}, e_{B2}$ in order to maximize this expression yields

$$c'(e_1^{TC^{ND}}) = c'(e_2^{TC^{ND}}) = g'(e_A^{TC^{ND}}) = g'(e_B^{TC^{ND}}). \tag{II.7}$$

It follows immediately that a time-consistent agent equates effort over tasks and smoothes effort over time, that is, $e_A = e_B$ and $e_1 = e_2$. Put differently, when $2e^{TC^{ND}}$ denotes the overall effort that a time-consistent agent invests over the two periods, then he invests $e^{TC^{ND}}$ in the first period and $e^{TC^{ND}}$ in the second period. Moreover, $e^{TC^{ND}}$ is spent on task A and $e^{TC^{ND}}$ is spent on task B. Note, however, that a time-consistent agent does not care about how he splits up his per period effort between the two tasks as long as he invests evenly in both tasks. This implies that being subject to a deadline does not help a time-consistent agent. When investment in task A is possible only in period 1, for a desired overall effort level $2e^{TC^{ND}}$ the time-consistent agent still can choose $e_A^{TC^D} = e_1^{TC^D} = e^{TC^{ND}}$ and $e_B^{TC^D} = e_2^{TC^D} = e^{TC^{ND}}$.

The Sophisticated Agent First consider a sophisticate who faces no deadline. Having two periods of time to work on two tasks is similar to having two periods of time to work on one task. The only additional question is how to divide the total effort on the two tasks. The reward function is identical for both tasks, thus it is optimal to invest half of the total effort in each task. From the single-task exercise we know that a sophisticate has a tendency to work more in period 2 than in period 1. By always working harder in the second period the agent can achieve effort smoothing over tasks in the second period irrespectively of the proportion of first period effort spend on a specific task.[31] This observation allows us to focus on the agent's effort choice over time. With effort being spread out evenly among the two tasks, the optimal second-period effort as a function of first-period effort, $e_2^{SND}(\hat{e}_1)$, is characterized by

$$c'(e_2^{SND}(\hat{e}_1)) = \beta g'((1/2)(\hat{e}_1 + e_2^{SND}(\hat{e}_1))). \tag{II.8}$$

The effort level chosen by a sophisticate in the first period, e_1^{SND}, is determined by the following first-order condition,[32]

$$\beta g'((1/2)(e_1^{SND} + e_2^{SND}(e_1^{SND}))) - c'(e_1^{SND}) + \frac{de_2^{SND}(e_1)}{de_1}\beta\left[g'((1/2)(e_1^{SND} + e_2^{SND}(e_1^{SND}))) - c'(e_2^{SND}(e_1^{SND}))\right] = 0\,. \tag{II.9}$$

Note that the two first-order conditions are very similar to those obtained in the single task case. Recapitulatory, when not facing a deadline, a sophisticated agent equates effort over tasks like a time-consistent agent, but does not achieve effort-smoothing over time, i.e. $e_1^{SND} < e_2^{SND}$ and $e_A = e_B = (1/2)(e_1^{SND} + e_2^{SND})$, where $e_2^{SND} = e_2^{SND}(e_1^{SND})$.

Next, consider a situation where a sophisticated agent faces a deadline in the sense described above: task A is due at the end of the first period, while task B is due at the end of the second period. Put differently, the agent can invest effort in task A only in period 1, whereas he can work for task B in both periods. Formally, $e_{A2} = 0$, $e_A = e_{A1}$ and $e_{B2} = e_2$. For given effort levels $\hat{e}_A$ and $\hat{e}_{B1}$, in the second period the agent's utility is given by

$$U_2^{SD} = -c(e_{B2}) + \beta g(\hat{e}_A) + \beta g(\hat{e}_{B1} + e_{B2})\,.$$

The optimal second-period effort invested in task B as a function of the first-period-effort invested in task B, $e_{B2}^{SD}(\hat{e}_{B1})$, satisfies

$$c'(e_{B2}^{SD}(\hat{e}_{B1})) = \beta g'(\hat{e}_{B1} + e_{B2}^{SD}(\hat{e}_{B1}))\,. \tag{II.10}$$

Differentiation of (II.10) yields

$$\frac{de_{B2}^{SD}(e_{B1})}{de_{B1}} = -\frac{\beta g''(e_{B1} + e_{B2}^{SD}(e_{B1}))}{\beta g''(e_{B1} + e_{B2}^{SD}(e_{B1})) - c''(e_{B2}^{SD}(e_{B1}))} \in (-1, 0)\,.$$

Correctly predicting his own future behavior, in period 1 a sophisticated agent chooses e_A and e_{B1} in order to maximize his intertemporal utility,

$$U_1^{SD} = -c(e_A + e_{B1}) - \beta c(e_{B2}^{SD}(e_{B1})) + \beta g(e_A) + \beta g(e_{B1} + e_{B2}^{SD}(e_{B1}))\,.$$

[31] In the proof of Proposition 4 we formally show, that there is no perception-perfect strategy where the first-period self commits to a first-period effort that high and that unevenly allocated among tasks, such that total second-period effort needed to achieve effort smothing over tasks becomes too costly.

[32] The first-order approach is valid according to the same reasoning as in the single-task case.

This utility maximization problem, however, does not have an interior solution.[33] When facing a deadline, a sophisticated agent considers it optimal to work exclusively on task A in the first period, that is, $e_{B1}^{SD} = 0$. Intuitively, the single-task case and the no-deadline case suggest that a present-biased agent will work harder in the second period. Hence, under a deadline, there is a tendency to invest more effort in task B anyway. But then investing in task B in the first period is not optimal, because, due to decreasing marginal rewards, the agent can benefit from shifting first-period effort from task B to task A. While intuitively plausible4, the formal proof of this statement is somewhat elaborate and therefore deferred to the appendix. The effort levels which are chosen strictly positive, e_A^{SD} and e_{B2}^{SD}, are characterized as follows:

$$c'(e_A^{SD}) = \beta g'(e_A^{SD}) \tag{II.11}$$

$$c'(e_{B2}^{SD}) = \beta g'(e_{B2}^{SD}) \tag{II.12}$$

From (II.11) and (II.12) it follows immediately that $e_A^{SD} = e_{B2}^{SD}$. To sum up: When facing a deadline, a sophisticated agent smoothes effort over time and equates effort over tasks. Moreover, he does not invest in task B in period 1. Let e^{SD} denote the effort level that is chosen under a regime of deadlines in each period and per task. Formally we have $e_1^{SD} = e_A^{SD} = e^{SD}$ and $e_B^{SD} = e_2^{SD} = e^{SD}$.

After all, we are interested in whether deadlines are helpful to overcome self-control problems and thereby to improve performance and the agent's satisfaction. The following proposition compares the behavior and well-being of a sophisticate under both regimes, deadlines and no deadlines.

Proposition II.4: *When facing a deadline, a sophisticated agent chooses a higher effort level in the first period and a higher total effort level than under a regime without a deadline, i.e.,* $e_1^{SND} < e^{SD}$ *and* $e_1^{SND} + e_2^{SND} < 2e^{SD}$. *Moreover, the sophisticated agent is strictly better off from a long-run perspective when facing a deadline, i.e.,* $U_0^{SD} > U_0^{SND}$.

The above proposition has a clear intuition: a deadline helps a sophisticate to better structure his work on the two tasks. He has to complete task A in the first period and therefore he cannot procrastinate finishing task A as he does without a deadline. Thus, the deadline helps the sophisticate to combat procrastination and thereby effort is allocated more efficiently over the two periods. This more efficient allocation reduces effort cost, which in turn leads to a higher overall effort and a better performance. The optimal total effort level from a long-run perspective is the one chosen by a TC. Furthermore, for any total effort level the optimal allocation is investing equal amounts

[33] With interior solution we refer to a pair of first-period effort choices (e_A, e_{B1}) with $0 < e_A, e_{B1} < \infty$.

in both tasks and exhibiting the same amount of effort in each period. Irrespectively of the regime, deadline or no deadline, the total effort a sophisticate invests in the tasks is below the optimal total effort of a TC. With a deadline, however, the level of total effort a sophisticate chooses is closer to a TC's total effort. Moreover, this more desirable level of total effort is more evenly allocated over the two periods. For this reason a sophisticate is better off when being constrained by a deadline.[34]

The Naive Agent Since the analysis for the naive agent is completely analogous to the one of the sophisticated agent for the regime with a deadline and to the single-task case for the regime without a deadline, we defer the formal analysis to the appendix. Here we briefly state the main results and then move on to a discussion of our findings.

When not facing a deadline, a naive agent equates efforts over tasks, but chooses a higher effort level in the second period, that is, $e_1^{NND} < e_2^{NND}$. When being subject to a deadline, a naive agent also equates effort over tasks, but - in contrast - smoothes effort over time. In particular, the first-period effort is spent exclusively on task A and the second-period effort is spent exclusively on task B. Formally, $e_1^{ND} = e_A^{ND} = e^{ND}$ and $e_B^{ND} = e_2^{ND} = e^{ND}$. As a consequence, under a deadline a naive agent achieves a more desirable allocation of his effort, which in turn leads to a higher level of total effort under deadlines. Hence, with the same reasoning as above, a deadline also makes a naive agent better off.

Proposition II.5: *When facing a deadline, a naive agent chooses a higher effort level in the first period and a higher total effort level than under a regime without a deadline, i.e.,* $e_1^{NND} < e^{ND}$ *and* $e_1^{NND} + e_2^{NND} < 2e^{ND}$. *Moreover, from a long-run perspective, being subject to a deadline makes a naive agent strictly better off, i.e.,* $U_0^{ND} > U_0^{NND}$.

One question is immediately at hand: Which type of hyperbolic discounter benefits more from being exposed to an interim deadline? As it turns out, under a deadline sophisticates and naifs choose the same allocation of effort, that is, $e^{SD} = e^{ND}$.[35] Thus, with long-run utility being the same for both types of hyperbolic discounters when facing a deadline, we just have to compare long-run utilities when there are no deadlines in order to answer the question of interest. With effort being evenly distributed over tasks no matter what, the situation without an interim deadline is comparable to the single-task case. Hence, from our earlier findings we know that in general it is undetermined which type of hyperbolic discounter benefits more from

[34] That restrictions on the choice set may help to reduce procrastination is also shown by O'Donoghue and Rabin (2001).

[35] This result, which is an artefact of our model where the agent faces as many deadlines and tasks as periods, is formally established in the proof of Proposition 5.

being exposed to deadlines. When $c'''(\cdot) \leq 0$ and $g'''(\cdot) \leq 0$, however, a naive agent will benefit at least as much from the imposition of a deadline as a sophisticated agent.

Discussion We have shown so far that simple deadlines can help people with self-control problems to improve their performance. The reason is that being exposed to deadlines allows people to allocate their effort more efficiently, which in turn leads to a higher amount of total effort and an overall better performance. Our findings are highly in line with the empirical observations of Ariely and Wertenbroch (2002). They demonstrate the value and effectiveness of deadlines for improving task performance in two different studies both conducted at MIT. In one study participants were "native English speakers [who were given the task to] proofread papers of other students to evaluate writing skills". Participants were randomly assigned to one of three conditions: evenly-spaced deadlines, end-deadline, or self-imposed deadlines.[36] In each condition a participant had to read three texts and payment was contingent on the quality of the proofreading with a penalty for each day of delay.[37] The number of errors correctly detected was highest in the evenly-spaced-deadlines condition, followed by the self-imposed-deadlines condition, with the lowest performance in the end-deadline condition. Moreover, participants were asked to estimate how much time they had spent on each of the three texts. Participants in the evenly-spaced-deadlines condition spent the highest amount of time on each text, followed by the participants of the self-imposed-deadlines condition, while participants of the end-deadline condition have invested the least amount of time. Ariely and Wertenbroch (2002) summarize these observations as follows: "[T]he results show that when deadline constraints increased, performance improved [and] time spend on the task increased" (p.223). These observations are predicted by our theoretical analysis of agents with self-control problems: a deadline increases total effort, which in turn improves performance. In the other study professionals participating in an executive-education course at MIT had the task to write three short papers. Participants were randomly assigned to one of two treatments: no-choice or free-choice. In the no-choice treatment deadlines were fixed and evenly spaced, in the free-choice treatment participants were free to choose the deadlines. In both treatments deadlines were binding and there was a penalty for late submission.[38]

[36] While the evenly-spaced deadlines condition is comparable to our deadline regime, our regime of no deadlines corresponds to the end-deadline condition.

[37] By setting their deadlines as late as possible, the participants would have the most time to work on the texts and the highest flexibility in arranging their workload.

[38] Besides giving the students the most time to work on the papers and the highest flexibility in arranging their workload, by setting their deadlines as late as possible they would also have the opportunity to learn the most about the topic before submitting the papers. Students also had an incentive to set submission dates late because the penalty would be applied only to late submissions and not to early ones. Finally, students who wanted to submit assignments early could privately plan to do so without precommitting to the instructor.

The main finding is that the grade in the no-choice treatment is significantly higher than the grade in the free-choice treatment. This observation also is in line with the theoretical results obtained in this paper.

The focus of the latter study is on self-imposed deadlines and inefficiencies arising due to suboptimal spacing of these deadlines. Even though we do not endogenize the timing of deadlines, our model also captures this result - in a highly stylized way. Let ΔU_0^S denote the long-run utility gain of a sophisticated agent from being exposed to a deadline. Formally, $\Delta U_0^S \equiv U_0^{SD} - U_0^{SND}$. Analogously define $\Delta U_1^S \equiv U_1^{SD} - U_1^{SND}$ to be the utility gain of a sophisticated agent from being exposed to a deadline as perceived from the beginning of the first period. Correctly predicting his future behavior, a sophisticate will always welcome being subject to a deadline in (fictitious) period zero. When asked in period 1, however, a sophisticate is not very enthusiastic about facing a deadline. Formally, $\Delta U_1^S < 0 < \Delta U_0^S$.[39] In period zero, a naive agent considers a deadline neither helpful nor harmful, that is, $\Delta U_0^N = 0$. In period 1, on the other hand, a naive agent considers a deadline an undesirable restriction. Formally we have $\Delta U_1^N < 0$. Thus, while both types of time-inconsistent agents may be willing to accept a deadline long before the task is to be performed, this will not be the case when the task is immediately at hand. Hence, when interpreting "suboptimal spacing of tasks" as not setting deadlines at all, asking present-biased agents too late whether they are willing to accept deadlines or to voluntarily impose deadlines on themselves may lead to agents rejecting this opportunity. Moreover, this finding illustrates what O'Donoghue and Rabin (2005) point out to be general principles when considering "incentives and present bias". Present-biased individuals are sensitive to exactly how decisions are made - e.g. choosing in advance vs. in the moment. When all consequences of a decision are sufficiently far in the future, however, present bias is not a problem and it may be possible to induce better behavior when people are given the opportunity to make decisions now about future behavior.

6. CONCLUSION

Empirical evidence suggests that people have self-control problems, in particular a tendency to procrastinate unpleasant tasks. Former research has shown that this procrastinative behavior can be explained by hyperbolic discounting. The focus of this paper is not on procrastination itself, but on the effects of hyperbolic discounting and awareness of the arising self-control problems on performance. We present a simple model in which an agent has two periods to work on a specific task. His performance

[39]This result is readily established by a simple revealed-preference argument.

depends on the total effort invested. We find that self-control problems reduce performance. Moreover, sophistication about one's own self-control problems not necessarily leads to better performance than naiveté.

In a next step, in a slightly augmented version of the basic model, we analyze the value and effectiveness of interim deadlines as commitment device. In line with recent empirical evidence we find that interim deadlines improve performance when individuals impulsively procrastinate. This improvement of performance, which makes a present-biased agent better off from a welfare point of view, is based on a more favorable allocation of effort. The restrictions imposed by deadlines help an agent to better structure his workload, which in turn leads to lower effort costs and an overall higher effort level. These results are of interest not only because they provide a theoretical underpinning of recent empirical work, but also because they explain many types of deadlines encountered in daily life. To get back to one of the examples that we have mentioned so far: Deadlines implemented by the "good-standing" rules of graduate schools make grad students work focused on each of their papers, finishing a paper thoroughly before starting another one, thereby improving chances to write high-quality papers. Without these deadlines, grad students cannot commit themselves to work in their last year in school exclusively on the their final paper. Instead, they possibly will end up spending effort on - perhaps unfinished - older papers, resulting in a bunch of low-quality papers that are finished in a hurry and written sloppy.

The model of this paper is simple in the sense that we consider the shortest possible time horizon that actually generates quasi-hyperbolic discounting effects. Without imposing further assumptions on cost and reward functions, analyzing a longer time horizon in a continuous action space framework, in particular the analysis of the behavior of sophisticated individuals, becomes very complicated very quickly. In the literature the arising complications are sidestepped by assuming instantaneous utility functions of the CRRA type. Facing the trade off between the analysis of a longer time horizon on the one hand, and less restrictive functional assumptions on the other hand, we opted for the latter. We think, however, that the main insights are to be obtained in our model.

Last, throughout the paper we focused on two extreme cases of awareness, total naiveté and full sophistication. As we show in Appendix A.2.2., the behavior of a partially naive person is somewhere between these two extremes. In consequence, with both extreme types of hyperbolic discounters benefiting from the presence of interim deadlines, it is little surprising that this result carries over to the case of partially naive individuals.

III. Relaxing Competition Through Quality and Tariff Differentiation

In this chapter, I compare two-part tariff competition to linear pricing in a vertically differentiated duopoly. Consumers have identical tastes for quality but differ in their preferences for quantity. The main finding is that quality differentiation occurs in equilibrium if and only if two-part tariffs are feasible. Furthermore, two-part tariff competition encourages entry, which in turn increases welfare. Nevertheless, two-part tariff competition decreases consumers' surplus compared to linear pricing.

1. Introduction

In many markets different firms sell products of different qualities. What are firms' strategic incentives for quality differentiation? In a seminal contribution, Shaked and Sutton (1982) show that firms can relax price competition via quality differentiation. This is possible due to heterogeneity of consumers with regard to their tastes for quality. In Shaked and Sutton, if consumers do not differ in their valuations for quality, firms have no strategic incentives for quality differentiation. This paper, in contrast, shows that firms may have incentives for quality differentiation even when consumers do not differ in their tastes for quality but differ in their preferences for quantity. It is shown that quality differentiation can relax competition when firms can offer two-part tariffs. If firms are restricted to linear pricing, however, quality differentiation does not relax competition.[1] The intuition is that selecting different qualities may facilitate market segmentation in which different classes of consumers are also served different quantities. Without quality differentiation the firms produce perfect substitutes and thus compete in a Bertrand fashion. Thus, the novel contribution of this paper to the theory of product differentiation is to identify differences in consumers' preferences for quantity as a reason for strategic quality differentiation by firms.[2]

The presented model is similar to the well-known Shaked and Sutton unit-demand model, or rather the Choi and Shin (1992) version with quasi-linear utility functions. The differences are that in this paper consumers have continuous demand and that next to linear pricing two-part tariff competition is considered. The analysis presented in this paper is based on a non-cooperative three-stage game. In the first stage, the two potential duopolists decide whether to enter the market. After observing the entry decisions, each firm in the market chooses a quality level for its product. Finally, observing entry and quality decisions, firms select a two-part tariff (or a linear price schedule). At each stage, firms act simultaneously and independently.

The main finding is that product differentiation occurs in equilibrium if and only if two-part tariffs are feasible. One example for quality differentiation and two-part tariffs is the health club industry: Health clubs often levy a membership fee plus a per-use charge and they offer a fixed quality (equipment and service). In many cities there is more than one health club in the city center (no spatial differentiation), and in most cases these clubs differ in quality.[3] Moreover, the implications of the feasibility of

[1] Linear pricing means that the tariff is a linear function of the quantity q, hence a linear tariff has the following form: $T(q) = p \cdot q$. Note, a two-part tariff $T(q) = A + p \cdot q$ is an affine function.

[2] That a higher pricing flexibility may increase the degree of product differentiation is not novel to the literature. Peitz and Valletti (2008) show for horizontally differentiated media platforms that content differentiation is higher if platforms can charge prices from viewers compared to the case where platforms are restricted to offer their service free to consumers.

[3] Other examples for two-part tariffs and vertical differentiation are discotheques, movie theaters

two-part tariffs on welfare, consumers' surplus, and industry profits are investigated. I show that welfare and industry profits are higher and consumers' surplus is lower if two-part tariffs are feasible. If two-part tariff competition is feasible, competition is relaxed and thus both firms enter the market, which in turn increases welfare.

The structure of the paper is as follows: After a brief review of the related literature, Section 2 describes the model and discusses the linear pricing case. In Section 3, the model with two-part tariff competition is solved by backwards induction. Section 4 compares the results of two-part tariff competition and linear pricing. Section 5 present two simple applications of the model to the theory of vertical restraints. The final section summarizes the main findings.

Related Literature: Nonlinear pricing often is observed in oligopolistic markets.[4] With several notable exceptions, the existing literature on nonlinear pricing focuses on the

monopoly problem (cf. Wilson (1992) for monopoly pricing). There exist only few papers on second-degree price discrimination or nonlinear pricing in oligopoly.[5] A classic paper on this topic is Katz (1984). Katz analyzes an economy with informed high demand consumers and uninformed low demand consumers. Informed consumers purchase from the cheapest store while uninformed consumers choose a store at random. In equilibrium, the firms choose tariffs to separate these two groups. Spulber (1989) studies nonlinear pricing in a free-entry circular-city model. He shows that nonlinear pricing leads to greater product variety than linear pricing. A seminal contribution to the literature on nonlinear pricing in oligopoly is Armstrong and Vickers (2001), who study a general framework with spatially differentiated firms that compete in nonlinear tariffs.[6] They show that under certain conditions firms choose welfare optimal two-part tariffs in equilibrium. Nonlinear pricing in spatial competition models is also analyzed by Stole (1995) and Rochet and Stole (2002).[7] The approaches of Stole and Rochet and

and several types of clubs. The main difference between movie theaters is the size of the silver screen. The ticket price is a fixed fee and the price for popcorn can be interpreted as the quantity depending part. As for discotheques, there typically is a fixed entrance fee whereas drinks are charged per bottles bought. Movie theaters and discotheques, respectively, are often located close to a rival.

[4]There is no commonly accepted definition of nonlinear pricing. Following Wilson (1992, p.5), I denote a tariff as nonlinear if the average charge is a function of the purchased quantity. In unit-demand models where firms offer various pairs of quality and price, however, these offers are often denoted as nonlinear pricing function. Based on the second definition, nonlinear pricing in a vertically differentiated duopoly is also analyzed by Champsaur and Rochet (1989) and Johnson and Myatt (2003).

[5]Pigou (1920) considers three kinds of price discrimination: first-degree price discrimination is perfect price discrimination, second degree price discrimination is discrimination across quantities and for third degree price discrimination the prices differ for distinguishable consumers.

[6]Armstrong and Vickers (2001) also analyze third-degree price discrimination.

[7]A similar framework is studied by Desai (2001) but with a different focus.

Stole are highly related to Armstrong and Vickers if the quality is interpreted as quantity. In these models consumers have unit-demand and firms discriminate via different quality-levels. Since quality and quantity have similar properties, these approaches can be reinterpreted as nonlinear pricing. A logit demand model with two-part tariffs is analyzed by Yin (2004).

In contrast to the articles mentioned so far, I analyze nonlinear pricing in a model of vertical rather than horizontal differentiation.[8] In a classic contribution on vertical differentiation, Gabszewicz and Thisse (1979) analyze a price equilibrium of an oligopoly game. Consumers differ in income m and obtain utility $U = s \cdot (m - p)$ when they buy quality s at price p. The qualities of the firms are fixed exogenously in this model. Shaked and Sutton (1982) extend the Gabszewicz-Thisse model by endogenizing quality levels. The main result is that in equilibrium firms produce distinct qualities and thereby relax price competition. Tirole (1988) shows robustness of these earlier results for the case of the Mussa-Rosen utility function, i.e., for the case in which $U = \theta \cdot s - p$, where θ denotes the consumer's type.[9] While Tirole focuses on parameter values such that the market is fully covered in equilibrium, Choi and Shin (1992) analyze the model when the market is not covered. A complete characterization of quality choices in a duopoly model in which consumers have a Mussa-Rosen utility function is given by Wauthy (1996). All these models assume that consumers have unit-demand. In contrast, I analyze the effects of quantity discounts in a vertically differentiated duopoly, which cannot be captured by one of the utility functions mentioned above. Consequently, I introduce a novel tractable utility function for the framework with vertically differentiated firms and multi-unit demand.

2. DESCRIPTION OF THE MODEL

There are two potential firms ($i = 1, 2$) producing (distinct) substitute goods. The two firms play a non-cooperative three-stage game.[10] At the first stage, they decide independently and simultaneously whether or not to enter the market. In case of entry, a firm incurs fixed cost $K > 0$. At stage two, each firm observes whether its rival has entered the market. Thereafter, both firms independently and simultaneously choose their respective quality level $s_i \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$.[11] This stage will be referred to as the

[8] Quality can be seen as a "vertical" product feature in the sense that each consumer prefers higher quality. The subdivision in vertical and horizontal differentiation is due to Lancaster (1979).

[9] Mussa and Rosen (1978) characterize the optimal price-quality schedule for a monopolist.

[10] This three stage game is similar to the one considered by Shaked and Sutton (1982) for their unit-demand approach.

[11] The presented analysis can be generalized to quality-levels $s_i \in [0, 1]$. The existence of tariff game equilibria, however, then is intricate to show.

quality game. At the third stage, both firms independently and simultaneously choose a tariff, having observed the rival's quality level. Each firm i chooses a two-part tariff: $T_i(q) = A_i + p_i \cdot q$.[12] Here, q denotes quantity, $p_i \geq 0$ is the price per unit, and $A_i \geq 0$ a fixed fee. The third stage will be called tariff game. If price discrimination is banned or infeasible, then $A_i = 0$. The focus of this paper is on firms' strategic incentives for product differentiation. To separate this effect, I assume that costs are zero for all quality levels. Put differently, there is only one reason for a firm not to produce the best possible quality, namely to relax tariff competition on the final stage.

There exists a continuum of consumers with measure one. I assume that consumers practice "one-stop shopping". That is, consumers make all their purchases from one firm. Thus, each consumer takes one of the following three actions: purchase from firm 1, purchase from firm 2, or do not purchase at all. Consumers differ in a taste parameter θ that is uniformly distributed on the unit-interval. It is assumed that consumers have quasi-linear utility functions and the reservation utility is normalized to zero. I assume that the utility of a consumer with type θ who purchases q units from firm i takes the following specific form[13]

$$U = s_i q - q^2/(2\theta) - T_i(q) \qquad \theta \in (0, 1] \; . \tag{III.1}$$

Consumers have the same tastes for quality, but differ in their preferences for quantity. Put differently, the marginal rate of substitution between quality and money is independent of the consumer's type, i.e., $\partial^2 U/\partial s_i \partial\theta = 0$. The type parameter θ determines a consumer's satiation point. With this utility function, an improved quality is more beneficial to large buyers. For instance, a frequent health club visitor who works out before going to work suffers more from a low quality, say a dirty shower room, than a health club member who sporadically goes to the gym in his spare time. Moreover, quality is not additive in the sense that being member in two low-quality health clubs, one with clean showers and the other with modern training equipment, is not equivalent to being member in a high-quality health club that offers both. From the above utility function I can derive a consumer's conditional demand function. When a consumer of type θ buys from firm i, his demand is given by:

$$q_i(p_i, \theta) = \theta(s_i - p_i) \qquad \text{for } p_i \leq s_i, \tag{III.2}$$

and zero otherwise. Inserting the demand into the utility function and ignoring the

[12] It is assumed that general nonlinear tariffs are infeasible.

[13] Granted this utility function is arbitrary, but it leads to a tractable linear demand function with reasonable properties. Assuming particular functional forms for consumers' utility is not uncommon in the literature on vertical differentiation. Jensen (2008) analyzes the problem of a monopolist who offers a menu of two-part tariffs and qualities. She also assumes a simple quadratic utility function in quantity.

fixed fee yields the surplus function

$$v_i(p_i, \theta) = (1/2)\theta(s_i - p_i)^2 \qquad \text{for } p_i \leq s_i, \tag{III.3}$$

and zero otherwise. This surplus is the maximum net utility a consumer of type θ can receive excluding a potential fixed fee payment when purchasing from firm i at marginal price p_i. The type parameter θ is a simple multiplier in the consumer's surplus function. A consumer's surplus is increasing in the quality he consumes and decreasing in the marginal price he has to pay. The consumer's surplus is a weighted quadratic function of the product's "net value" $(s_i - p_i)$. Since a large buyer does benefit more as quality improves and since demand is increasing in the consumer's type, the increase in indirect utility due to an improved quality is larger for a consumer with a higher type, i.e., $\partial^2 v_i / \partial s_i \partial\theta > 0$. If firms practice marginal cost pricing ($p_i = 0$), then the indirect utility of a consumer with type θ is given by $V = \frac{1}{2}\theta \cdot s_i^2 - A_i$.[14] It is assumed that consumers have full information about the tariffs and the quality levels in the market. Consumers' tastes are private information, only the distribution is known by the firms. If a consumer is indifferent between firm 1 and firm 2, he purchases the higher-quality product.[15] If quality levels are equal, the consumer chooses a store at random. The equilibrium concept employed is subgame perfect Nash equilibrium in pure strategies. Thus, the game is solved by backwards induction.

2.1. Linear Pricing

Suppose price discrimination is infeasible, that is, firms are restricted to offer linear prices. It is assumed, without loss of generality, that firm 2 is the high-quality supplier and that firm 1 is the low-quality supplier, i.e., $s_1 \leq s_2$. Given that both firms are active in the market, it follows immediately from consumers' surplus functions that all consumers purchase from the firm that offers the product with the higher "net value" $(s_i - p_i)$. In contrast to former models of vertical differentiation, like Shaked and Sutton (1982), where consumers with different types have different rankings for price-quality pairs, here all consumers have the same ranking for price-quality pairs. Thus, firms cannot relax price competition via quality differentiation.

Lemma III.1: *Suppose firms are restricted to linear prices and that both firms have entered the market. Then it is impossible that both firms realize strictly positive profits.*

[14]In this case the consumer's indirect utility function is similar to the well-known Mussa and Rosen (1978) utility function for unit demand models with distinct qualities, where $U = \theta \cdot s - p$. The Mussa and Rosen utility function is used in various models of vertically differentiated markets, for instance Choi and Shin (1992), or in an augmented version in Rochet and Stole (2002).

[15]This assumption is not crucial, however, it simplifies some proofs.

Proof: All proofs are given in the appendix.

The outcome of the two-stage game (quality game and tariff game) when only linear pricing is feasible, depends on the behavior of firm 1. When firm 1 chooses a quality $s_1 < s_2 \leq 1$, then in equilibrium firm 1 has no positive market share and consequently zero profits. On the other hand, if firm 1 chooses $s_1 = s_2$, there is perfect competition and $\pi_1 = \pi_2 = 0$. In either case firm 1 earns zero profits. Consequently, when price discrimination is infeasible at most one firm enters.

3. Two-part Tariff Competition

Now suppose firms can use two-part tariffs. This section shows that when two-part tariff competition is feasible firms can relax price competition via tariff and quality differentiation.

3.1. Preliminary Remarks on the Quality Game

In this subsection, I establish that when both firms are active in the market, they produce distinct quality levels.

Lemma III.2: *Suppose that both firms produce the same quality and two-part tariffs are feasible. Then in the unique tariff game equilibrium both firms use the cost-based linear tariff $T^* = 0 \cdot q$ and earn zero profits.*

The intuition is similar to the reasoning behind the well-known Bertrand paradox. Assume, for a sake of contradiction, that both firms produce the same quality and at least one firm makes positive profits. Then the firm with lower profits can increase its profits by slightly undercutting the rival's tariff.[16] Firm i undercutting firm j's tariff means that in an expenditure-quantity diagram the tariff of firm i is completely below the tariff of firm j. This logic is still true when the firms have equal positive profits. When firm i slightly undercuts firm j's tariff, firm i obtains all customers of firm j and almost always keeps some of its former customers. Consequently, slightly undercutting the rival's tariff increases profits. Hence, for equal qualities I obtain that the well-known Bertrand result also holds when two-part tariffs are feasible.

Without quality differentiation price discrimination is infeasible and perfect competition leads to zero profits for both firms. Therefore, due to the presence of entry costs, I obtain the following result for the subgame perfect equilibrium.

Lemma III.3: *In any subgame perfect equilibrium in which both firms enter, the firms produce distinct quality levels.*

[16] The tariff T_i undercuts tariff T_j if $\forall q \;\; p_i q + A_i < p_j q + A_j$.

This result extends Proposition 1 in Shaked and Sutton (1982) to the multi-unit approach with two-part tariffs.

3.2. The Tariff Game

Suppose that $s_2 > s_1$, so that there is a high-quality supplier (firm 2) and a low-quality supplier (firm 1). The net surplus of a consumer of type θ, given the tariffs of the two firms, depending on the consumer's quality decision is:

$$V(\theta) = \begin{cases} \frac{1}{2}\theta(s_2 - p_2)^2 - A_2 & \text{if he buys from firm 2 (high-quality),} \\ \frac{1}{2}\theta(s_1 - p_1)^2 - A_1 & \text{if he buys from firm 1 (low-quality),} \\ 0 & \text{otherwise.} \end{cases} \tag{III.4}$$

Due to the following lemma, one can focus on the case where the high-quality firm serves the consumers with high valuations for the good.[17]

Lemma III.4: *There does not exist a pure strategy Nash equilibrium of the tariff game with (i) the low-quality firm selling to consumers with high valuations* $\theta \in [\tilde{\theta}, 1]$, *and (ii) the high-quality firm selling to consumers with relatively low valuations* $\theta \in [\hat{\theta}, \tilde{\theta}]$, *where* $0 \leq \hat{\theta} \leq \tilde{\theta} < 1$.

According to Lemma 4, if both firms share the market in equilibrium, then consumers with relatively strong tastes for the product buy from the high-quality firm. For "middle-type" consumers, the high-quality firm is too expensive, hence they purchase the low-quality product at a (very) cheap tariff. Consumers with relatively low tastes do not purchase at all. I will call consumers who are indifferent between two options "marginal consumers". Therefore, in the economy there exist two kinds of marginal consumers: one is indifferent between buying from firm 1 and firm 2, whereas the other marginal consumer is indifferent between buying from firm 1 and not buying at all (see Figure 1). The first type of marginal consumer will be denoted by $\tilde{\theta}$, the latter type by $\hat{\theta}$.

This market structure implies that $A_2 > A_1$ and $(s_2 - p_2) > (s_1 - p_1)$ in equilibrium, see Figure 1.[18] From the definitions of the marginal consumers, one immediately obtains the following characterization of the fixed fees charged by the two firms in an

[17] Readers should be aware that the proof of Lemma 4 requires some results presented in Section 3 later on. Therefore, I recommend to postpone reading the proof to the end of Section 3.

[18] For the relations of the fixed fees and the products' net values one can distinguish four cases. Two of these cases are ruled out by Lemma 4 as equilibrium candidates. The first of the two remaining cases is depicted in Figure 1, whereas the second case, in which firm 2 serves the whole market, is analyzed later on.

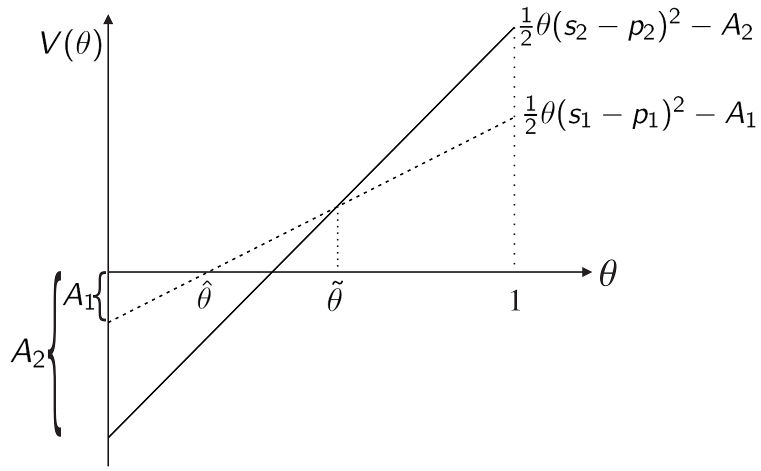

Figure III.1.: Optimal choices for different consumer types.

equilibrium in which both firms share the market:

$$A_1 = \frac{1}{2}\hat{\theta}(s_1 - p_1)^2 \;, \tag{III.5}$$

$$A_2 = \frac{1}{2}\tilde{\theta}\left[(s_2 - p_2)^2 - (s_1 - p_1)^2\right] + A_1 \;. \tag{III.6}$$

Note that given p_i and the rival's tariff, the fixed fee A_i is uniquely determined by $\tilde{\theta}$. Put differently, the choice of the fixed fee is equivalent to choosing the marginal consumer $\tilde{\theta}$.[19] Each firm i chooses $(p_i, \tilde{\theta})$ to maximize profits. The firms' profit functions, for a given tariff of the competitor, are[20]

$$\pi_2(p_2, \tilde{\theta}) = (1 - \tilde{\theta}) \cdot A_2(p_2, \tilde{\theta}) + p_2 \cdot \int_{\tilde{\theta}}^{1} q_2(p_2, \theta)\, d\theta \;, \tag{III.7}$$

$$\pi_1(p_1, \tilde{\theta}) = \left(\tilde{\theta} - \hat{\theta}(\tilde{\theta}, p_1)\right) \cdot A_1(p_1, \tilde{\theta}) + p_1 \cdot \int_{\hat{\theta}(\tilde{\theta}, p_1)}^{\tilde{\theta}} q_1(p_1, \theta)\, d\theta \;. \tag{III.8}$$

For now, assume that firms' maximization problems have interior solutions. In the appendix it is shown that this indeed is the case.

Profit Maximization Problem of Firm 2 (High-Quality Supplier)

More specifically, the profit function of the high quality firm is given by

$$\pi_2(p_2, \tilde{\theta}) = (1 - \tilde{\theta}) \cdot A_2 + \frac{1}{2} p_2 (s_2 - p_2)(1 - \tilde{\theta}^2) \tag{III.9}$$

[19] Note that in equilibrium the marginal consumers $\tilde{\theta}$ chosen by the two firms are the same.

[20] More precisely, $A_i = A_i(p_i, p_j, A_j, \tilde{\theta})$ for $i \neq j$, but I suppress in the following the rival's tariff parameters.

where

$$A_2 = \frac{1}{2}\tilde{\theta}\left[(s_2 - p_2)^2 - (s_1 - p_1)^2\right] + A_1 \ .$$

First, setting the partial derivative of π_2 with respect to p_2 equal to zero allows to solve for the optimal marginal price of firm 2,

$$p_2^* = \frac{1}{2}(1 - \tilde{\theta})s_2 \ . \tag{III.10}$$

The optimal marginal price p_2^* is determined by the marginal consumer $\tilde{\theta}$. Hence, p_2^* depends only indirectly on rival's tariff via $\tilde{\theta}$. Obviously, p_2^* decreases in $\tilde{\theta}$. Intuitively, a greater market share is accompanied by a lower fixed fee, i.e., less of the surplus of the served consumers can be extracted by the fixed fee. This in turn leads to a raise in the optimal marginal price. Note that the optimal marginal price exceeds marginal cost. This result is in contrast to several models of horizontal differentiation, where in equilibrium marginal prices equal marginal costs. For instance, Armstrong and Vickers (2001) show for spatially differentiated markets that the firms offer cost-based two-part tariffs in equilibrium. On the other hand, Yin (2004) points out that in the context of a Hotelling model, marginal prices are higher than marginal costs if the transportation costs are shipping costs.[21] The observation that in the model of Armstrong and Vickers consumers' types do not interact with quantity, whereas in Yin's model and the one presented here there is an interaction between consumers' types and quantity leads me to the following conclusion: regardless of the differentiation framework, marginal prices exceed marginal costs if for a given marginal price consumers with different types prefer different quantities.

The first-order condition for profit maximization of firm 2 with respect to $\tilde{\theta}$ is given by

$$\frac{\partial \pi_2}{\partial \tilde{\theta}} = 0 \iff -A_2 + (1 - \tilde{\theta})\frac{1}{2}\left[(s_2 - p_2)^2 - (s_1 - p_1)^2\right] - p_2(s_2 - p_2)\tilde{\theta} = 0 \ . \tag{III.11}$$

Inserting the optimal marginal price, p_2^*, into (III.11) and rearranging yields

$$A_2^* = \frac{1}{2} \cdot (1 - \tilde{\theta})\left[\frac{1}{4}s_2^2(1 - \tilde{\theta}^2) - (s_1 - p_1)^2\right] \ . \tag{III.12}$$

Profit Maximization Problem of Firm 1 (Low-Quality Supplier)

The profit function and the optimization constraints of firm 1 are given by

$$\pi_1(p_1, \tilde{\theta}) = (\tilde{\theta} - \hat{\theta}) \cdot A_1 + \frac{1}{2}p_1(s_1 - p_1)(\tilde{\theta}^2 - \hat{\theta}^2) \ , \tag{III.13}$$

[21] Anderson and Engers (1994) describe two types of transportation costs. A shipping cost depends on the quantity which is "shipped" and a shopping cost is independent of the amount purchased.

where

$$\hat{\theta} = \tilde{\theta}\left[1 - \left(\frac{s_2 - p_2}{s_1 - p_1}\right)^2\right] + \frac{2A_2}{(s_1 - p_1)^2}$$
$$A_1 = \frac{1}{2}\tilde{\theta}\left[(s_1 - p_1)^2 - (s_2 - p_2)^2\right] + A_2 .$$

Setting the partial derivative of π_1 with respect to p_1 equal to zero yields an implicit condition for the optimal marginal price p_1^*:

$$\frac{\partial \pi_1}{\partial p_1} = 0 \quad \Longleftrightarrow \quad 3\hat{\theta} = \tilde{\theta} . \tag{III.14}$$

In equilibrium the low-quality firm serves $\frac{2}{3}$ of the residual demand. Equation (III.14) can be rewritten as

$$p_1^* = s_1 - \sqrt{\frac{3}{\tilde{\theta}}\left(\frac{1}{2}\tilde{\theta}(s_2 - p_2)^2 - A_2\right)} .$$

The optimal marginal price of firm 1 is higher than marginal costs if the net surplus of the marginal consumer $\tilde{\theta}$ is sufficiently small. The net surplus of marginal consumer $\tilde{\theta}$ is quite small if competition between the two firms is not very intense. Hence, when the products of the two firms are sufficiently differentiated the marginal price of firm 1 is positive. The other first-order condition of firm 1 is obtained by setting the partial derivative of the profit function with respect to $\tilde{\theta}$ equal to zero. Rewriting this equation and inserting p_1^* yields the following formulation for the optimal fixed fee:

$$A_1^* = \left(\frac{1}{2}\tilde{\theta}(s_2 - p_2)^2 - A_2\right)\left[\frac{-3A_2}{\tilde{\theta}(s_2 - p_2)^2} + \frac{7}{2} - \frac{2s_1}{(s_2 - p_2)^2}\sqrt{\frac{3}{2}(s_2 - p_2)^2 - \frac{3A_2}{\tilde{\theta}}}\right] - \frac{1}{3}\tilde{\theta}s_1\sqrt{\frac{3}{2}(s_2 - p_2)^2 - \frac{3A_2}{\tilde{\theta}}} . \tag{III.15}$$

3.3. Characterization of the Equilibria

Any pure strategy Nash equilibrium of the tariff game in which both firms have a positive market share, has to satisfy the equations (III.5), (III.6), (III.10), (III.12), (III.14), and (III.15). In a Nash equilibrium both firms choose the best response given the rival's tariff. Hence, for each firm the two first-order conditions must hold.[22] Equation (III.6) ensures that both firms choose the same marginal consumer $\tilde{\theta}$, which is necessary for an equilibrium. Condition (III.5) determines the optimal $\hat{\theta}$ for a given tariff of firm 1. More precisely, the Nash equilibrium is characterized by a system of equations with six equations and six unknown variables $(p_1, p_2, A_1, A_2, \tilde{\theta}, \hat{\theta})$. This system of equations can

[22] The second-order necessary conditions (SOCs) are checked in the appendix, see A.3.2. I suggest to postpone reading A.3.2. until the end of Section 3.

be condensed into a single equation, a polynomial of sixth order in $\tilde{\theta}$. This polynomial and therefore the Nash equilibrium cannot be solved analytically. Fortunately, if the degree of product differentiation is sufficiently large ($s_2/s_1 \geq 2$), it can be shown numerically that this polynomial has exactly one root in $[0, 1]$, which is the solution for $\tilde{\theta}$. If, however, the degree of product differentiation is low, then there does not exist an equilibrium in which both firms share the market.

If only one firm has a positive market share in equilibrium, then this is necessarily the high-quality firm. Note that, for any feasible tariff of firm 1, the tariff $T_2(q) = (s_2 - s_1)q$ leads to strictly positive profits for firm 2. Furthermore, in an equilibrium in which firm 1 makes non-positive profits there cannot be an unsatisfied residual demand, thus it has to be a best response for firm 2 to serves the whole market. The unique tariffs corresponding to an equilibrium where the high-quality firm serves the whole market are $T_1^*(q) = 0$ and $T_2^*(q) = (s_2 - s_1)q$. These tariffs constitute an equilibrium of the tariff game, if the degree of quality differentiation is not too high ($s_2/s_1 \leq 2$). If, on the other hand, the degree of quality differentiation is sufficiently high, firm 1 is not a tough competitor and thus firm 2 behaves more like a monopolist and leaves an unserved residual demand. The following proposition summarizes the tariff game equilibria for all relevant quality pairs. Note that $s_1 = 0$ is as if firm 1 is not present in the market.

Proposition III.1 (Tariff Game Equilibria): *(i) For $s_2 = 1$ and $s_1 = 2/3$ (low differentiation) there exists a unique pure strategy equilibrium. In equilibrium, both firms offer linear tariffs. The low-quality firm has a zero market share and makes zero profits, while the high-quality firm serves the whole market and makes strictly positive profits.*
(ii) For $s_2 = 2/3$ and $s_1 = 1/3$ (middle differentiation) there are two pure strategy equilibria. In one equilibrium, both firms use linear tariffs; firm 2 serves the whole market and makes strictly positive profits, while firm 1 has a zero market share and makes no profits. In the other equilibrium, both firms offer two part tariffs and realize strictly positive profits. In this equilibrium the market is not fully covered.
(iii) For $s_2 = 1$ and $s_1 = 1/3$ (high differentiation) there exists a unique pure strategy equilibrium. In equilibrium, the market is not fully covered and both firms offer a two-part tariff. The two firms share the market and both firms realize strictly positive profits.

By using Proposition 1, it is straightforward to obtain firms' optimal behavior at the quality stage. In case (i), it is profitable for firm 1 to deviate to $s_1 = 1/3$. In case (ii), irrespective of which equilibrium is played at the final stage, at least firm 1 has an incentive to deviate to $s_1 = 1$. Being the leading firm in the market is clearly better than serving a small residual demand. Thus, in the unique pure strategy subgame perfect equilibrium of the quality game and the tariff game, one firm chooses a quality level of 1, whereas the other firm produces a quality of 1/3. Since a firm decides to enter the industry if its expected profits exceed the entry cost, the next result follows immediately.

Proposition III.2: *Suppose the entry cost, K, is sufficiently small. Then there exists a unique subgame perfect equilibrium in pure strategies. Both firms enter the industry and make strictly positive profits in equilibrium.*

Applying numerical methods to solve for the root of the polynomial of sixth order, that characterizes the value of $\tilde{\theta}$ in equilibrium, allows one to compute the equilibrium profits, prices, and fixed fees:

$$\begin{aligned} \pi_1 &= .00138222 - K & \pi_2 &= .142833 - K \\ p_1 &= .0657433 & p_2 &= .351995 \\ A_1 &= .00353261 & A_2 &= .0550837 \\ \hat{\theta} &= .0986701 & \tilde{\theta} &= .29601 \end{aligned}$$

4. Two-Part Tariff Competition versus Linear Pricing

With the above analysis the characterization of firms' optimal entry decisions, depending on whether price discrimination is feasible or not, follows immediately.

Proposition III.3: *For sufficiently small entry cost, K, two firms enter the market if price discrimination is permitted, and a ban on price discrimination leads to a monopoly.*

If two-part tariffs are infeasible firms are not able to relax price competition. In contrast, if two-part tariffs are feasible the two firms produce distinct qualities and realize strictly positive profits. Relaxing price competition via quality differentiation is possible if and only if two-part tariff competition is feasible. What are the driving forces behind this result? An explanation can be given by analyzing the consumers' utility functions. Since all consumers have the same ranking for price-quality pairs firms cannot relax competition via quality differentiation when being restricted to linear pricing. Nevertheless, consumers with high types benefit more from an improved quality. Thus, a consumer's willingness to pay an upfront fee to obtain the right to purchase the high-quality product is increasing in his type. This is the reason why firms can relax competition if two-part tariffs are feasible. Here, firms' incentives for quality differentiation are based on heterogeneity in consumers' preferences for quantity, which is novel in the literature. Former research studied vertical differentiation only with linear tariffs, for instance Shaked and Sutton (1982). In these unit-demand frameworks, consumers are heterogeneous with regard to their tastes for quality.[23]

Now I address the question how a ban on price discrimination affects welfare, consumers' surplus and industry profits. Social welfare is defined as the sum of consumers' surplus and industry profits.

Proposition III.4: *Suppose the entry cost, K, is sufficiently small. Then social welfare and industry profits are higher and aggregate consumers' surplus is lower if two-part tariff competition is feasible.*

Permitting price discrimination encourages entry, which in turn increases welfare.[24] This finding is in line with the result obtained by Armstrong and Vickers (2001), who analyze non-linear pricing in a free entry circular city model. In their model, as in the one presented here, permitting price discrimination increases industry profits which in turn encourages entry, thereby increasing welfare. Surprisingly, from consumers' point of view it is better to have a monopolist who is restricted to linear pricing than facing two competing firms that are able to charge two-part tariffs.

It is worthwhile to point out, that imposing a minimum quality-standard can have the same entry effect as banning price discrimination. If the quality standard is too high, so that the possible profits of firm 1 are lower than the entry cost, then in the subgame perfect equilibrium only one firm will enter the market.

5. Motivation of the Utility Function: Applications to Vertical Relations

The utility function (III.1) can also be interpreted as a profit function of a retailer. Two-part tariffs are widely used and discussed in the context of vertical relations.[25] Assume that firms, called manufacturers in this section, do not sell their products to consumers directly but to a retail firm. A retail firm can only sell the product of a single manufacturer. As far as vertical relations are concerned, when the manufacturer sets a fixed fee, this fixed fee can be interpreted as franchise fee. For instance, the decision of a potential retailer may be to open either a McDonald's, a Pizza Hut, or

[23]See, for example, Choi and Shin (1992). In Shaked and Sutton (1982), consumers have the same utility function: $u(s,m) = s \cdot m$, and different incomes (m). Note that the income is like a taste parameter for quality.

[24]While market entry in a vertical product differentiation model is also analyzed by Donnenfeld and Weber (1992, 1995) and in particular by Johnson and Myatt (2003), these models are unit-demand models.

[25]Classic contributions to the theory of vertical restraints, in particular on upstream competition and two-part tariffs, are Rey and Stiglitz (1988) and Bonnano and Vickers (1988).

neither. A model of vertical differentiation and market power on the upstream market is also analyzed in Avenel and Caprice (2006). In their model a monopolist produces a high quality product and the low quality product is produced by a competitive fringe. Product differentiation on the upstream market is exogenous. The focus of Avenel and Caprice is on retailers' product lines. In the reinterpretation of the model introduced so far, both upstream firms have market power and the degree of quality differentiation is endogenous. But the retailers are exclusive dealers, who cannot sell the products of both manufacturers.

In the following, I will give two examples of how the utility function (III.1) can be interpreted as a profit function of a retailer who sells manufacturer i's products.

Example 1 (Retailer is price taker): Consider a retail firm that operates as a price taker. Suppose there exists a competitive fringe selling the same qualities as produced by the manufacturers. The market price depends on the quality the retailer sells and is given by

$$P(s_i) = s_i \; . \tag{III.16}$$

For selling q units of a product purchased by manufacturer i, a retailer of type $\theta \in (0.1]$ has costs

$$c(q,\theta) = \frac{1}{2}\frac{q^2}{\theta} \; , \tag{III.17}$$

where θ measures how efficient the retailer is. If θ is high, the retailer has relatively low costs for serving consumers. The profit of the retailer when selling the products of firm i is then given by

$$\pi_R^{PT}(q,i,\theta) = s_i q - \frac{1}{2}\frac{q^2}{\theta} - p_i q - A_i \; , \tag{III.18}$$

where A_i is a franchise fee the retailer has to pay for the permission to sell firm i's products and p_i is firm i's wholesale price. Obviously, the above profit function is equivalent to the utility function assumed throughout the whole paper.

Example 2 (Retailer is local monopolist): Consider a retailer of type $\theta \in (0,1]$ who sells the products of manufacturer i. Assume that the retailer is a local monopolist and operates without costs. The demand of a retailer of type θ with retail price P is $q = \theta D(P)$, where $D(P) = 2[s_i - P]$. A higher θ corresponds to a retailer with higher demand. Put differently, θ measures the "market size" of the retailer in the downstream market. Thus, the inverse demand function he faces is

$$P(q,\theta) = s_i - \frac{1}{2}\frac{q}{\theta} \; .$$

Given a tariff T_i offered by manufacturer i to each potential retailer, the profit of a retailer of type θ who sells firm i's products is

$$\pi_R^{LM}(q,i,\theta) = \left(s_i - \frac{1}{2}\frac{q}{\theta}\right) q - p_i q - A_i \; . \tag{III.19}$$

Hence, the profit function of the retailer is equivalent to the utility function (III.1).

In the main body of the paper, I focused on the case where each producer ($i = 1, 2$) offers his products to consumers directly. Nevertheless, the above applications allow for an interpretation of consumers as retail outlets.

6. CONCLUDING REMARKS

In this paper, a vertically differentiated duopoly with endogenous degree of product differentiation and two-part tariffs is analyzed. The main finding is that product differentiation occurs in equilibrium if and only if two-part tariff competition is feasible. With consumer heterogeneity regarding their preferences for quality being absent in the model presented, firms' incentives for quality differentiation are solely based on consumers' heterogeneity in their preferences for quantity. This finding is in contrast to earlier models of vertically differentiated industries, where consumers differ only in their tastes for quality and quality differentiation can always relax price competition.

The utility function assumed in this paper is arbitrary. The main findings, however, can also be obtained for surplus functions of the form $v_i(p_i, \theta) = \psi(\theta) \cdot \rho(s_i, p_i)$. Allowedly, the assumption that consumers differ only in their tastes for quantity and not in their tastes for quality is strong. By focusing on this extreme case, I established that the possibility of price discrimination may increase firms' strategic incentives for quality differentiation. Moreover, the model of the paper is simple in the sense that firms can produce only quality levels $s_i \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. For qualities $s_i \in [0, 1]$ the existence of a pure strategy equilibrium of the tariff game is complicated to show. Trusting that firms' first-order conditions with respect to marginal prices and marginal consumers characterize the tariff game equilibrium, application of numerical methods allows to compute the subgame perfect quality levels. Firm 2 produces quality $s_2 = 1$ and firm 1's quality level is $s_1 \approx 0.36$. All results remain qualitatively the same.

For real-world applications, this model represents just an extreme case, similar as pure spatial competition models. I believe, however, that focusing on pure vertical differentiation sheds some light on firms' incentives for quality differentiation in real life. From an elementary perspective, this paper demonstrates that when consumers are heterogeneous, firms can always relax price competition as long as there are enough dimensions for firms to differentiate their products.

IV. Uncertain Demand, Consumer Loss Aversion, and Flat-Rate Tariffs

The so called flat-rate bias is a well-known phenomenon caused by consumers' desire to be insured against fluctuations in their billing amount. This paper shows that consumer loss aversion provides a formal explanation for this bias. We investigate into two-part tariffs offered by oligopolistic firms to loss-averse consumers who are uncertain about their demand. The main finding is that, in equilibrium, firms offer a flat-rate tariff to those consumers whose degree of loss aversion is sufficiently high compared to marginal cost. Moreover, we show that differently loss-averse consumers may not impose an informational externality on each other. Thus, firms may be able to screen differently loss-averse consumers at no cost.

1. Introduction

Consumers facing the choice between several tariffs often do not select the optimal one given their consumption patterns. In particular, consumers often prefer a flat-rate tariff even though they would save money with a measured tariff. Train (1991) referred to this phenomenon as "flat-rate bias". Given the fact that consumers are willing to pay a "flat-rate premium" it is unsurprising that this tariff form is widely utilized in many industries, e.g., telephone services, Internet access, car rental, car leasing, DVD rental, amusement parks, health clubs, and many others.

That flat rates are such a favorable pricing scheme is hard to reconcile with orthodox economic theory, in particular for industries where marginal costs are non-negligible.[1,2] Train et al. (1989) point out that "customers do not choose tariffs with complete knowledge of their demand, but rather choose tariffs [...] on the basis of the insurance provided by the tariff in the face of uncertain consumption patterns". Since variations in the monthly billing rates are small compared to a consumer's income, standard risk aversion cannot capture this insurance motive, however.[3] Already Train (1991) states that "[t]he existence of this [flat-rate] bias is problematical. Standard theory of consumer behavior does not incorporate it." Therefore, to capture first-order risk aversion we posit that consumers are loss averse.[4] A loss averse consumer dislikes even small deviations from his reference point. In our model, the consumer's demand is uncertain at the point where he selects a tariff. We assume that a consumer forms rational expectations about his invoice, which determine his reference point. The consumer feels a loss if his actual invoice amount is above his reference point, and he feels a gain if it is below his reference point. We follow Kőszegi and Rabin (2006, 2007) and assume that the reference point is a full distribution of the possible billing rates. To illustrate this concept, suppose the consumer's amount invoiced is either \$15, \$20, or \$30. Then, receiving a bill of \$ 20 generates a mixture of feelings, a gain of \$5 and a loss of \$10. We show that a consumer with these preferences is biased in favor of flat-rate tariffs, since flat rates insure against the risk of losses in periods of greater than average usage.

Because observed tariffs are the result of strategic interactions of firms and consumers, we set up an oligopoly model where firms compete for loss averse consumers. Moreover, we allow for consumer heterogeneity with respect to their degree of loss aver-

[1] Sundararajan (2004) shows that it is always optimal for a monopolistic firm to offer a flat rate next to usage-based tariffs if marginal costs are zero and there are positive transaction costs accompanied with usage-based pricing.

[2] It is important to notice that, despite conventional wisdom, the marginal cost of a telephone call is not zero, see for instance Faulhaber and Hogendorn (2000). Moreover, one should keep in mind that telephone companies pay access charges on a per minute basis for off-net calls.

[3] Cf. Clay et al. (1992) or Miravete (2002).

[4] That consumer loss aversion could potentially explain the flat-rate bias is argued in the marketing literature, see Lambrecht and Skiere (2006).

sion. The timing is as follows: (i) Firms offer a menu of two-part tariffs to consumers; (ii) Each consumer selects either one tariff or none; (iii) The consumer observes the realization of the state of the world which determines his preferences for the good and chooses a demanded quantity. We analyze the symmetric information case in which firms observe a consumer's degree of loss aversion, as well as the asymmetric information case in which the degree of loss aversion is private information. In equilibrium of the symmetric information benchmark, firms offer a flat-rate tariff to those consumers whose degree of loss aversion compared to marginal cost exceeds a certain threshold. Consumers with a lower degree of loss aversion are assigned to a metered tariff, i.e., a two-part tariff with a strictly positive unit price. These findings turn out to carry over likewise to the asymmetric information case. Interestingly, offering the appropriate flat-rate tariff next to usage-based tariffs never reduces the firm's profits obtained from serving those consumers who are assigned to a usage-based contract. Moreover, under certain conditions differently loss averse consumers do not impose informational externalities on each other. Put differently, firms may be able to screen a consumer's degree of loss aversion at no cost.

The structure of the paper is as follows: After presenting some evidence on the prevalence of flat-rate biases and a discussion of the related literature, Section 2 presents the model and in particular consumer's preferences. Section 3 analyzes the demand function of a loss averse consumer who signed an arbitrary two-part tariff. Sections 4 and 5 solve firms' profit maximization problems and characterize the equilibrium outcome for the symmetric information and the asymmetric information case. The final section summarizes our main findings and discusses some shortcomings of the presented model.

Evidence and Related Literature

Existence and Causes of Tariff-Choice Biases.—The existence of tariff-choice biases was first documented for U.S. households among telephone service options. Train et al. (1987) provide evidence for U.S. households favoring flat-rate tariffs over measured services for local telephone calls. Conducting a logit model with a tariff specific constant, the authors find that this constant is highest for the flat-rate option. Similar results are obtained by Train et al. (1989). The authors argue that consumers choose a tariff that ends up not being cost-minimizing for the customer's level of consumption because consumers also care about the insurance provided by the tariff option. Given uncertain consumption patterns "the flat-rate tariff provides complete insurance" (Train et al., 1989). A tendency of households to prefer flat-rate tariffs for telephone services is also reported by Hobson and Spady (1988) for single-person households, by Kling and van der Ploeg (1990) who evaluate a tariff experiment of AT&T, by Mitchell and Vogelsang (1991), and Kridel et al. (1993). For instance, Kridel et al. (1993) find that 55% of all customers who choose a flat-rate service would have achieved higher surplus if they had chosen measured service instead. The authors also hold an insurance motive of the customers responsible for this finding. They report that customers exhibit substantial risk aversion when faced with bill uncertainty. Miravete (2003) rejects the thesis that customers are subject to a flat-rate bias when selecting telephone service tariffs. In his data set, however, the flat-rate option is optimal for the vast majority of households.

The flat-rate bias is documented also for other telecommunication services. Lambrecht and Skiera (2006) analyze transactional data of over 10,000 customers of an Internet service provider in Germany. They find that over 50% of these customers are biased in favor of a flat-rate option. Moreover, they provide evidence that the flat-rate bias is at least partially due to an insurance motive of the consumers. In a follow up paper, Lambrecht et al. (2007) argue "[c]onsumers may prefer a tariff that leads to fewer month-to-month fluctuations in their bill". For mobile telephone services, a preference for flat-rate tariffs that cannot be explained by customers' usage is documented by Gerpot (2009) and Mitomo et al. (2009). These papers, however, rely on survey data.

Relying on survey data, Nunes (2000) finds strong evidence for a flat-rate bias outside telecommunications services (grocery shopping online, access payment for a swimming pool of an apartment building). Presumably the most powerful demonstration of the flat-rate bias outside the telecommunications service sector is DellaVigna and Malmendier (2006). They analyze a data set from three U.S. health clubs and show that a large fraction of health club members who are enrolled in a flat-fee contract (either monthly or annually) paid on average more per visit than they would have paid with a pay-per-visit option. According to the authors, the leading explanation for these observations is consumers' overconfidence about future self-control. A low per usage price is a commitment device for higher attendance in case of self-control problems, when consumption leads to immediate costs and delayed rewards. Such motives of selecting an option that provides commitment to higher usage rates obviously cannot explain the prevalence of flat rates for telecommunications services, car rental, car leasing and amusement parks. Moreover, from a theoretical point of view, consumer self-control problems can only explain why marginal prices may be below marginal cost but not why marginal prices are exactly zero, see DellaVigna and Malmendier (2004).[5]

[5]In DellaVigna and Malmendier (2004), generically the optimal per-unit price is unequal to zero. If one does not allow for negative unit prices, then there is a broad range of parameter combinations such that the optimal unit price equals zero.

(Behavioral) Models of Pricing Strategies.—Since Oi's (1971) analysis of an optimal two-part tariff for a monopolist, this pricing scheme is intensely analyzed in the economic literature. Leland and Meyer (1976) show that a firm, regardless of its objective, always does at least as well with a two-part tariff as with linear pricing. Pareto-optimal menus of tariffs are analyzed by Willig (1978). He shows that a Pareto-optimal menu includes a cost-based two-part tariff. The pricing literature of the 80's solves for the optimal nonlinear tariff. Notable works on this topic are Maskin and Riley (1984), Goldman et al. (1984) as well as the book by Wilson (1993). This literature established the now well-known no distortion at the top result, i.e., marginal prices exceed marginal cost for all but the last unit. While these classic screening models focus on deterministic demand, there are some papers analyzing sequential screening problems. In these papers, a consumer first chooses a contract and then he learns his true preferences before making a quantity choice. See, for instance, Courty and Li (2000) or Miravete (2002).

This paper is more related to the recent and growing literature investigating how rational firms respond to consumer biases. A seminal contribution in this field is DellaVigna and Malmendier (2004). They consider a market, either monopolistic or perfectly competitive, with homogeneous time-inconsistent consumers. Their main finding is that the unit price of the optimal two-part tariff is above marginal cost for leisure goods (usage of rental car) and below marginal cost for investment goods (health club attendance). Likewise presuming that consumers are quasi-hyperbolic discounters, Heidhues and Kőszegi (2008b) set up a model of a perfectly competitive market for credit-cards. They allow for consumer heterogeneity and pay particularly attention to welfare implications of possible policy interventions. Using a different notion of time-inconsistency, Eliaz and Spiegler (2006) solve for the optimal menu of tariffs for a monopolist who faces consumers that differ in their degree of sophistication. The optimal contract exploits those consumers who are sufficiently naive about their self-control problems. Moreover, they show that the optimal menu can be implemented by a menu of three-part tariffs. The optimal menu does not include a flat-rate tariff. The optimal nonlinear pricing scheme for a monopolist who sells to consumers with self-control problems is also analyzed by Esteban et al. (2007). Instead of assuming time inconsistency, they model self-control problems by applying the concept of Gul and Pesendorfer (2001). The optimal tariff resembles the one in the standard nonlinear pricing literature except for a price ceiling. Similar results are obtained by Esteban and Miyagawa (2006) for a perfectly competitive market where consumers have temptation preferences according to Gul-Pesendorfer.

Next to time inconsistency, there are a few papers dealing with the optimal selling strategy for overconfident consumers. Grubb (forthcoming) analyzes the optimal menu of nonlinear price schedules for a monopolist as well as for a perfectly competitive market. Consumers in his model are overconfident in the sense that they underestimate fluctuations in their demand. The optimal menu is close to a menu of three-part tariffs which is often observed in the cellular phone service industry. The optimal tariff is completely flat up to the allowance only if marginal costs are zero, however. A similar model where firms screen consumers at the basis of their priors is considered by Uthemann (2005). In his model firms are differentiated à la Hotelling. Unlike Grubb, he does not assume that consumption is satiated at a finite level, and therefore he obtains that marginal prices are always above marginal cost. Focusing on only two-states of the world but without imposing any differentiability assumptions on the consumer's utility function, Eliaz and Spiegler (2008) analyze the problem of a monopolist who faces consumers with biased beliefs regarding the probability assignment to the two states of nature. Optimistic consumers, who assign too much weight on the state of nature that is characterized by larger gains from trade, sign exploitative contracts. In a stylized example, the authors show that the optimal menu may include a flat-rate tariff. The authors, however, do not derive conditions under which there model predicts flat-rate contracts.

To the best of our knowledge, there is no paper analyzing nonlinear tariffs when consumers are loss averse. Nevertheless, loss aversion and in particular the concept developed by Kőszegi and Rabin (2006, 2007) is used in models of industrial organization. Heidhues and Kőszegi (2005) apply this concept to provide an explanation why monopoly prices react less sensitive to cost shocks than predicted by orthodox theory. Moreover, Heidhues and Kőszegi (2008a) introduce consumer loss aversion into a model of horizontally differentiated firms. They show that in equilibrium asymmetric competitors charge identical focal prices for differentiated products. Next to industrial organization, the Kőszegi and Rabin formulation is applied to contract theory by Herweg et al. (2008). Considering a moral hazard framework, they provide an explanation for the frequent usage of lump-sum bonus contracts.

2. The Model

Market Framework

We consider a market for a single good. In the market there are two firms, A and B, facing a continuum of potential consumers with measure normalized to one.

Consumers.—The consumers can be partitioned into two groups that differ in a parameter $\lambda \geq 1$ (degree of loss aversion). Let the two groups be denoted by $j = 1, 2$ with

$\lambda_1 < \lambda_2$. A consumer's demand for the good is continuous and depends on the state of the world $\theta \in [\underline{\theta}, \bar{\theta}] \equiv \Theta$. The state of the world is continuously distributed according to the commonly known and twice differentiable cumulative distribution function $F(\cdot)$. Let the probability density function be $f(\cdot)$. The state of the world is unknown to consumers and firms at the point of contracting. For instance, a consumer may sign a contract with a car rental company today for his holidays in a few weeks. How frequently he will use the rented car depends on the weather. If the sun is always shinning the consumer uses the car only to drive to the nearby beach. But if the weather is bad he takes longer sight-seeing trips.

Firms & Timing.—Both profit maximizing firms produce the single good at constant marginal cost $c > 0$ and without fixed cost. Each firm $i = A, B$ offers a two-part tariff to each group of consumers $j = 1, 2$. The tariff is given by $T^i_j(q) = L^i_j + p^i_j q$, where $q \geq 0$ is the quantity, and L^i_j and p^i_j denote the fixed fee and the per unit price, respectively, charged by firm i from consumers of type j. We will analyze the symmetric information case in which firms observe λ, as well as the asymmetric information case in which λ is private information of the consumer.

The timing is as follows: (1) Firms simultaneously and independently offer a tupel of two-part tariffs $\{(L^i_j, p^i_j)\}_{j=1,2}$ to consumers. (2) Each consumer either signs exactly one contract or none. (3) Consumers, but not firms, observe the realization of the state of the world. A consumer chooses his demand to maximize his utility for the given state of the world if he signed a contract. (4) Finally, payments are made according to the demanded quantities and the concluded contracts.

Discrete Choice Framework.—The products of the two firms are symmetrically differentiated. We assume that, next to λ, consumers are heterogeneous with respect to their brand preferences. Each consumer has idiosyncratic preferences for differing brands of the product (firms), which are parameterized by $\zeta = (\zeta^0, \zeta^A, \zeta^B)$. A consumer with brand preferences ζ has net utility $v^i + \zeta^i$ if he buys from firm i, and net utility ζ^0 if no contract is signed. The brand preferences $\zeta = (\zeta^0, \zeta^A, \zeta^B)$ are independently and identically distributed according to a known distribution among the two groups of consumers.

To solve for the tariffs that are offered in the pure-strategy Nash equilibrium by the two firms, we follow the approach of Armstrong and Vickers (2001) and model firms as offering utility directly to consumers. Each two-part tariff can be considered as a deal of a certain expected value that is offered by a firm to its consumers. Thus, firms compete over customers by trying to offer them better deals, i.e., a two-part tariff that yields higher utility (including gain-loss utility). Put differently, we decompose a firm's problem into two parts. First, we solve for the two-part tariff that maximizes profits subject to the constraint that the consumer receives a certain utility level. Thereafter, we solve for the utility levels (v_1^i, v_2^i) a firm i offers to its customers. It is important to note that, when λ is unobservable, the two-part tariffs have to be designed such that each group of consumers prefers the offer that is dedicated to them. Suppose the utility offered to consumers of group j by firm A and firm B is v_j^A and v_j^B, respectively. Furthermore, assume that the incentive constraints are satisfied. Then, the market share of firm A in the submarket j is $m_j(v_j^A, v_j^B)$ and the market share of firm B is $m_j(v_j^B, v_j^A)$, with $m_j(v_j^A, v_j^B) + m_j(v_j^B, v_j^A) \leq 1$. The market share function $m_j(\cdot)$ is increasing in the first argument and decreasing in the second. Since the brand preferences are identically distributed among the two groups, the market share functions are identical for the two submarkets, i.e., $m_1(\cdot) = m_2(\cdot) = m(\cdot)$. Following Armstrong and Vickers, we impose some regularity conditions in order to guarantee existence of equilibrium. First, we assume that

$$\frac{\partial m(v^A, v^B)/\partial v^A}{m(v^A, v^B)} \text{ is non-decreasing in } v^B.$$

Second, we assume that for each submarket the collusive utility level $\tilde{v}_j$ exists which maximizes (symmetric) joint profits.[6]

Consumers' Preferences

We assume that consumers are loss averse, in the sense that a consumer is disappointed if the payment he has to make exceeds his reference payment. For instance, consumers typically feel a loss if at the end of the month the invoice from their telecommunication provider is larger than expected. Since, for the situations we have in mind, it is natural to assume that the reference point incorporates forward looking expectations, we apply the approach of reference-dependent preferences developed by Kőszegi and Rabin (2006, 2007). First, this concept posits that overall utility has two additively separable components, consumption utility (intrinsic utility) and gain-loss utility. Second, the consumer's reference point is determined by his rational expectations about outcomes. Finally, a given outcome is evaluated by comparing it to each possible outcome, where each comparison is weighted with the ex-ante probability with which the alternative outcome occurs.

The consumer's intrinsic utility is quasi linear in money; formally, intrinsic utility equals $u(q, \theta) - T^i(q) - \zeta^i$ if he purchases from firm i. For the markets we have in mind, like rental cars or Internet services, even if the price per unit is zero, demand is bounded. Therefore, we assume that there exists a satiation point, $q^S(\theta)$, and that

[6]For a detailed description of the competition-in-utility-space framework and the needed assumptions see Armstrong and Vickers (2001).

overconsumption is harmless, i.e., free disposal is possible. Additionally, it is assumed that a higher state of the world is associated with a stronger need for the good. The restrictions on the consumer's intrinsic utility for the good are summarized in the following assumption.

Assumption (A1): *It is assumed that for all $\theta \in \Theta$: (i) not consuming the good yields zero intrinsic utility $u(0,\theta) \equiv 0$. (ii) $u(q,\theta)$ is C^3 for $q \leq q^S(\theta)$. Intrinsic utility for the good has the following properties,*

$$\begin{aligned} &\partial u(q,\theta)/\partial q > 0 \quad \text{for } q < q^S(\theta), && \partial^2 u(q,\theta)/\partial q^2 < 0 \quad \text{for } q \leq q^S(\theta), \\ &\partial u(q,\theta)/\partial q = 0 \quad \text{for } q \geq q^S(\theta), && \partial^2 u(q,\theta)/\partial q \partial\theta > \kappa \quad \text{for } q \leq q^S(\theta), \end{aligned}$$

where $\kappa > 0$. (iii) $\partial u(0,\theta)/\partial q = \infty$.

Some comments to Assumption $(A1)$ are in order: It is assumed that $u(q,\theta)$ is thrice differentiable in q and θ for $q \leq q^S(\theta)$. There exists a satiation point $q^S(\theta)$ such that $u(q^S(\theta),\theta) \geq u(q;\theta)$ for all $q \geq 0$. Moreover, overconsumption is harmless, i.e., $u(q;\theta) = u(q^S(\theta),\theta)$ for $q \geq q^S(\theta)$. Finally, since the intrinsic utility of zero consumption is normalized to zero and marginal utility is increasing in the state of the world it holds that $\partial u(q,\theta)/\partial\theta > 0$ for $q \in (0, q^S(\theta)]$.

By Assumption $(A1)$, the satiation point $q^S(\theta)$ is defined by

$$q^S(\theta) = \min\{q \in \mathbb{R}^+ | \partial u(q,\theta)/\partial q = 0\} \,. \tag{IV.1}$$

The satiation point is increasing in the state of the world θ, formally:[7]

$$\frac{dq^S(\theta)}{d\theta} = -\frac{\partial^2 u(q^S(\theta),\theta)/\partial q\partial\theta}{\partial^2 u(q^S(\theta),\theta)/\partial q^2} > 0 \,. \tag{IV.2}$$

We depart from the Kőszegi and Rabin concept by assuming that the consumer feels gains and losses only in the money dimension. This means that a consumer does not feel a loss if the weather is nice and he uses the rented car less often than expected. Similarly, he does not feel a gain when using the car more often than expected due to bad weather. The consumer feels a loss, however, if the rental price depends on the driven miles and he used the car more often than expected. In accordance with the majority of applied-loss aversion papers, it is assumed that the consumer's gain-loss function is piece-wise linear.[8] Suppose the consumer pays T, but expected to pay $\hat{T}$, then his gain-loss utility is given by

$$\mu(\hat{T} - T) = \begin{cases} \hat{T} - T &, \quad \text{for } \hat{T} \geq T \\ -\lambda(T - \hat{T}), & \quad \text{for } T > \hat{T} \end{cases} ,$$

[7]Strictly speaking, we have to take the left-hand limit when q approaches $q^S(\theta)$ to obtain the stated derivative.

[8]See, for instance, Heidhues and Kőszegi (2005, 2008a).

where $1 \leq \lambda \in \{\lambda_1, \lambda_2\}$.[9] Note that a consumer's expected demand conditional on the state of the world fully determines the distribution of his expected payments, and thus in turn his reference point. Suppose the consumer signed a contract with firm i. Then his overall utility from this deal when purchasing q units, given the state of the world is ϕ and his expected consumption is $\langle q(\theta)\rangle_{\theta\in\Theta}$, is given by

$$U_j(q|\phi, \langle q(\theta)\rangle) + \zeta^i \ ,$$

where

$$U_j(q|\phi, \langle q(\theta)\rangle) = u(q,\phi) - T^i(q) + \int_{X(q)} [T^i(q(\theta)) - T^i(q)] f(\theta)\, d\theta \\ - \lambda_j \int_{X^c(q)} [T^i(q) - T^i(q(\theta))] f(\theta)\, d\theta \ , \quad \text{(IV.3)}$$

with $X(q) \equiv \{\theta \in \Theta | T^i(q) < T^i(q(\theta))\}$ and $X^c(q) \equiv \{\theta \in \Theta | T^i(q) > T^i(q(\theta))\}$. Note that X^c is not the complementary set to X, since both sets are defined by strict inequalities, i.e., $X(q) \cup X^c(q) \subseteq \Theta$. Moreover, observe that for $z > 0$, $X(q) \supseteq X(q+z)$ and $X^c(q) \subseteq X(q+z)$ as long as $T(q)$ is non-decreasing. To deal with the resulting interdependence between actual consumption and expected consumption, we use the personal equilibrium concept, which requires the strategy that generates expectations to be optimal conditional on these expectations.[10]

Definition IV.1 (Personal Equilibrium): *For a given per unit price p the demand profile $\langle \hat{q}_j(\theta;p)\rangle_{\theta\in\Theta}$ is a personal equilibrium if for all $\phi \in \Theta$,*

$$\hat{q}_j(\phi;p) \in arg\max_{q\geq 0} \ U_j(q|\phi, \langle \hat{q}_j(\theta;p)\rangle) \ .$$

3. THE DEMAND FUNCTION

In this section, we characterize the consumer's demand conditional on having accepted the offer (p^i, L^i) of firm i. Since, the consumer's demand is independent of his idiosyncratic brand preferences, we will omit ζ until we discuss the consumer's brand choice. Moreover, to cut down on notation we suppress the superscript $i \in \{A, B\}$ indicating from which firm the consumer purchases and the subscript $j \in \{1, 2\}$ denoting the

[9] With this formalization the weight on gain-loss utility is implicitly normalized to one. This normalization has no qualitative impact on our results. Even for $\lambda = 1$ (no loss aversion), however, the consumer has reference-dependent preferences. His utility for a given state of the world ϕ is then $u(q,\phi) - 2T(q) + \text{constant}$. The consumer values money twice at the moment of the purchasing decision, since paying one dollar more reduces intrinsic utility and reduces gain-loss utility either by reducing gains or by increasing losses. Ex ante, when making plans the consumer's expected utility for $\lambda = 1$ equals the expected utility of a consumer without gain-loss utility. These important observations will become clearer after deriving the consumer's utility.

[10] See Kőszegi and Rabin (2006, 2007) for a general description and a defense of this concept of consumer behavior.

consumer's degree of loss aversion in the following. We focus on $p \geq 0$ since negative unit prices cannot be optimal due to harmless overconsumption.

In order not to render flat-rate tariffs completely infeasible, it is assumed that in case of being indifferent between two or more quantities the consumer chooses the lowest of these quantities. Alternatively, one could assume that overconsumption is not harmless. Since a higher state of the world is associated with a stronger preference for the good, it seems reasonable that the demand profile is increasing in the state of the world. The following lemma shows that this is indeed the case.

Lemma IV.1: *For any two states of the world $\phi_1, \phi_2 \in \Theta$ with $\phi_1 < \phi_2$, $\hat{q}(\phi_1; p) \leq \hat{q}(\phi_2; p)$.*

Unless specified otherwise, all proofs are presented in the appendix. Since in any personal equilibrium demand is increasing in the state of the world and $p \geq 0$, the consumer feels losses compared to lower states and gains compared to higher ones. Thus, the consumer's utility for a given state ϕ in a personal equilibrium can be written as

$$U(\hat{q}(\phi;p)|\phi, \langle \hat{q}(\theta;p)\rangle) = u(\hat{q}(\phi;p), \phi) - T(\hat{q}(\phi;p)) \\ + \int_{\phi}^{\bar{\theta}} [T(\hat{q}(\theta;p)) - T(\hat{q}(\phi;p))] f(\theta)\, d\theta - \lambda \int_{\underline{\theta}}^{\phi} [T(\hat{q}(\phi;p)) - T(\hat{q}(\theta;p))] f(\theta)\, d\theta\,. \quad \text{(IV.4)}$$

Taking the expected value with respect to the state of the world of the above formula yields the consumer's ex ante expected utility on the equilibrium path,

$$\mathbb{E}_{\theta}[U(\hat{q}(\theta;p)|\theta, \langle q(\theta;p)\rangle)] = \int_{\underline{\theta}}^{\bar{\theta}} [u(\hat{q}(\theta;p);\theta) - T(\hat{q}(\theta;p))] f(\theta)\, d\theta \\ - (\lambda - 1) \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} [T(\hat{q}(\phi;p)) - T(\hat{q}(\theta;p))] f(\phi) f(\theta)\, d\phi d\theta\,. \quad \text{(IV.5)}$$

The first integral of the above formula represents standard expected intrinsic utility. The second term is the ex ante expected net loss of the consumer, which is weighted by $\lambda - 1$. The reason is that the consumer compares a given outcome with each possible other outcome. Thus, the comparison of any two possible payments enters the consumer's expected utility exactly twice, once as a loss and once as an equally-sized gain. For $\lambda = 1$ the consumer puts equal weights on gains and losses, so ex ante all these comparisons cancel out. When making the purchasing decision, however, even for $\lambda = 1$ these comparisons do not cancel out since then expectations are fixed.

To further characterize the personal equilibria it is useful to distinguish two cases, namely, whether or not the marginal price p is strictly positive. We start with the case of a flat-rate tariff where $p = 0$.

Marginal Price Equals Zero

Obviously, with $p = 0$ the consumer does neither feel any sensation of gains nor of losses. Thus, the consumer maximizes for each state of the world his intrinsic utility for the good.

Lemma IV.2: *Suppose $p = 0$, then there is a unique personal equilibrium. In the personal equilibrium the consumer demands for all states of the world $\theta \in \Theta$ his satiation quantity, i.e., $\hat{q}(\theta; 0) = q^S(\theta)$.*

Proof: Follows directly from the observation that the consumer's utility for an arbitrary state of the world is independent of the expected demand for all other states of the world. Formally, $U(\hat{q}(\phi; p)|\phi, \langle q(\theta; p)\rangle) = u(q; \phi) - L$ which is maximized for $q \geq q^S(\phi)$. By our assumption that the consumer does not overconsume, it follows immediately that demand equals $q^S(\phi)$.

Q.E.D.

With the marginal price being zero the consumer's expected utility is

$$\mathbb{E}_\theta[U(q^S(\theta)|\theta, \langle q^S(\theta)\rangle)] = \int_{\underline{\theta}}^{\bar{\theta}} u(q^S(\theta), \theta) f(\theta)\, d\theta - L\ .$$

Let $S(p)$ be the expected joint surplus of a firm and a consumer when contracting at marginal price p. Formally

$$S(p) \equiv \mathbb{E}_\theta[U(\hat{q}(\theta; p)|\theta, \langle q(\theta; p)\rangle)] + (p - c)\mathbb{E}_\theta[\hat{q}(\theta; p)] + L\ . \tag{IV.6}$$

Thus, the generated joint surplus from a flat-rate tariff amounts to

$$\begin{aligned} S(0) &= \int_{\underline{\theta}}^{\bar{\theta}} \left[u(q^S(\theta), \theta) - cq^S(\theta)\right] f(\theta)\, d\theta \\ &= \int_{\underline{\theta}}^{\bar{\theta}} \int_0^{q^S(\theta)} \left[\frac{\partial u(q, \theta)}{\partial q} - c\right] dq f(\theta)\, d\theta \end{aligned} \tag{IV.7}$$

The surplus generated by a flat rate, $S(0)$, becomes arbitrarily negative for sufficiently large marginal costs c and approaches the first-best surplus S^{FB} for $c \to 0$, where $S^{FB} := \max_{\langle q(\theta)\rangle_\theta \in \Theta} \int_{\underline{\theta}}^{\bar{\theta}} (u(q(\theta), \theta) - cq(\theta)) f(\theta) d\theta$.

Positive Marginal Prices

First, we present a technical result: The consumer's demand when playing a personal equilibrium does not "jump" if the state of the world changes slightly.

Lemma IV.3: *Suppose $p > 0$. Then any personal equilibrium $\langle \hat{q}(\theta; p)\rangle_{\theta\in\Theta}$ is continuous in the state of the world θ.*

Since the demand profile is non-decreasing and continuous, we can conclude that it is differentiable almost everywhere. Before exploiting this property of the personal equilibrium, it is shown that demand is strictly increasing in the state of the world if the marginal price is small. To establish this result we define the function $\tilde{q}(\theta; p)$, which is implicitly characterized by

$$\frac{\partial u(\tilde{q}(\theta; p), \theta)}{\partial q} = p\ [2 + (\lambda - 1)F(\theta)]\ . \tag{IV.8}$$

In fact, as we will show, the function $\tilde{q}(\theta; p)$ characterizes the unique personal equilibrium if p is sufficiently small. It turns out that whether or not $\tilde{q}(\theta; p)$ is strictly increasing in the state of the world plays an important role for the characterization of personal equilibria. Implicit differentiation of (IV.8) with respect to θ yields

$$\frac{d\tilde{q}(\theta; p)}{d\theta} = -\frac{\partial^2 u(\tilde{q}(\theta; p), \theta)/\partial q \partial\theta - p(\lambda - 1)f(\theta)}{\partial^2 u(\tilde{q}(\theta; p), \theta)/\partial q^2}\ .$$

The function $\tilde{q}(\theta; p)$ is strictly increasing in θ if and only if the following condition is satisfied:

Condition: *For all $\theta \in \Theta$,*

$$p < \frac{\partial^2 u(\tilde{q}(\theta; p), \theta)/\partial q \partial\theta}{(\lambda - 1)f(\theta)}\ . \tag{C1}$$

The condition is more likely to be satisfied if the distribution of the state of the world is not very dense. Put differently, if the environment is sufficiently unpredictable then Condition 1 holds. Furthermore, the above condition is satisfied if the per unit price is sufficiently small. To see this, note that $\partial^2 u(q, \theta)/\partial q \partial\theta > \kappa$ for all $\theta \in \Theta$ and all $q \geq 0$. Now, we are prepared to establish the result that in a personal equilibrium there cannot be a set of types that consumes the same amount if the per unit price is low.

Lemma IV.4: *Suppose $p > 0$. Then in a personal equilibrium it holds that for any two types $\phi_1, \phi_2 \in \Theta$ with $\phi_1 \neq \phi_2$, $\hat{q}(\phi_1; p) \neq \hat{q}(\phi_2; p)$ if and only if $(C1)$ holds.*

Unique Personal Equilibrium

We know that the personal equilibrium $\hat{q}(\theta; p)$ is differentiable almost everywhere. Since, given $(C1)$ holds, the personal equilibrium is strictly increasing we can conclude that $d\hat{q}(\theta; p)/d\theta > 0$. Put differently, the consumer's overall utility is differentiable with respect to q for $q \in [\hat{q}(\underline{\theta};p), \hat{q}(\bar{\theta};p)]$, except at a finite number of kinks. In the following, we derive a candidate personal equilibrium where higher types consume strictly more by construction. It is shown that there exists exactly one candidate. This equilibrium candidate is strictly increasing given $(C1)$ holds.

The utility of a consumer in state ϕ who consumes $q \in [\hat{q}(\underline{\theta};p), \hat{q}(\bar{\theta};p)]$ units, given he expected to play a personal equilibrium where consumption is strictly higher for higher states, is given by

$$U(q|\phi, \langle\hat{q}(\theta;p)\rangle) = u(q;\phi) - pq - L \\ + p\int_{\alpha(q)}^{\bar{\theta}} [\hat{q}(\theta;p) - q] f(\theta)\, d\theta - \lambda p \int_{\underline{\theta}}^{\alpha(q)} [q - \hat{q}(\theta;p)] f(\theta)\, d\theta \;, \quad \text{(IV.9)}$$

where $\alpha(q)$ is implicitly defined by $\hat{q}(\alpha(q);p) \equiv q$. Note that the derivative $\alpha'(q) = (d\hat{q}(\alpha(q);p)/d\theta)^{-1} > 0$ almost everywhere if $(C1)$ holds. Taking the partial derivative of $U(q|\cdot)$ with respect to q yields

$$\frac{\partial U(q|\cdot)}{\partial q} = \frac{\partial u(q,\phi)}{\partial q} - p - \lambda p[q - \hat{q}(\alpha(q);p)] f(\alpha(q))\alpha'(q) - \lambda p \int_{\underline{\theta}}^{\alpha(q)} f(\theta)\, d\theta \\ - p[\hat{q}(\alpha(q);p) - q] f(\alpha(q))\alpha'(q) - p\int_{\alpha(q)}^{\bar{\theta}} f(\theta)\, d\theta \;. \quad \text{(IV.10)}$$

Taking into account that $\hat{q}(\alpha(q);p) - q = 0$, the above derivative can be simplified to

$$\frac{\partial U(q|\cdot)}{\partial q} = \frac{\partial u(q,\phi)}{\partial q} - p - p[1 - F(\alpha(q))] - p\lambda F(\alpha(q)) \;.$$

Hence, the consumer's utility is strictly concave for $q \in [\hat{q}(\underline{\theta};p), \hat{q}(\bar{\theta};p)]$ since

$$\frac{\partial^2 U(q|\cdot)}{\partial q^2} = \frac{\partial^2 u(q,\phi)}{\partial q^2} - p(\lambda - 1) f(\alpha(q))\alpha'(q) < 0 \;.$$

A necessary condition for $\langle\hat{q}(\theta;p)\rangle_{\theta\in\Theta}$ to constitute a personal equilibrium is that for all $\theta \in \Theta$ the first-order condition $\partial U(\hat{q}(\theta;p)|\theta,\cdot)/\partial q = 0$ is satisfied. Thus, the following condition is necessary for a personal equilibrium with a strictly increasing demand function: $\forall\, \theta \in \Theta$,

$$\frac{\partial u(\hat{q}(\theta;p),\theta)}{\partial q} = p\;[2 + (\lambda - 1)F(\theta)] \;. \quad \text{(IV.11)}$$

Note that (IV.11) gives us a unique candidate for a personal equilibrium with strictly increasing demand. Equation (IV.11) characterizes the unique personal equilibrium if $d\hat{q}(\theta;p)/dq > 0$ for all $\theta \in \Theta$, which is satisfied if and only if condition $(C1)$ holds. Furthermore, note that $\hat{q}(\theta;p) = q^S(\theta)$ for $p = 0$.

Proposition IV.1: *Suppose $(C1)$ holds. Then there exists a unique personal equilibrium $\langle\hat{q}(\theta;p)\rangle_{\theta\in\Theta}$. The personal equilibrium is characterized by $\partial u(\hat{q}(\theta;p),\theta)/\partial q = p[2 + (\lambda - 1)F(\theta)]$.*

The consumer's expected utility from signing a two-part tariff (p, L) is given by

$$\mathbb{E}_\phi[U(\hat{q}(\phi;p)|\phi,\langle\hat{q}(\theta;p)\rangle)] = \int_{\underline{\theta}}^{\bar{\theta}} \Big\{ u(\hat{q}(\phi;p),\phi) - p\hat{q}(\phi;p)$$
$$+ p\int_{\phi}^{\bar{\theta}} [\hat{q}(\theta;p) - \hat{q}(\phi;p)]f(\theta)\, d\theta - \lambda p \int_{\underline{\theta}}^{\phi} [\hat{q}(\phi;p) - \hat{q}(\theta;p)]f(\theta)\, d\theta \Big\} f(\phi)\, d\phi - L\,. \tag{IV.12}$$

Personal Equilibrium with Bunching

How does the personal equilibrium look like if condition $(C1)$ fails to hold? In this case there exists an interval of states of the world for which demand is the same. Before characterizing the personal equilibrium candidates, we show that any personal equilibrium demand is bounded from above and from below. Let the lower and the upper bound be denoted by q^{MIN}, respectively q^{MAX}. Clearly, these bounds depend on the marginal price p. It is straightforward to show that the bounds are characterized by the following equations,[11]

$$\frac{\partial u(q^{MIN},\underline{\theta})}{\partial q} = (\lambda+1)p \qquad \text{and} \qquad \frac{\partial u(q^{MAX},\bar{\theta})}{\partial q} = 2p.$$

For $q < q^{Min}$ even the lowest type, $\underline{\theta}$, has an incentive to deviate to a higher quantity. Similarly, for $q > q^{MAX}$ it is optimal for all types, even for the highest type, $\bar{\theta}$, to deviate to a lower quantity.

By Lemmas 2 and 3, any personal equilibrium is continuous in the state of the world even if $(C1)$ does not hold. Furthermore, if the personal equilibrium consists of flat parts as well as strictly increasing parts, then for the strictly increasing parts the personal equilibrium is given by $\tilde{q}(\theta;p)$. Thus, if the flat-part is an interior interval of Θ, then at the boundary points condition (IV.8) has to hold. On the other hand, if the flat segment starts at $\underline{\theta}$ or ends at $\bar{\theta}$, then the "bunching quantity" $\bar{q}$ has to satisfy an inequality constraint: Given $\bar{q}$ a downward (upward) deviation has to reduce the utility of the type $\underline{\theta}$ (respectively $\bar{\theta}$). The following lemma characterizes these cases.

Lemma IV.5: *Consider a personal equilibrium $\langle\hat{q}(\theta;p)\rangle_{\theta\in\Theta}$ with bunching in at least one interval $I \subseteq \Theta$ with bounds θ_1 and θ_2 where $\underline{\theta} \leq \theta_1 < \theta_2 \leq \bar{\theta}$, i.e., $\hat{q}(\theta;p) = \bar{q}\ \forall\ \theta \in I$. Then the constant quantity $\bar{q}$ and the bounds, θ_1 and θ_2, are characterized by*

$$\partial u(\bar{q},\theta_1)/\partial q - p[2+(\lambda-1)F(\theta_1)] = 0 \qquad \textit{if } \theta_1 > \underline{\theta}$$
$$\partial u(\bar{q},\theta_1)/\partial q - p[2+(\lambda-1)F(\theta_1)] \geq 0 \qquad \textit{if } \theta_1 = \underline{\theta}$$

[11] Suppose $\hat{q}(\phi;p) < q^{MIN}\ \forall \phi \in \Theta$. Then, if a consumer of type θ chooses a quantity $q \geq q^{MIN}$ his utility is $u(q,\theta) - pq - \lambda p \int_{\underline{\theta}}^{\bar{\theta}} [q - \hat{q}(\phi;p)]f(\phi)d\phi$. Thus, type θ has an incentive to deviate if $\partial u(q^{MIN},\theta)/\partial q - (\lambda+1)p > 0$. The lowest incentive for an upward deviation has type $\underline{\theta}$, which characterizes the bound q^{MIN}. The upper bound, q^{MAX}, is obtained by a similar reasoning.

and

$$\partial u(\bar{q}, \theta_2)/\partial q - p[2 + (\lambda - 1)F(\theta_2)] = 0 \qquad \textit{if } \theta_2 < \bar{\theta}$$
$$\partial u(\bar{q}, \theta_2)/\partial q - p[2 + (\lambda - 1)F(\theta_2)] \leq 0 \qquad \textit{if } \theta_2 = \bar{\theta}.$$

For the parts where the personal equilibrium is strictly increasing $\hat{q}(\theta; p) = \tilde{q}(\theta; p)$.

The situation described in the above lemma is depicted in Figure 1.

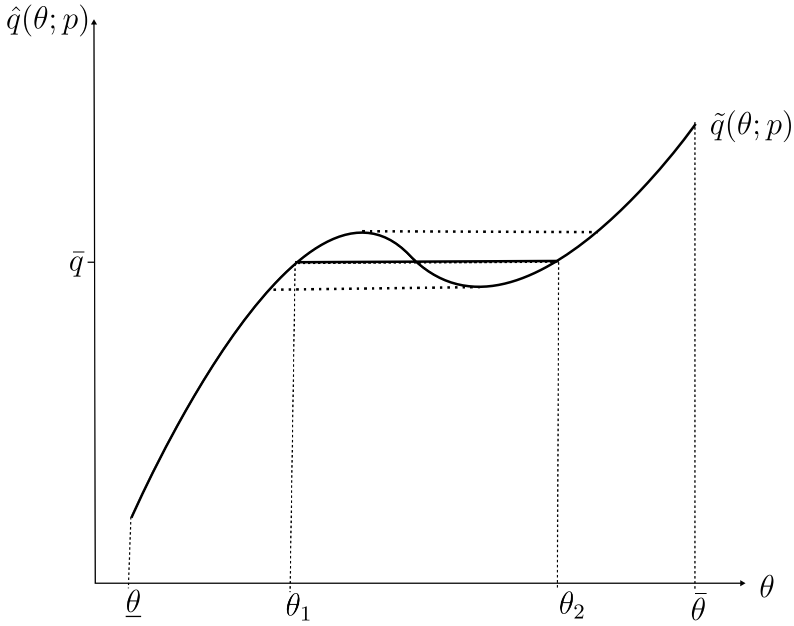

Figure IV.1.: Personal Equilibrium with Bunching

Without further assumptions on the utility function and the distribution function of the state of the world, the "bunching" regions are intricate to characterize. Moreover, if condition $(C1)$ does not hold and thus the personal equilibrium consists of flat parts, there are typically multiple personal equilibria. Since we are interested in situations where it is optimal for firms to offer flat-rate tariffs, there is no need to further discuss the personal equilibria for high per-unit prices. A flat-rate can only be optimal if marginal costs are not too high. If marginal costs are sufficiently low, however, then the gains from trade for high marginal prices such that bunching occurs are lower than the joint surplus generated from a flat rate. Hence, it is never optimal for a firm to set such high per unit prices. When we analyze firms' behavior we provide a sufficient condition that allows us to focus on the case where $(C1)$ is satisfied.

For illustrative purposes, we characterize all personal equilibria for a special case. Suppose the state of the world is uniformly distributed and that the cross derivative is constant. Then, $d\tilde{q}(\theta; p)/d\theta$ is either strictly increasing for all $\theta \in \Theta$ or non-increasing for all θ. Thus, depending on the per unit price, the personal equilibrium is either strictly increasing or constant over all states of the world.

Corollary IV.1: *Suppose* $\partial^2 u(q,\theta)/\partial q \partial\theta = K > 0$ *and* $\theta \sim U[\underline{\theta},\bar{\theta}]$. *Then (i) for* $p < K(\bar{\theta}-\underline{\theta})/(\lambda-1)$ *there exists a unique personal equilibrium which is characterized by* $\partial u(\hat{q}(\theta;p),\theta)/\partial q = p[2+(\lambda-1)(\theta-\underline{\theta})/(\bar{\theta}-\underline{\theta})]$, *(ii) for* $p \geq K(\bar{\theta}-\underline{\theta})/(\lambda-1)$ *in any personal equilibrium demand is independent of the state of the world, i.e.,* $\hat{q}(\theta;p) = \bar{q}$ *for all* $\theta \in [\underline{\theta},\bar{\theta}]$. *In this case there are multiple personal equilibria and* $\bar{q}$ *satisfies* $[\partial u(\bar{q},\bar{\theta})/\partial q]/(\lambda+1) \leq p \leq [\partial u(\bar{q},\underline{\theta})/\partial q]/2$.

4. Firm's Subproblem: Joint Surplus Maximization

For this section, suppose firms can observe consumers' types $\lambda \in \{\lambda_1,\lambda_2\}$. With the consumers' types being observable, the two market segments of types λ_1 and λ_2 can be viewed as distinct markets. Thus, for the analysis we can focus on one market where consumers are homogeneous with respect to their degree of loss aversion, which is denoted by λ.

Suppose firm $i \in \{A,B\}$ offers consumers a "deal" using a two-part tariff (L^i,p^i) that gives them utility v^i. Then, if a consumer with brand preferences $\zeta = (\zeta^0,\zeta^A,\zeta^B)$ purchases from firm i his net utility is $v^i+\zeta^i$. Let $\pi_j(v^i)$ be firm i's maximum profit per customer of type j when offering them a deal that yields utility v^i. The per consumer profit function is the same for both firms but it depends on the consumer's degree of loss aversion λ. In this section, we focus on one market segment and therefore the type subscript can be omitted without confusion. Since $\pi(\cdot)$ is the same for both firms, we will omit firm's superscript in the following. With this notation, $\pi(v)$ is given by the solution to the problem:

$$\pi(v) \quad = \quad \max_{L,p\geq 0} \quad : \quad \left\{ L + (p-c)\int_{\underline{\theta}}^{\bar{\theta}} \hat{q}(\theta,p)f(\theta)d\theta \;\middle|\; \mathbb{E}_\theta[U(\hat{q}(\theta,p)|\theta,\langle q(\phi,p)\rangle)] = v \right\}. \tag{IV.13}$$

First, we study the firm's subproblem, that is, we derive the optimal two-part tariff that solves the above problem. In the next section, we solve for the utility levels and the corresponding tariffs which are offered by the two firms in equilibrium. Now, we maximize a firm's profit over the choice variables p and L subject to the constraint that the consumer's utility from the offered deal is v. The constraint of (IV.13) can be written as

$$L = \int_{\underline{\theta}}^{\bar{\theta}} \Bigg\{ u(\hat{q}(\phi,p),\phi) - p\hat{q}(\phi,p) + p\int_{\phi}^{\bar{\theta}} [\hat{q}(\theta,p)-\hat{q}(\phi,p)]f(\theta)\,d\theta \\ - \lambda p \int_{\underline{\theta}}^{\phi} [\hat{q}(\phi,p)-\hat{q}(\theta,p)]f(\theta)\,d\theta \Bigg\} f(\phi)\,d\phi - v. \tag{IV.14}$$

Hence, the firm's tariff choice problem can be restated as a problem of choosing only a per unit price p. The firm chooses p to maximize $S(p) - v$, where

$$S(p) = \int_{\underline{\theta}}^{\bar{\theta}} \Big\{ u(\hat{q}(\phi,p),\phi) - c\hat{q}(\phi,p) + p\int_{\phi}^{\bar{\theta}} [\hat{q}(\theta,p) - \hat{q}(\phi,p)] f(\theta)\, d\theta \\ - \lambda p \int_{\underline{\theta}}^{\phi} [\hat{q}(\phi,p) - \hat{q}(\theta,p)] f(\theta)\, d\theta \Big\} f(\phi)\, d\phi. \quad \text{(IV.15)}$$

The firm chooses the marginal price p such that the joint surplus of the two contracting parties, the consumer and the firm, is maximized. The optimal marginal price $\hat{p}$ is independent of the utility the firm offers to the consumer v. This immediately implies that $\pi'(v) = -1$.

Given condition $(C1)$ does not hold then $S(p)$ is not unambiguously defined, since the personal equilibrium is not unique. Due to the next lemma one can focus on the case where there is a unique personal equilibrium given that the marginal cost is low. Let $\bar{p} := \min_{\theta}\{\kappa[(\lambda - 1)f(\theta)]^{-1}\}$. Note that if $p < \bar{p}$ then condition $(C1)$ is satisfied.

Lemma IV.6: *Suppose marginal cost, $c > 0$, is sufficiently low. Then the joint surplus, $S(p)$, is maximized for a unit price $p \in [0, \bar{p})$.*

The condition under which Lemma 6 is applicable is not very restrictive if one is prepared to assume that the distribution of the states of the world is not very dense. With $f(\theta)$ being small and thus $\bar{p}$ being high, the possible gains from trade with unit prices larger than $\bar{p}$ are small, since demand is decreasing in p. Keep in mind that the price at which "bunching" occurs, $\bar{p}$, is independent of marginal cost. If marginal cost is relatively low compared to the price $\bar{p}$, then unit prices $p \geq \bar{p}$ lead to higher distortions in demand than a unit price of zero compared to the first-best quantities. Because the firm tries to maximize the joint surplus, and $p = 0$ leads to no losses whereas prices $p \geq \bar{p}$ do, the optimal unit price is below $\bar{p}$ for not too high marginal cost. In all what follows it is assumed that c is such that Lemma 6 is applicable.

Hence, we can focus on the case where the personal equilibrium $\langle \hat{q}(\theta,p)\rangle_{\theta\in\Theta}$ is characterized by $\partial u(\hat{q}(\theta,p),\theta)/\partial q = p[2 + (\lambda - 1)F(\theta)]$. The derivative of the joint surplus with respect to the marginal price p is

$$S'(p) = \\ \int_{\underline{\theta}}^{\bar{\theta}} \Big\{ (p - c)\frac{d\hat{q}(\theta,p)}{dp} + p\int_{\theta}^{\bar{\theta}} \frac{d\hat{q}(\phi,p)}{dp} f(\phi)\, d\phi + \lambda p \int_{\underline{\theta}}^{\theta} \frac{d\hat{q}(\phi,p)}{dp} f(\phi)\, d\phi \Big\} f(\theta)\, d\theta \\ - (\lambda - 1)\int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi,p) - \hat{q}(\theta,p)] f(\phi) f(\theta)\, d\phi d\theta\,. \quad \text{(IV.16)}$$

The change in the demanded quantities due to a change in the marginal price is

$$\frac{d\hat{q}(\theta,p)}{dp} = \frac{2+(\lambda-1)F(\theta)}{\partial^2 u(\hat{q}(\theta,p),\theta)/\partial q^2} < 0 \,. \tag{IV.17}$$

Obviously, for unit prices $p \geq c$ the joint surplus is strictly decreasing in p. Thus, the optimal marginal price $\hat{p} \in [0, c)$. In order to guarantee that $S(p)$ is well behaved, we need an additional assumption. In this regard, we define

$$\Psi(p) \equiv (p-c)\int_{\underline{\theta}}^{\bar{\theta}} \frac{d\hat{q}(\theta,p)}{dp} f(\theta)d\theta - (\lambda-1)\int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta}^{\bar{\theta}} [\hat{q}(\phi,p)-\hat{q}(\theta,p)]f(\phi)f(\theta)\, d\phi d\theta \,.$$

Assumption (A2): *For $\lambda \in \{\lambda_1, \lambda_2\}$ and $p \in [0, c)$, $\Psi(p)$ is non-increasing in p.*

Assumption 2 is satisfied, for instance, if $d^2\hat{q}(p,\theta)/dp^2 \geq 0$ and $d^2\hat{q}(p,\theta)/dpd\theta \geq 0$. In particular, we have to rule out that a higher marginal price leads to a reduction in expected losses, due to a highly compressed demand profile. To cut back on our lengthy formulas we define

$$\Sigma(\lambda) \equiv (\lambda-1)\,\frac{\int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta}^{\bar{\theta}}[\hat{q}(\phi,0)-\hat{q}(\theta,0)]f(\phi)f(\theta)d\phi d\theta}{-\int_{\underline{\theta}}^{\bar{\theta}}[d\hat{q}(\theta,0)/dp]f(\theta)d\theta} \,.$$

Note that $\hat{q}(\theta, p)$ does also depend on λ. A few words to $\Sigma(\lambda)$ are in order, since this term is a crucial determinant for the optimality of flat-rate tariffs. Obviously, $\Sigma(1) = 0$. Moreover, it can be shown that $\Sigma(\cdot)$ is strictly increasing in λ. The numerator is a measure for the demand variation under a flat-rate contract. If the firm increases the unit price slightly above zero this reduces the consumer's utility by imposing an expected net loss on him, which is proportional to the numerator of $\Sigma(\lambda)$. Even a small increase in p above zero has a negative first-order impact on the consumer's gain-loss utility. The denominator measures how strong on average the consumer's demand reacts due to an increase of the unit price slightly above zero. Since a flat-rate contract leads to overconsumption which is costly, the firm has an incentive to choose a positive unit price if price increases cause sharp reductions in demand. To sum up, we can expect flat-rate tariffs to be optimal when $\Sigma(\lambda)$ is large. This is the case if either fluctuation in demand is high, or if demand reacts relatively inelastic on price changes.

With this notation, we are prepared to state the main result of this section.

Proposition IV.2: *Suppose (A2) holds. Then, the joint surplus of a firm and a consumer is maximized via a flat-rate tariff, i.e., $\hat{p} = 0$, if and only if $\Sigma(\lambda) \geq c$, where $\Sigma'(\lambda) > 0$.*

Economically, the proposition states that a flat-rate tariff is optimal when the marginal costs are sufficiently low compared to the consumer's degree of loss aversion. On the one hand, a flat-tariff eliminates losses on the side of the consumer, but, on the other hand, it leads to a too high consumption level. If marginal costs are sufficiently high, the negative effect due to overconsumption outweighs the positive effect of minimized losses. Moreover, a flat-rate tariff can only be optimal if there is enough variation in the consumer's demand. The numerator of $\Sigma(\lambda)$ is a measure for the degree of demand variation.[12]

Unfortunately, the proposition requires that condition $(A2)$ holds. It is worthwhile to point out, however, that a flat-rate tariff is optimal even if $(A2)$ does not hold given that the following condition is satisfied for all $p \in [0, c)$:

$$(\lambda - 1)\,\frac{\int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta}^{\bar{\theta}}[\hat{q}(\phi,p) - \hat{q}(\theta,p)]f(\phi)f(\theta)d\phi d\theta}{-\int_{\underline{\theta}}^{\bar{\theta}}[d\hat{q}(\theta,p)/dp]f(\theta)d\theta} > c\,.$$

Example: To illustrate the optimality of flat-rate tariffs we now discuss an example. Suppose the consumer's intrinsic utility for the good is given by $u(q,\theta) = \theta q - (1/2)q^2$ for $q \leq \theta = q^S(\theta)$ and $u(q,\theta) = (1/2)\theta^2$ otherwise. Let the state of the world be uniformly distributed on $[2, 3]$. Suppose the marginal cost of the firm $c = .05$. The demanded quantities in a personal equilibrium for this example are characterized by Corollary 1. For $p < (\lambda - 1)^{-1}$ the demand function is strictly increasing and given by $\hat{q}(\theta, p) = \theta[1 - p(\lambda - 1)] - 2p(2 - \lambda)$. For $p \geq (\lambda - 1)^{-1}$ the consumer's demand is independent of the state of the world. In this case, there are multiple personal equilibria. Here, it can easily be verified that the joint surplus is always maximized for prices below $(\lambda - 1)^{-1}$.

Nevertheless, we briefly characterize the joint surplus for all p values. For $p \geq (\lambda - 1)^{-1}$ it can be shown that the preferred personal equilibrium is to demand the highest possible quantity, i.e., $\bar{q} = \max\{2(1 - p), 0\}$. The preferred personal equilibrium is the plan among the consistent plans (personal equilibria) that maximizes the consumer's expected utility. Here, the preferred personal equilibrium is also optimal from the firm's perspective. The joint surplus, $S(p)$, is depicted below for the case $\lambda = 3$. Observe that $S(p)$ is continuous at $p = (\lambda - 1)^{-1}$ which is a general feature of the model and not due to the specific example.

In this example the function $\Sigma(\cdot)$ takes the following simple form, $\Sigma(\lambda) = (1/3)(\lambda - 1)(\lambda + 3)^{-1}$. Thus, by applying Proposition 2 a flat-rate tariff is optimal if $\lambda \geq 1.706$. Figure 3 depicts the joint surplus, $S(p)$, for $\lambda = 1, 1.5, 2, 3, 5$. Lower curves correspond to higher values of λ.

[12]Empirical studies about the flat-rate bias who support the so-called "ratio rule" often argue that a higher variance in the consumer's demand does not necessarily increase the consumer's preferences for a flat-rate option, see for instance Nunes (2000). Similarly, for a loss averse consumer an invoice profile is more risky if it has a higher average self distance (numerator of $\Sigma(\cdot)$), which does not imply a higher variance.

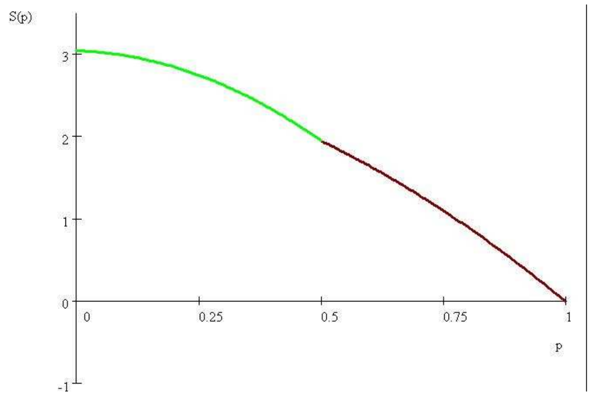

Figure IV.2.: Joint Surplus for $\lambda = 3$.

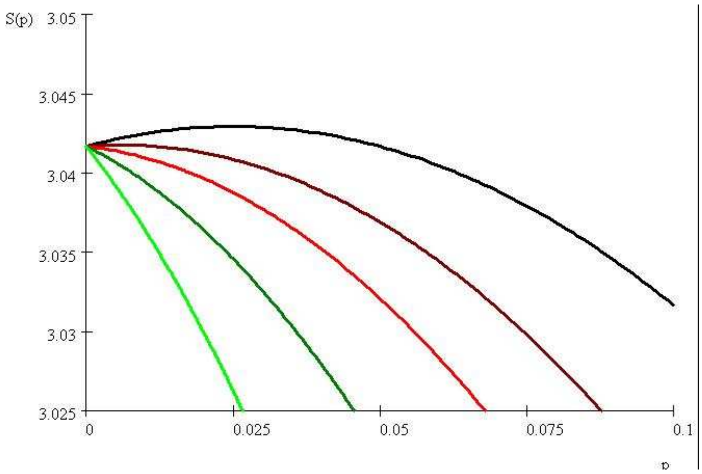

Figure IV.3.: Joint Surplus for $\lambda = 1, 1.5, 2, 3, 5$.

Without loss aversion ($\lambda = 1$) the optimal marginal price $p = (1/2)c$. Note that even for $\lambda = 1$ the consumer has reference-dependent preferences and therefore his marginal utility for money is two at the moment where he makes his purchasing decision. With the consumer being loss averse ($\lambda > 1$), in most cases a flat-rate tariff is optimal.

5. Competitive Equilibrium

Since firms are symmetric we focus on the profit maximization problem of firm A.

Symmetric Information Case

With λ being observable we can regard the two market segments as distinct markets. For a given market segment $j \in \{1, 2\}$ firm A offers utility v_j^A to these consumers to maximize profits. For a given utility v_j^B offered by firm B the profit maximization problem of firm A is given by

$$\max_{v_j^A} m(v_j^A, v_j^B)\pi_j(v_j^A) . \tag{IV.18}$$

The necessary first-order condition for profit maximization amounts to

$$[\partial m(v_j^A, v_j^B)/\partial v_j^A]\pi_j(v_j^A) + m(v_j^A, v_j^B)\pi_j'(v_j^A) = 0 . \tag{IV.19}$$

Remember that $\pi_j'(v^A) = -1$. The optimal marginal price is unaffected by the choice of v_j^A. If firm A offers one unit utility more to consumers, then this is optimally be achieved by lowering the fixed fee by one unit. The fixed fee is a one-to-one transfer from the consumer to the firm. Define

$$\Phi(v) \equiv \frac{m(v,v)}{\partial m(v,v)/\partial v^A} .$$

Applying Proposition 1 of Armstrong and Vickers (2001), the firm's per customer profit in submarket j in the symmetric equilibrium is given by

$$\pi_j(\hat{v}_j) = \Phi(\hat{v}_j) ,$$

where $\hat{v}_j$ denotes the utility offered by both firms to consumers of type λ_j in equilibrium. As is shown by Armstrong and Vickers, there are no asymmetric equilibria. Moreover, the equilibrium is often unique.[13] The following proposition summarizes the tariffs offered by the two firms to consumers in equilibrium.

Proposition IV.3 (Full Information): *Suppose (A2) holds. Then, in equilibrium, (i) if $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$ both firms offer the tariff $(\hat{p}, \hat{L})$ with a positive unit price to consumers of type λ_1, and a flat-rate tariff $(0, L^F)$ to consumers of type λ_2. (ii) If $c \leq \Sigma(\lambda_1) < \Sigma(\lambda_2)$, then both firms offer the flat-rate tariff $(0, L^F)$ to both types of loss averse consumers. The tariffs $(\hat{p}, \hat{L})$ and $(0, L^F)$ are characterized by: $S(\hat{p}) = 0$,*

$$\begin{aligned} \hat{L} &= \Phi(\hat{v}_1) - (\hat{p} - c)\int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_1(\theta, \hat{p}) f(\theta) d\theta \\ \textit{and} \qquad L^F &= \Phi(\hat{v}_2) + c\int_{\underline{\theta}}^{\bar{\theta}} q^S(\theta) f(\theta) d\theta , \end{aligned}$$

respectively, with $\hat{p} \in (0, c)$.

If the degree of loss aversion of the less loss-averse consumers is below the threshold, then firms offer to these consumers a usage-based pricing scheme. Next to the usage-based scheme firms offer a flat-rate tariff to the more loss-averse consumers. Thus,

[13] See Armstrong and Vickers (2001) for sufficient conditions for a unique equilibrium.

consumer heterogeneity with respect to their first-order risk preferences provides one possible answer why firms offer flat-rates next to usage-based tariffs. If the degree of loss aversion of both types is above the threshold, then firms offer only a single tariff, which is a flat-rate tariff. It is worthwhile to point out that the structure of the tariffs offered in equilibrium does not depend on the degree of competition. The degree of competition only influences the size of the fixed fee. In a more competitive market firms offer tariffs with lower fixed fees, the unit price is unaffected by the degree of competition. Even in the limit, when we approach a perfectly competitive market, the equilibrium tariffs do not converge to marginal cost pricing. Note that in this model the degree of competition (or the degree of product differentiation) is measured by $\Phi(\cdot)$. A lower $\Phi(\cdot)$ corresponds to a more competitive market. $\Phi(\cdot)$ is the inverse semi-elasticity of demand evaluated at the equilibrium utility level. Thus, the higher $\Phi(\cdot)$ the less elastic is the demand of a firm. To make this point even clearer, suppose firms are located at the two extreme points of a Hotelling line of length one. Consumers' ideal brands are uniformly distributed on this line. If a consumer incurs "transport cost" of t times the distance between his ideal brand and the firm he purchases from, then $\Phi(\hat{v}) = t$ given the market is fully covered in equilibrium.

A final comment to the offered tariffs is in order: Here, firms offer a flat-rate tariff to those consumers who are willing to pay an extra amount to be insured against unexpected high bills. The flat-rate tariff, however, is not offered to exploit consumers' behavioral bias. Here, firms offer flat-rate tariffs to consumers in the cases where these tariffs also maximize the joint surplus. This is in contrast to several models with biased consumers where firms design tariffs to exploit consumers' biases, see for instance Grubb (forthcoming) or Eliaz and Spiegler (2008).

For completeness, the following result states the equilibrium outcome for the case where consumers are not loss averse.

Corollary IV.2: *Suppose consumers do not exhibit loss aversion, i.e., $\lambda_1 = \lambda_2 = 1$. Then, in equilibrium both firms offer the two-part tariff with marginal price $\hat{p} = (1/2)c$ and fixed fee $\hat{L} = \Phi(\hat{v}) + (1/2)c\int_{\underline{\theta}}^{\bar{\theta}} q^{FB}(\theta)d\theta$, where $q^{FB}(\theta) \equiv arg\max_q\{u(q,\theta) - cq\}$. In this case the joint surplus equals the first-best surplus, $\hat{v} + \pi(\hat{v}) = S^{FB}$.*

Without loss aversion, due to ex ante contracting, firms choose a tariff that implements the first-best allocation. Depending on the degree of competition, the first-best surplus is shared between firms and consumers. Since with $\lambda = 1$ consumers still have reference-dependent preferences, the per-unit price does not equal marginal cost. Due to the reference-dependent preferences the consumer's marginal utility for money is two, since the weight on gain-loss utility is assumed to be one. It is important to point out that reference-dependent preferences without loss aversion have only quantitative effects on the equilibrium outcome but not qualitative effects.

Asymmetric Information Case

In this subsection, we investigate the tariffs offered by the two firms when facing a screening problem, i.e., the degree of loss aversion is private information of the consumer. The following lemma states that a consumer's expected utility is decreasing in his degree of loss aversion.

Lemma IV.7: *Consider a two-part tariff (p, L) and suppose $(C1)$ holds. Then,*

$$\frac{d}{d\lambda}\left[\mathbb{E}_\theta[U(\hat{q}(\theta;p)|\theta,\langle q(\theta;p)\rangle)]\right] \leq 0.$$

Suppose firms offer the tariffs as in the full information benchmark. Due to Lemma 7, consumers who are less loss averse may have an incentive to choose the tariff that is designed for the more loss-averse consumers. Note that when choosing a flat-rate tariff the consumer does neither feel a loss nor a gain. Thus, the expected utility from a flat-rate contract is independent of the consumer's degree of loss aversion. Hence, if firms offer a flat-rate tariff to the types with a high degree of loss aversion, then a consumer of type λ_1 does not necessarily benefit from choosing the tariff that is designed for consumers of type λ_2. If firms' profits from the market segment of λ_1 types is lower than their profits from λ_2 types, however, then consumers of type λ_2 may have an incentive to choose the tariff $(\hat{p}, \hat{L})$. Because, if this is the case, then $\hat{v}_1$ is considerably larger than $\hat{v}_2$. We rule this out by assuming that $\Phi(\cdot)$ is non-decreasing.[14] With this assumption, both firms' profits and consumers' surplus increase in equilibrium, if the joint surplus from contracting increases. Assuming that an increase in joint surplus is shared between consumers and firms seems to be natural for imperfectly competitive markets.

With this assumption, the two types of loss averse consumers may not exert any informational externality on each other. Put differently, if this is the case, firms can screen the consumer's type at no cost.

Proposition IV.4 (Asymmetric Information): *Suppose $(A2)$ holds and $\Phi'(v) \geq 0$. Then, (i) if $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$ both firms offering tariff $(\hat{p}, \hat{L})$ with a positive unit price to consumers of type λ_1, and flat-rate tariff $(0, L^F)$ to consumers of type λ_2 is an equilibrium. (ii) If $c \leq \Sigma(\lambda_1) < \Sigma(\lambda_2)$, then in equilibrium both firms offer the flat-rate tariff $(0, L^F)$ to both types of loss averse consumers. The tariffs, $(\hat{p}, \hat{L})$ and $(0, L^F)$, are characterized in Proposition 3.*

[14] For instance, this assumption is satisfied for the standard Hotelling model and the logit demand model, see Appendix A.4.2.

As in the symmetric information case, if λ_1 is below and λ_2 above the threshold, then firms offer a usage-based pricing scheme to the less loss-averse types and a flat-rate tariff to the more loss-averse consumers. The fixed fee of the flat-rate tariff is higher than the fixed fee of the usage-based pricing scheme. In this case, we do not make any claims about the uniqueness of this equilibrium.[15] If the degree of loss aversion of both types exceeds the threshold, then we obtain a pooling equilibrium: Each firm offers only a single tariff that is accepted by both types of consumers.

If $\Sigma(\lambda_2) < c$ then there is an information externality. In this case, if firms can observe λ they would offer to each type a different usage-based tariff. When offering these tariffs in the asymmetric information case, then type λ_1 obtains a higher expected utility from signing the contract that is designed for the types λ_2. We refrain from characterizing the equilibrium tariffs for this case, since this case is intricate to analyze in the applied competition-in-utility-space framework.

It is important to point out that offering a flat-rate tariff next to usage based tariffs does not impose some additional incentive constraints. If the degree of loss aversion, λ, is continuously distributed on $[\underline{\lambda}, \bar{\lambda}]$ with $\Sigma(\underline{\lambda}) < c < \Sigma(\bar{\lambda})$, then in equilibrium firms offer the flat-rate tariff $(0, L^F)$ which is chosen at least by types $\lambda \in [\tilde{\lambda}, \bar{\lambda}]$, where $\Sigma(\tilde{\lambda}) = c$.

6. Concluding Remarks

The goal of this article is to provide one possible explanation for the frequent usage of flat-rate tariffs. Since empirical evidence suggests that consumers choose flat rates because these tariffs provide insurance in case of uncertain consumption patterns, we posit that consumers are first-order risk averse. First-order risk aversion is captured by reference-dependent preferences of the consumer in combination with loss aversion. This paper shows that a flat-rate contract is optimal when consumers are sufficiently loss averse, marginal cost is small, or demand is rather unpredictable. Moreover, in an imperfectly competitive market firms offer flat-rate tariffs to those consumers whose degree of loss aversion exceeds a certain threshold. Consumers with a lower degree of loss aversion sign a metered tariff in equilibrium. Interestingly, offering a flat-rate contract next to usage-based pricing schemes does not introduce additional sorting constraints into a firm's optimization problem. Thus, this paper predicts that in markets with low marginal cost and uncertain consumption patterns, a firm's tariff menu includes a flat-rate option.

[15] To analyze all equilibria we cannot apply the competition in utility space framework, since we have to take the sorting constraints explicitly into account. Note that each firm has 4 choice variables which makes the calculation of firm A's best response to firm B's tariff offers intricate.

We departed from the Kőszegi-Rabin concept by positing that the consumer does not feel any sensations of gains and losses in the good dimension. If one considers gain-loss utility in both dimensions and assumes that higher states are always associated with higher utility, then a flat-rate tariff eliminates only the losses in the money but not in the good dimension. Depending on the particular form of the intrinsic utility function, a flat-rate contract may increase or decrease the expected losses in the good dimension. Alternatively, one could assume that intrinsic utility evaluated at the satiation quantity is constant for all states, but marginal utility is still increasing in the state. With this formulation, for a given state ϕ the consumer feels a loss in the good as well as in the money dimension compared to states $\theta < \phi$. In this case, a flat-rate tariff eliminates any losses in both dimensions; the good and the money dimension. Hence, the formulation of this paper can be viewed as an intermediate case between the two possible approaches with gain-loss utility in both dimensions. Moreover, focusing on the case with gain-loss utility only in the money dimension helps to make the analysis of the personal equilibria clearer and shorter.

An obvious drawback of our model is that firms are restricted to two-part tariffs. With the consumer being loss averse according to Kőszegi-Rabin, his utility in a given state of the world also depends on his payments in all other states. Thus, the standard procedure of the nonlinear pricing literature, where the tariff in the firm's objective typically is replaced by the consumer's net surplus, does not work here. This in turn makes the analysis of nonlinear tariffs more complicated. We believe, however, that focusing on two-part tariffs provides some insights on the forces at play when consumers are loss averse. In particular, the identified insurance motive of loss averse consumers should also play a major role when firms can offer more sophisticated contracts.

It is rather obvious that imposing a quantity limit on the flat-rate option can improve the joint surplus. If the quantity limit equals the first-best quantity for the highest state then, compared to a flat rate with unlimited usage, standard efficiency is improved without imposing additional losses on the consumer. Flat-rate tariffs with limited usage are often observed for the Internet service industry. Hence, investigation of optimal nonlinear pricing schedules for firms facing loss-averse consumers is an interesting question for future research.

A. Appendices

1. Appendix to Chapter I

1.1. Proofs of Propositions and Lemmas

Proof of Lemma I.1: Suppose that signals are ordered according to their likelihood ratio, that is, $s > s'$ if and only if $\gamma_s^H/\gamma_s^L > \gamma_{s'}^H/\gamma_{s'}^L$. Consider a contract of the form

$$u_s = \begin{cases} \underline{u} & \text{if} \quad s < \hat{s} \\ \underline{u} + b & \text{if} \quad s \geq \hat{s} \end{cases},$$

where $b > 0$ and $1 < \hat{s} \leq S$. Under this contractual form and given that the first-order approach is valid, (IC) can be rewritten as

$$b \Bigg\{ \left[\sum_{s=\hat{s}}^{S} (\gamma_s^H - \gamma_s^L) \right] \left(1 - (\lambda - 1) \sum_{s=1}^{\hat{s}-1} \gamma_s(\hat{a}) \right) - (\lambda - 1) \left(\sum_{s=1}^{\hat{s}-1} (\gamma_s^H - \gamma^L) \right) \left(\sum_{s=\hat{s}}^{S} \gamma_s(\hat{a}) \right) \Bigg\} = c'(\hat{a}).$$

Since signals are ordered according to their likelihood ratio, we have $\sum_{s=\hat{s}}^{S} (\gamma_s^H - \gamma_s^L) > 0$ and $\sum_{s=1}^{\hat{s}-1} (\gamma_s^H - \gamma^L) < 0$ for all $1 < \hat{s} \leq S$. This implies that the term in curly brackets is strictly positive for $\lambda \leq 2$. Hence, with $c'(\hat{a}) > 0$, b can alway be chosen such that (IC) is met. Rearranging the participation constraint,

$$\underline{u} \geq \bar{u} + c(\hat{a}) - b \left(\sum_{s=\hat{s}}^{S} \gamma_s(\hat{a}) \right) \left[1 - (\lambda - 1) \left(\sum_{s=1}^{\hat{s}-1} \gamma_s(\hat{a}) \right) \right],$$

reveals that (IR) can be satisfied for any b by choosing $\underline{u}$ appropriately. This concludes the proof.

Q.E.D.

Proof of Proposition I.1: It is readily verified that Assumptions 1-3 from Grossman and Hart (1983) are satisfied. Thus, the cost-minimization problem is well defined, in the sense that for each action $a \in (0, 1)$ there exists a second-best incentive scheme.

Suppose the principal wants to implement action $\hat{a} \in (0,1)$ at minimum cost. Since the agent's action is not observable, the principal's problem is given by

$$\min_{\{u_s\}_{s=1}^{S}} \sum_{s=1}^{S} \gamma_s(\hat{a}) h(u_s) \tag{MR}$$

subject to

$$\sum_{s=1}^{S} \gamma_s(\hat{a}) u_s - c(\hat{a}) \geq \bar{u}\,, \tag{IR$_R$}$$

$$\sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) u_s - c'(\hat{a}) = 0\,. \tag{IC$_R$}$$

where the first constraint is the individual rationality constraint and the second is the incentive compatibility constraint. Note that the first-order approach is valid, since the agent's expected utility is a strictly concave function of his effort. The Lagrangian to the resulting problem is

$$\mathcal{L} = \sum_{s=1}^{S} \gamma_s(a) h(u_s) - \mu_0 \left\{ \sum_{s=1}^{S} \gamma_s(a) u_s - c(a) - \bar{u} \right\} - \mu_1 \left\{ \sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) u_s - c'(a) \right\},$$

where μ_0 and μ_1 denote the Lagrange multipliers of the individual rationality constraint and the incentive compatibility constraint, respectively. Setting the partial derivative of $\mathcal{L}$ with respect to u_s equal to zero yields

$$\frac{\partial \mathcal{L}}{\partial u_s} = 0 \iff h'(u_s) = \mu_0 + \mu_1 \frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}, \quad \forall s \in \mathcal{S}. \tag{A.1}$$

Irrespective of the value of μ_0, if $\mu_1 > 0$, convexity of $h(\cdot)$ implies that $u_s > u_{s'}$ if and only if $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) > (\gamma_{s'}^H - \gamma_{s'}^L)/\gamma_{s'}(\hat{a})$, which in turn is equivalent to $\gamma_s^H/\gamma_s^L > \gamma_{s'}^H/\gamma_{s'}^L$. Thus it remains to show that μ_1 is strictly positive. Suppose, in contradiction, that $\mu_1 \leq 0$. Consider the case $\mu_1 = 0$ first. From (A.1) it follows that $u_s = u^f$ for all $s \in \mathcal{S}$, where u^f satisfies $h'(u^f) = \mu_0$. This, however, violates (IC$_R$), a contradiction. Next, consider $\mu_1 < 0$. From (A.1) it follows that $u_s < u_{s'}$ if and only if $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) > (\gamma_{s'}^H - \gamma_{s'}^L)/\gamma_t(\hat{a})$. Let $\mathcal{S}^+ \equiv \{s | \gamma_s^H - \gamma_s^L > 0\}$, $\mathcal{S}^- \equiv \{s | \gamma_s^H - \gamma_s^L < 0\}$, and $\hat{u} \equiv \min\{u_s | s \in \mathcal{S}^-\}$. Since $\hat{u} > u_s$ for all $s \in \mathcal{S}^+$, we have

$$\begin{aligned}
\sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) u_s &= \sum_{\mathcal{S}^-} (\gamma_s^H - \gamma_s^L) u_s + \sum_{\mathcal{S}^+} (\gamma_s^H - \gamma_s^L) u_s \\
&< \sum_{\mathcal{S}^-} (\gamma_s^H - \gamma_s^L) \hat{u} + \sum_{\mathcal{S}^+} (\gamma_s^H - \gamma_s^L) \hat{u} \\
&= \hat{u} \sum_{s=1}^{S} (\gamma_s^H - \gamma_s^L) \\
&= 0,
\end{aligned}$$

again a contradiction to (IC_R). Hence, $\mu_1 > 0$ and the desired result follows.

Q.E.D.

Proof of Proposition I.2: The problem of finding the optimal contract $\boldsymbol{u}^*$ to implement action $\hat{a} \in (0,1)$ is decomposed into two subproblems. First, for a given incentive feasible ordering of signals, we derive the optimal nondecreasing incentive scheme that implements action $\hat{a} \in (0,1)$. Then, in a second step, we choose the ordering of signals for which the ordering specific cost of implementation is lowest.

Step 1: Remember that the ordering of signals is incentive feasible if $\beta_s(\cdot) > 0$ for at least one signal s. For a given incentive feasible ordering of signals, in this first step we solve Program ML. First, note that it is optimal to set $b_s = 0$ if $\beta_s(\cdot) < 0$. To see this, suppose, in contradiction, that in the optimum (IC$'$) holds and $b_s > 0$ for some signal s with $\beta_s(\cdot) \leq 0$. If $\beta_s(\cdot) = 0$, then setting $b_s = 0$ leaves (IC$'$) unchanged, but leads to a lower value of the objective function of Program ML, contradicting that the original contract is optimal. If $\beta_s(\cdot) < 0$, then setting $b_s = 0$ not only reduces the value of the objective function, but also relaxes (IC$'$), which in turn allows to lower other bonus payments, thereby lowering the value of the objective function even further. Again, a contradiction to the original contract being optimal. Let $\mathcal{S}_\beta \equiv \{s \in \mathcal{S} | \beta_s(\cdot) > 0\}$ denote the set of signals for which $\beta_s(\cdot)$ is strictly positive under the considered ordering of signals, and let S_β denote the number of elements in this set. Thus, Program (ML) can be rewritten as

Program ML$^+$:

$$\min_{\{b_s\}_{s\in\mathcal{S}_\beta}} \sum_{s\in\mathcal{S}_\beta} b_s \rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$$

$$\text{subject to} \quad (i) \sum_{s\in\mathcal{S}_\beta} b_s \beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \qquad (\text{IC}^+)$$

$$(ii) \;\; b_s \geq 0, \;\; \forall s \in \mathcal{S}_\beta \,.$$

Program ML$^+$ is a linear programming problem. It is well-known that if a linear programming problem has a solution, it must have a solution at an extreme point of the constraint set. Generically, there is a unique solution and this solution is an extreme point. Since the constraint set of Program ML$^+$, $\mathcal{M} \equiv \{\{b_s\}_{s\in\mathcal{S}_\beta} \in \mathbb{R}_+^{S_\beta} | \sum_{s\in\mathcal{S}_\beta} b_s \beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a})\}$, is closed and bounded, Program ML$^+$ has a solution. Hence, generically $\sum_{s\in\mathcal{S}_\beta} b_s \rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$ achieves its greatest lower bound at one of the extreme points of $\mathcal{M}$. (We comment on genericity below.) With $\mathcal{M}$ describing a hyperplane in $\mathbb{R}_+^{S_\beta}$, all extreme points of $\mathcal{M}$ are characterized by the following property: $b_s > 0$ for exactly one signal $s \in S_\beta$ and $b_t = 0$ for all $t \in S_\beta$, $t \neq s$. It remains to determine for which signal the bonus is set strictly positive. The size of the bonus payment, which is set strictly positive, is uniquely determined by (IC$^+$):

$$b_s \beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \iff b_s = \frac{c'(\hat{a})}{\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})}. \tag{A.2}$$

Therefore, from the objective function of Program ML$^+$ it follows that, for the signal ordering under consideration, the optimal signal for which the bonus is set strictly positive, $\hat{s}$, is characterized by

$$\hat{s} = \arg\min_{s \in \mathcal{S}_\beta} \frac{c'(\hat{a})}{\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})} \rho_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}).$$

Step 2: From all incentive feasible signal orders, the principal chooses the one which minimizes her cost of implementation. With the number of incentive feasible signal orders being finite, this problem clearly has a solution. Let s^* denote the resulting cutoff, i.e.,

$$u_s^* = \begin{cases} u^* & \text{if} \quad s < s^* \\ u^* + b^* & \text{if} \quad s \geq s^* \end{cases},$$

where $b^* = c'(\hat{a})/\beta_{s^*}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$ and $u^* = \bar{u} + c(\hat{a}) - b^* \left[\sum_{\tau=s^*}^{S} \gamma_\tau(\hat{a}) - \rho_{s^*}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})\right]$. Letting $u_L^* = u^*$, $u_H^* = u^* + b^*$, and $\mathcal{B}^* = \{s \in \mathcal{S} | s \geq s^*\}$ establishes the desired result.

On genericity: We claimed that, for any given feasible ordering of signals, generically Program ML$^+$ has a unique solution at one of the extreme points of the constraint set. To see this, note that a necessary condition for the existence of multiple solutions is $\beta_s/\beta_{s'} = \rho_s/\rho_{s'}$ for some $s, s' \in \mathcal{S}_\beta$, $s \neq s'$. This condition is characterized by the action to be implemented, $\hat{a}$, the structure of the performance measure, $\{(\gamma_s^H, \gamma_s^L)\}_{s=1}^S$, and the agent's degree of loss aversion, λ. Now, fix $\hat{a}$ and $\{(\gamma_s^H, \gamma_s^L)\}_{s=1}^S$. With both $\beta_s > 0$ and $\rho_s > 0$ for all $s \in \mathcal{S}_\beta$, it is readily verified, that exactly one value of λ equates $\beta_s/\beta_{s'}$ with $\rho_s/\rho_{s'}$. Since λ is drawn from the interval $(1, 2]$, and with the number of signals being finite, this necessary condition for Program ML$^+$ having multiple solutions for a given feasible ordering of signals generically will not hold. With the number of feasible orderings being finite, generic optimality of a corner solution carries over to the overall problem.

Q.E.D.

Proof of Proposition I.3: $\mathcal{B}^*$ maximizes $X(\mathcal{B}) := \left[\sum_{s \in \mathcal{B}} (\gamma_s^H - \gamma_s^L)\right] \times Y(P_\mathcal{B})$, where

$$Y(P_\mathcal{B}) := \frac{1}{(\lambda - 1) P_\mathcal{B}(1 - P_\mathcal{B})} - \frac{1}{P_\mathcal{B}} + \frac{1}{1 - P_\mathcal{B}}.$$

Suppose for the moment that $P_{\mathcal{B}}$ is a continuous decision variable. Accordingly,

$$\frac{dY(P_{\mathcal{B}})}{dP_{\mathcal{B}}} = \frac{1}{P_{\mathcal{B}}^2(1-P_{\mathcal{B}})^2}\left[2P_{\mathcal{B}}^2 + \frac{2-\lambda}{\lambda-1}(2P_{\mathcal{B}}-1)\right] . \tag{A.3}$$

It is readily verified that $dY(P_{\mathcal{B}})/dP_{\mathcal{B}} < 0$ for $0 < P_{\mathcal{B}} < \bar{P}(\lambda)$ and $dY(P_{\mathcal{B}})/dP_{\mathcal{B}} > 0$ for $\bar{P}(\lambda) < P_{\mathcal{B}} < 1$, where

$$\bar{P}(\lambda) \equiv \frac{\lambda - 2 + \sqrt{\lambda(2-\lambda)}}{2(\lambda-1)} .$$

Note that for $\lambda \leq 2$ the critical value $\bar{P}(\lambda) \in [0, 1/2)$. Hence, excluding a signal of $\mathcal{B}$ increases $Y(P_{\mathcal{B}})$ if $P_{\mathcal{B}} < \bar{P}(\lambda)$, whereas including a signal to $\mathcal{B}$ increases $Y(P_{\mathcal{B}})$ if $P_{\mathcal{B}} \geq \bar{P}(\lambda)$. With these insights the next two implications follow immediately.

(i) $P_{\mathcal{B}^*} < \bar{P}(\lambda) \implies \mathcal{B}^* \subseteq \mathcal{S}^+$

(ii) $P_{\mathcal{B}^*} \geq \bar{P}(\lambda) \implies \mathcal{S}^+ \subseteq \mathcal{B}^*$

We prove both statements in turn by contradiction. (i) Suppose $P_{\mathcal{B}^*} < \bar{P}(\lambda)$ and that there exists a signal $\hat{s} \in \mathcal{S}^-$ which is also contained in $\mathcal{B}^*$, i.e., $\hat{s} \in \mathcal{B}^*$. Clearly, $\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) < \sum_{s\in\mathcal{B}^*\backslash\{\hat{s}\}}(\gamma_s^H - \gamma_s^L)$ because $\hat{s}$ is a bad signal. Moreover, $Y(\mathcal{B}^*) < Y(\mathcal{B}^*\backslash\{\hat{s}\})$ because $Y(\cdot)$ increases when signals are excluded of $\mathcal{B}^*$. Thus $X(\mathcal{B}^*) < X(\mathcal{B}^*\backslash\{\hat{s}\})$, a contradiction to the assumption that $\mathcal{B}^*$ is the optimal partition. (ii) Suppose $P_{\mathcal{B}^*} \geq \bar{P}(\lambda)$ and that there exists a signal $\tilde{s} \in \mathcal{S}^+$ that is not contained in $\mathcal{B}^*$, i.e., $\mathcal{B}^* \cap \{\tilde{s}\} = \emptyset$. Since $\hat{s}$ is a good signal $\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) < \sum_{s\in\mathcal{B}^*\cup\{\hat{s}\}}(\gamma_s^H - \gamma_s^L)$. $P_{\mathcal{B}^*} \geq \bar{P}(\lambda)$ implies that $Y(\mathcal{B}^* \cup \{\tilde{s}\}) > Y(\mathcal{B}^*)$. Thus, $X(\mathcal{B}^*) < X(\mathcal{B}^* \cup \{\tilde{s}\})$ a contradiction to the assumption that $\mathcal{B}^*$ maximizes $X(\mathcal{B}^*)$. Finally, since for any $\mathcal{B}^*$ we are either in case (i) or in case (ii), the desired result follows.

Q.E.D.

Proof of Proposition I.4: Suppose, in contradiction, that in the optimum there are signals $s, t \in \mathcal{S}$ such that $s \in \mathcal{B}^*$, $t \notin \mathcal{B}^*$ and $\frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})} < \frac{\gamma_t^H - \gamma_t^L}{\gamma_t(\hat{a})}$. We derive a contradiction by showing that exchanging signal s for signal t reduces the principal's cost, which implies that the original contract cannot be optimal. Let $\bar{\mathcal{B}} \equiv (\mathcal{B}^* \setminus \{s\}) \cup \{t\}$.It suffices to show that $X(\bar{\mathcal{B}}) > X(\mathcal{B}^*)$, where $X(\mathcal{B})$ is defined as in the proof of Proposition 3. $X(\bar{\mathcal{B}}) > X(\mathcal{B}^*)$ is equivalent to

$$\left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H - \gamma_j^L) + (\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] >$$
$$\left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H - \gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\mathcal{B}^*})}{(\lambda-1)P_{\mathcal{B}^*}(1-P_{\mathcal{B}^*})}\right] .$$

Rearranging yields

$$\left[(\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L)\right] \left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] > \left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H - \gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\mathcal{B}^*})}{(\lambda-1)P_{\mathcal{B}^*}(1-P_{\mathcal{B}^*})} - \frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] . \quad (A.4)$$

With $Y(P_{\mathcal{B}})$ being defined as in the proof of Proposition 3, we have to consider two cases, (i) $dY(P_{\mathcal{B}^*})/P_{\mathcal{B}} \geq 0$, and (ii) $dY(P_{\mathcal{B}^*})/P_{\mathcal{B}} < 0$.

Case (i): Since $\gamma_s(\hat{a})-\gamma_t(\hat{a}) \leq \kappa$, we have $P_{\mathcal{B}^*} \leq P_{\bar{\mathcal{B}}}+\kappa$. With $Y(P_{\mathcal{B}})$ being (weakly) increasing at $P_{\mathcal{B}^*}$, inequality (A.4) is least likely to hold for $P_{\mathcal{B}^*} = P_{\bar{\mathcal{B}}} + \kappa$. Inserting $P_{\mathcal{B}^*} = P_{\bar{\mathcal{B}}} + \kappa$ into (A.4) yields

$$\left[(\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L)\right] \left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] > \left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H - \gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}}-2\kappa)}{(\lambda-1)[P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})+\kappa(1-2P_{\bar{\mathcal{B}}})]-\kappa^2} - \frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] . \quad (A.5)$$

The right-hand side of (A.5) becomes arbitrarily close to zero for $\kappa \to 0$, thus it remains to show that

$$\left[(\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L)\right] \left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right] > 0 . \quad (A.6)$$

For (A.6) to hold, we must have $(\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L) > 0$. From the proof of Proposition 3 we know that $\mathcal{S}^+ \subseteq \mathcal{B}^*$ if $Y(P_{\mathcal{B}})$ is increasing at $\mathcal{B}^*$. Since the principal will end up including all good signals in the set $\mathcal{B}^*$ anyway, the question of interest is whether she can benefit from swapping two bad signals. Therefore, we consider case $s, t \in \mathcal{S}^-$, where $\mathcal{S}^- \equiv \{s \in \mathcal{S} | \gamma_s^H - \gamma_s^L < 0\}$. With $s, t \in \mathcal{S}^-$, we have

$$\left[(\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L)\right] \geq \gamma_t(\hat{a})\gamma_s(\hat{a}) \left[\frac{1}{\gamma_s(\hat{a})}\frac{\gamma_t^H - \gamma_t^L}{\gamma_t(\hat{a})} - \frac{1}{\gamma_s(\hat{a})+\kappa}\frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}\right] , \quad (A.7)$$

where the inequality holds because $\gamma_t(\hat{a}) - \gamma_s(\hat{a}) \leq \kappa$. Note that for $\kappa \to 0$ the right-hand side of (A.7) becomes strictly positive, thus $(\gamma_t^H - \gamma_t^L) - (\gamma_s^H - \gamma_s^L) > 0$ for $\kappa \to 0$. Hence, for κ sufficiently small, $X(\mathcal{B}^*) < X(\bar{\mathcal{B}})$, a contradiction to $\mathcal{B}^*$ being optimal.

Case (ii): Since $\gamma_t(\hat{a}) - \gamma_s(\hat{a}) \leq \kappa$, we have $P_{\mathcal{B}^*} \geq P_{\bar{\mathcal{B}}} - \kappa$. With $Y(P_{\mathcal{B}})$ being decreasing at $P_{\mathcal{B}^*}$, inequality (A.4) is least likely to hold for $P_{\mathcal{B}^*} = P_{\bar{\mathcal{B}}} - \kappa$. Inserting $P_{\mathcal{B}^*} = P_{\bar{\mathcal{B}}} - \kappa$ into (A.4), and running along the lines of case (i) allows us to establish that, for κ sufficiently small, $X(\mathcal{B}^*) < X(\bar{\mathcal{B}})$, a contradiction to $\mathcal{B}^*$ being optimal.

To sum up, for κ sufficiently small we have

$$\max_{s\in\mathcal{S}\setminus\mathcal{B}^*}\{(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a})\} < \min_{s\in\mathcal{B}^*}\{(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a})\} ,$$

or equivalently,

$$\max_{s\in\mathcal{S}\setminus\mathcal{B}^*}\{\gamma_s^H/\gamma_s^L\} < \min_{s\in\mathcal{B}^*}\{\gamma_s^H/\gamma_s^L\}\ .$$

Letting $K \equiv \min_{s\in\mathcal{B}^*}\{\gamma_s^H/\gamma_s^L\}$ establishes the desired result.

Q.E.D.

Proof of Proposition I.5: We first prove part (ii). Suppose that a small change in λ leaves the optimal partition $\mathcal{B}^*$ of the set of all signals unchanged. Rearranging (IC$'$) yields

$$b^* = \frac{c'(\hat{a})}{\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) - (\lambda-1)\left[\sum_{s\in\mathcal{B}^*}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\mathcal{B}^*}]}. \tag{A.8}$$

Straight-forward differentiation reveals that

$$\frac{db^*}{d\lambda} = \frac{c'(\hat{a})\left[\sum_{s\in\mathcal{B}^*}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\mathcal{B}^*}]}{\left\{\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) - (\lambda-1)\left[\sum_{s\in\mathcal{B}^*}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\mathcal{B}^*}]\right\}^2}.$$

Since under the second-best contract $\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) > 0$, the desired result follows. To prove part (i), let $\mathcal{B}^+ \equiv \left\{\mathcal{B}\subset\mathcal{S} \,|\, \sum_{s\in\mathcal{B}}(\gamma_s^H-\gamma_s^L) > 0\right\}$. For any $\tilde{\mathcal{B}} \in \mathcal{B}^+$, let

$$b_{\tilde{\mathcal{B}}} = \frac{c'(\hat{a})}{\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H - \gamma_s^L) - (\lambda-1)\left[\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H-\gamma_s^L)\right][1-2P_{\tilde{\mathcal{B}}}]}$$

and

$$\underline{u}_{\tilde{\mathcal{B}}} = \bar{u} + c(\hat{a}) - b_{\tilde{\mathcal{B}}}P_{\tilde{\mathcal{B}}} + (\lambda-1)P_{\tilde{\mathcal{B}}}(1-P_{\tilde{\mathcal{B}}})b_{\tilde{\mathcal{B}}}.$$

The cost of implementing action $\hat{a}$ when paying $\underline{u}_{\tilde{\mathcal{B}}}$ for signals in $\mathcal{S}\setminus\tilde{\mathcal{B}}$ and $\underline{u}_{\tilde{\mathcal{B}}} + b_{\tilde{\mathcal{B}}}$ for signals in $\tilde{\mathcal{B}}$ is given by

$$C_{\tilde{\mathcal{B}}} \;=\; \underline{u}_{\tilde{\mathcal{B}}} + b_{\tilde{\mathcal{B}}}P_{\tilde{\mathcal{B}}} \;=\; \bar{u} + c(\hat{a}) + \frac{c'(\hat{a})(\lambda-1)P_{\tilde{\mathcal{B}}}(1-P_{\tilde{\mathcal{B}}})}{\left[\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H-\gamma_s^L)\right][1-(\lambda-1)(1-2P_{\tilde{\mathcal{B}}})]}. \tag{A.9}$$

Differentiation of $C_{\tilde{\mathcal{B}}}$ with respect to λ yields

$$\frac{dC_{\tilde{\mathcal{B}}}}{d\lambda} = \frac{c'(\hat{a})P_{\tilde{\mathcal{B}}}(1-P_{\tilde{\mathcal{B}}})}{\left[\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H-\gamma_s^L)\right][1-(\lambda-1)(1-2P_{\tilde{\mathcal{B}}})]^2}.$$

Obviously, $dC_{\tilde{\mathcal{B}}}/d\lambda > 0$ for all $\mathcal{B}\in\mathcal{B}^+$. Since the optimal partition of $\mathcal{S}$ may change as λ changes, the minimum cost of implementing action $\hat{a}$ is given by

$$C(\hat{a}) = \min_{\mathcal{B}\in\mathcal{B}^+} C_{\mathcal{B}}.$$

Put differently, $C(\hat{a})$ is the lower envelope of all $C_{\mathcal{B}}$ for $\mathcal{B}\in\mathcal{B}^+$. With $C_{\mathcal{B}}$ being continuous and strictly increasing in λ for all $\mathcal{B}\in\mathcal{B}^+$, it follows that also $C(\hat{a})$ is continuous and strictly increasing in λ. This completes the proof.

Q.E.D.

Proof of Lemma I.2: We show that program (MG) has a solution, i.e., $\sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s)$ achieves its greatest lower bound. First, from Lemma 1 we know that the constraint set of program (MG) is not empty for action $\hat{a}\in(0,1)$. Next, note that from (IR_G) it follows that $\sum_{s=1}^{S}\gamma_s(\hat{a})u_s$ is bounded below. Following the reasoning in the proof of Proposition 1 of Grossman and Hart (1983), we can artificially bound the constraint set – roughly spoken because unbounded sequences in the constraint set make $\sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s)$ tend to infinity by a result from Bertsekas (1974). Since the constraint set is closed, the existence of a minimum follows from Weierstrass' theorem.

Q.E.D.

Proof of Lemma I.3: Since (IR_G) will always be satisfied with equality due to an appropriate adjustment of the lowest intrinsic utility level offered, relaxing (IR_G) will always lead to strictly lower costs for the principal. Therefore, the shadow value of relaxing (IR_G) is strictly positive, so $\mu_{IR}>0$.

Next, we show that relaxing (IC_G) has a positive shadow value, $\mu_{IC}>0$. We do this by showing that a decrease in $c'(\hat{a})$ leads to a reduction in the principal's minimum cost of implementation. Let $\{u_s^*\}_{s\in\mathcal{S}}$ be the optimal contract under (the original) Program MG, and suppose that $c'(\hat{a})$ decreases. Now the principal can offer a new contract $\{u_s^N\}_{s\in\mathcal{S}}$ of the form

$$u_s^N=\alpha u_s^*+(1-\alpha)\sum_{t=1}^{S}\gamma_t(\hat{a})u_t^*\ , \tag{A.10}$$

where $\alpha\in(0,1)$, which also satisfies (IR_G), the relaxed (IC_G), and (OC_G), but yields strictly lower costs of implementation than the original contract $\{u_s^*\}_{s\in\mathcal{S}}$.

Clearly, for $\hat{\alpha}\in(0,1)$, $u_s^N<u_{s'}^N$ if and only if $u_s^*<u_{s'}^*$, so (OC_G) is also satisfied under contract $\{u_s^N\}_{s\in\mathcal{S}}$.

Next, we check that the relaxed (IC_G) holds under $\{u_s^N\}_{s\in\mathcal{S}}$. To see this, note that for $\alpha=1$ we have $\{u_s^N\}_{s\in\mathcal{S}}\equiv\{u_s^*\}_{s\in\mathcal{S}}$. Thus, for $\alpha=1$, the relaxed (IC_G) is oversatisfied under $\{u_s^N\}_{s\in\mathcal{S}}$. For $\alpha=0$, on the other hand, the left-hand side of (IC_G) is equal to zero, and the relaxed (IC_G) in consequence is not satisfied. Since the left-hand side of (IC_G) is continuous in α under contract $\{u_s^N\}_{s\in\mathcal{S}}$, by the intermediate-value theorem there exists $\hat{\alpha}\in(0,1)$ such that the relaxed (IC_G) is satisfied with equality.

Last, consider (IR_G). The left-hand side of (IR_G) under contract $\{u_s^N\}_{s\in\mathcal{S}}$ with $\alpha=\hat{\alpha}$

amounts to

$$\begin{aligned}
&\sum_{s=1}^{S}\gamma_s(\hat{a})u_s^N - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^N - u_s^N\right] \\
&= \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - \tilde{\alpha}(\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&> \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&= \bar{u} + c(\hat{a}) \ , \qquad \text{(A.11)}
\end{aligned}$$

where the last equality follows from the fact that $\{u_s^*\}_{s\in\mathcal{S}}$ fulfills the (IR$_g$) with equality. Thus, contract $\{u_s^N\}_{s\in\mathcal{S}}$ is feasible in the sense that all constraints of program (MG) are met. It remains to show that the principal's costs are reduced. Since $h(\cdot)$ is strictly convex, the principal's objective function is strictly convex in α, with a minimum at $\alpha = 0$. Hence, the principal's objective function is strictly increasing in α for $\alpha \in (0, 1]$. Since $\{u_s^N\}_{s\in\mathcal{S}} \equiv \{u_s^*\}_{s\in\mathcal{S}}$ for $\alpha = 1$, for $\alpha = \hat{\alpha}$ we have

$$\sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s^*) > \sum_{s=1}^{S}\gamma_s(\hat{a})h(u_s^N),$$

which establishes the desired result.

Q.E.D.

Proof of Proposition I.6: For the agent's intrinsic utility function being sufficiently linear, the principal's costs are approximately given by a second-order Taylor polynomial about $r = 1$, thus

$$C(\boldsymbol{u}|r) \approx \sum_{s\in\mathcal{S}}\gamma_s(\hat{a})u_s + \Omega(\boldsymbol{u}|r) \ , \qquad \text{(A.12)}$$

where

$$\Omega(\boldsymbol{u}|r) \equiv \sum_{s\in\mathcal{S}}\gamma_s(\hat{a})\left[(u_s \ln u_s)(r-1) + (1/2)u_s(\ln u_s)^2(r-1)^2\right] . \qquad \text{(A.13)}$$

Relabeling signals such that the wage profile is increasing allows us to express the incentive scheme in terms of increases in intrinsic utility. The agent's binding participation constraint implies that

$$u_1 = \bar{u} + c(\hat{a}) - \sum_{s=2}^{S} b_s\left\{\sum_{\tau=s}^{S}\gamma_\tau(\hat{a}) - (\lambda-1)\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right]\left[\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right]\right\} \equiv u_1(\boldsymbol{b}) \qquad \text{(A.14)}$$

and $u_s = u_1(\boldsymbol{b}) + \sum_{t=2}^{s} b_t \equiv u_s(\boldsymbol{b})$ for all $s = 2, \ldots, S$. Inserting the binding participation constraint into the above cost function and replacing $\Omega(\boldsymbol{u}|r)$ equivalently by $\tilde{\Omega}(\boldsymbol{b}|r) \equiv \Omega(u_1(\boldsymbol{b}), \ldots, u_S(\boldsymbol{b})|r)$ yields

$$C(\boldsymbol{b}|r) \approx \bar{u} + c(\hat{a}) + (\lambda - 1)\sum_{s=2}^{S} b_s \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a})\right]\left[\sum_{t=1}^{s-1} \gamma_t(\hat{a})\right] + \tilde{\Omega}(\boldsymbol{b}|r) . \tag{A.15}$$

Hence, for a given increasing wage profile the principal's cost minimization problem is:

PROGRAM ME:

$$\min_{\boldsymbol{b}\in\mathbb{R}_+^{S-1}} \; \boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) + \tilde{\Omega}(\boldsymbol{b}|r)$$

$$\text{subject to} \quad \boldsymbol{b}'\boldsymbol{\beta}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \tag{IC$'$}$$

If r is sufficiently close to 1, then the incentive scheme that solves program ML also solves program ME. Note that generically program ME is solved only by bonus schemes. Put differently, even if there are multiple optimal contracts for program ML, all these contracts are generically simple bonus contracts. Thus, from Proposition 2 it follows that generically for r close to 1 the optimal incentive scheme entails a minimum of wage differentiation. Note that for $\lambda = 1$ the principal's problem is to minimize $\tilde{\Omega}(\boldsymbol{b}|r)$ even for r sufficiently close to 1.

Q.E.D.

Proof of Proposition I.7: First consider $b \geq 0$. We divide the analysis for $b \geq 0$ into three subcases.

Case 1 ($\boldsymbol{a_0 < 0}$): For the effort level $\hat{a}$ to be chosen by the agent, this effort level has to satisfy the following incentive compatibility constraint:

$$\hat{a} \in \arg\max_{a\in[0,1]} \; u + \gamma(a)b - \gamma(a)(1-\gamma(a))b(\lambda - 1) - \frac{k}{2}a^2 \tag{IC}$$

For $\hat{a}$ to be a zero of $dE\left[U(a)\right]/da$, the bonus has to be chosen according to

$$b^*(\hat{a}) = \frac{k\hat{a}}{(\gamma^H - \gamma^L)\left[2 - \lambda + 2\gamma(\hat{a})(\lambda - 1)\right]} .$$

For $a > a_0$, $b^*(a)$ is a strictly increasing and strictly concave function with $b^*(0) = 0$. Hence, each $\hat{a} \in [0, 1]$ can be made a zero of $dE\left[U(a)\right]/da$ with a non-negative bonus. By choosing the bonus according to $b^*(\hat{a})$, $\hat{a}$ satisfies, by construction, the first-order condition. Inserting $b^*(\hat{a})$ into the $d^2E\left[U(a)\right]/da^2$ shows that expected utility is strictly concave function if $a_0 < 0$. Hence, with the bonus set equal to $b^*(\hat{a})$, effort level $\hat{a}$ satisfies the second-order condition for optimality and therefore is incentive compatible.

Case 2 ($a_0 = 0$): Just like in the case where $a_0 < 0$, each effort level $a \in [0, 1]$ turns out to be implementable with a non-negative bonus. To see this, consider bonus

$$b_0 = \frac{k}{2(\gamma^H - \gamma^L)^2(\lambda - 1)}.$$

For $b \leq b_0$, $dE[U(a)]/da < 0$ for each $a > 0$, that is, lowering effort increases expected utility. Hence, the agent wants to choose an effort level as low as possible and therefore exerts no effort at all. If, on the other hand, $b > b_0$, then $dE[U(a)]/da > 0$. Now, increasing effort increases expected utility, and the agent wants to choose effort as high as possible. For $b = b_0$, expected utility is constant over all $a \in [0, 1]$, that is, as long as his participation constraint is satisfied, the agent is indifferent which effort level to choose. As a tie-breaking rule we assume that, if indifferent between several effort levels, the agent chooses the effort level that the principal prefers.

Case 3 ($a_0 > 0$): If $a_0 > 0$, the agent either chooses $a = 0$ or $a = 1$. To see this, again consider bonus b_0. For $b \leq b_0$, $dE[U(a)]/da < 0$ for each $a > 0$. Hence, the agent wants to exert as little effort as possible and chooses $a = 0$. If, on the other hand, $b > b_0$, then $d^2E[U(a)]/da^2 > 0$, that is, expected utility is a strictly convex function of effort. In order to maximize expected utility, the agent will choose either $a = 0$ or $a = 1$ depending on whether $E[U(0)]$ exceeds $E[U(1)]$ or not.

Negative Bonus $b < 0$: Let $b^- < 0$ denote the monetary punishment that the agent receives if the good signal is observed. With a negative bonus, the agent's expected utility is

$$E[U(a)] = u + \gamma(a)b^- + \gamma(a)(1 - \gamma(a))\lambda b^- + (1 - \gamma(a))\gamma(a)(-b^-) - \frac{k}{2}a^2. \quad \text{(A.16)}$$

The first derivative with respect to effort,

$$\frac{dE[U(a)]}{da} = \underbrace{(\gamma^H - \gamma^L)b^- [\lambda - 2\gamma(a)(\lambda - 1)]}_{MB^-(a)} - \underbrace{ka}_{MC(a)},$$

reveals that $MB^-(a)$ is a positively sloped function, which is steeper the harsher the punishment is, that is, the more negative b^- is. It is worthwhile to point out that if bonus and punishment are equal in absolute value, $|b^-| = b$, then also the slopes of $MB^-(a)$ and $MB(a)$ are identical. The intercept of $MB^-(a)$ with the horizontal axis, a_0^- again is completely determined by the model parameters:

$$a_0^- = \frac{\lambda - 2\gamma^L(\lambda - 1)}{2(\gamma^H - \gamma^L)(\lambda - 1)}$$

Note that $a_0^- > 0$ for $\gamma^L \leq 1/2$. For $\gamma^L > 1/2$ we have $a_0^- < 0$ if and only if $\lambda > 2\gamma^L/(2\gamma^L - 1)$. Proceeding in exactly the same way as in the case of a non-negative bonus yields a familiar results: effort level $\hat{a} \in [0, 1]$ is implementable with a strictly negative bonus if and only if $a_0^- \leq 0$. Finally, note that $a_0 < a_0^-$. Hence a negative bonus does not improve the scope for implementation.

Q.E.D.

Proof of Proposition I.8: Throughout the analysis we restricted attention to non-negative bonus payment. It remains to be shown that the principal cannot benefit from offering a negative bonus payment: implementing action $\hat{a}$ with a negative bonus is at least as costly as implementing action $\hat{a}$ with a positive bonus. In what follows, we make use of notation introduced in the paper as well as in the proof of Proposition 7. Let $a_0(p)$, $a_0^-(p)$, $b^*(\hat{a};p)$, and $u^*(\hat{a};p)$ denote the expressions obtained from a_0, a_0^-, $b^*(\hat{a})$, and $u^*(\hat{a})$, respectively, by replacing $\gamma(\hat{a})$, γ^L, and γ^H with $\gamma(\hat{a};p)$, $\gamma^L(p)$, and $\gamma^H(p)$. From the proof of Proposition 6 we know that (i) action $\hat{a}$ is implementable with a non-negative bonus (negative bonus) if and only if $a_0(p) \leq 0$ ($a_0^-(p) \leq 0$), (ii) $a_0^-(p) \leq 0$ implies $a_0(p) < 0$. We will show that, for a given value of p, if $\hat{a}$ is implementable with a negative bonus then it is less costly to implement $\hat{a}$ with a non-negative bonus.

Consider first the case where $a_0^-(p) < 0$. The negative bonus payment satisfying incentive compatibility is given by

$$b^-(\hat{a};p) = \frac{k\hat{a}}{(\gamma^H(p) - \gamma^L(p))\left[\lambda - 2\gamma(\hat{a};p)(\lambda - 1)\right]}.$$

It is easy to verify that the required punishment to implement $\hat{a}$ is larger in absolute value than than the respective non-negative bonus which is needed to implement $\hat{a}$, that is, $b^*(\hat{a};p) < |b^-(\hat{a};p)|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$. When punishing the agent with a negative bonus $b^-(\hat{a};p)$, $u^-(\hat{a};p)$ will be chosen to satisfy the corresponding participation constraint with equality, that is,

$$u^-(\hat{a};p) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(\hat{a};p)b^-(\hat{a};p)\left[\lambda - \gamma(\hat{a},p)(\lambda - 1)\right].$$

Remember that, if $\hat{a}$ is implemented with a non-negative bonus, we have

$$u^*(\hat{a};p) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(\hat{a};p)b^*(\hat{a};p)\left[2 - \lambda + \gamma(\hat{a};p)(\lambda - 1)\right].$$

It follows immediately that the minimum cost of implementing $\hat{a}$ with a non-negative bonus is lower than the minimum implementation cost with a strictly negative bonus:

$$\begin{aligned}
C^-(\hat{a};p) &= u^-(\hat{a};p) + \gamma(\hat{a};p)b^-(\hat{a};p) \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(\hat{a};p)b^-(\hat{a};p)\left[\lambda - \gamma(\hat{a};p)(\lambda-1) - 1\right] \\
&> \bar{u} + \frac{k}{2}\hat{a}^2 + \gamma(\hat{a};p)b^*(\hat{a};p)\left[\lambda - \gamma(\hat{a};p)(\lambda-1) - 1\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(\hat{a};p)b^*(\hat{a};p)\left[1 - \lambda + \gamma(\hat{a};p)(\lambda-1)\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(\hat{a};p)b^*(\hat{a};p)\left[2 - \lambda + \gamma(\hat{a};p)(\lambda-1)\right] + \gamma(\hat{a};p)b^*(\hat{a};p) \\
&= u^*(\hat{a};p) + \gamma(\hat{a};p)b^*(\hat{a};p) \\
&= C(\hat{a};p).
\end{aligned}$$

The same line of argument holds when $a_0^- = 0$: the bonus which satisfies the (IC) is

$$b_0^-(\hat{a};p) = -\frac{k}{2(\gamma^H(p) - \gamma^L(p))^2(\lambda-1)},$$

and so $b^*(\hat{a};p) < |b_0^-(\hat{a};p)|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$.

Q.E.D.

Proof of Corollary I.1: Let $p \in (0,1)$. With $\hat{\boldsymbol{\zeta}}$ being a convex combination of $\hat{\boldsymbol{\gamma}}$ and $\mathbf{1}$ we have $(\zeta^H, \zeta^L) = p(1,1) + (1-p)(\gamma^H, \gamma^L) = (\gamma^H + p(1-\gamma^H), \gamma^L + p(1-\gamma^L))$. The desired result follows immediately from Proposition 3: Consider $\lambda > 2$. Implementation problems are less likely to be encountered under $\hat{\boldsymbol{\zeta}}$ than under $\hat{\boldsymbol{\gamma}}$. Moreover, if implementation problems are not an issue under both performance measures, then implementation of a certain action is less costly under $\hat{\boldsymbol{\zeta}}$ than under $\hat{\boldsymbol{\gamma}}$. For $\lambda = 2$ implementation problems do not arise and implementation costs are identical under both performance measures. Last, if $\lambda < 2$, implementation problems are not an issue under either performance measure, but the cost of implementation is strictly lower under $\hat{\boldsymbol{\gamma}}$ than under $\hat{\boldsymbol{\zeta}}$.

Q.E.D.

1.2. Validity of the First-Order Approach

Lemma A.1: *Suppose (A1)-(A3) hold, then the incentive constraint in the principal's cost minimization problem can be represented as* $E[U'(\hat{a})] = 0$.

Proof: Consider a contract $(u_1, \{b_s\}_{s=2}^S)$ with $b_s \geq 0$ for $s = 2, \ldots, S$. In what follows, we write β_s instead of $\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$ to cut back on notation. The proof proceeds in two steps. First, we show that for a given contract with the property $b_s > 0$ only if $\beta_s > 0$, all actions that satisfy the first-order condition of the agent's utility maximization problem characterize a local maximum of his utility function. Since the utility function is continuous and all extreme points are local maxima, if there exists some action that fulfills the first-order condition, this action corresponds to the unique maximum. In the second step we show that under the optimal contract we cannot have $b_s > 0$ if $\beta_s \leq 0$.

Step 1: The second derivative of the agent's utility with respect to a is

$$E[U''(a)] = -2(\lambda - 1)\sum_{s=2}^{S} b_s\sigma_s - c''(a) \ , \tag{A.17}$$

where $\sigma_s := (\sum_{i=1}^{s-1}\gamma_i^H - \gamma_i^L)(\sum_{i=s}^{S}\gamma_i^H - \gamma_i^L) < 0$. Suppose action $\hat{a}$ satisfies the first-order condition. Formally

$$\sum_{s=2}^{S} b_s\beta_s = c'(\hat{a}) \iff \sum_{s=2}^{S} b_s\frac{\beta_s}{\hat{a}} = \frac{c'(\hat{a})}{\hat{a}} \ . \tag{A.18}$$

Action $\hat{a}$ locally maximizes the agent's utility if

$$-2(\lambda - 1)\sum_{s=2}^{S} b_s\sigma_s < c''(\hat{a}) \ . \tag{A.19}$$

Under Assumption (A3), we have $c''(\hat{a}) > c(\hat{a})/\hat{a}$. Therefore, if

$$\sum_{s=2}^{S} b_s\big[-2(\lambda-1)\sigma_s - \beta_s/\hat{a}\big] < 0 \ , \tag{A.20}$$

then (A.18) implies (A.19), and each action $\hat{a}$ satisfying the first-order condition of the agent's maximization problem is a local maximum of his expected utility. Inequality (A.20) obviously is satisfied if each element of the sum is negative. Summand s is negative if and only if

$$\begin{aligned} &-2(\lambda-1)\left(\sum_{i=1}^{s-1}(\gamma_i^H - \gamma_i^L)\right)\left(\sum_{i=s}^{S}(\gamma_i^H - \gamma_i^L)\right)\hat{a} \\ &\qquad - \left(\sum_{\tau=s}^{S}(\gamma_\tau^H - \gamma_\tau^L)\right)\left[1 - (\lambda-1)\left(\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right)\right] \\ &\qquad\qquad - (\lambda-1)\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right]\left(\sum_{t=1}^{s-1}(\gamma_t^H - \gamma_t^L)\right) < 0 \ . \end{aligned}$$

Rearranging the above inequality yields

$$\begin{aligned} \left(\sum_{i=s}^{S}(\gamma_i^H - \gamma_i^L)\right)\left\{\lambda + 2(\lambda-1)\left[\hat{a}\sum_{i=1}^{s-1}(\gamma_i^H - \gamma_i^L) - \sum_{i=1}^{s-1}\gamma_i(\hat{a})\right]\right\} &> 0 \\ \iff \left(\sum_{i=s}^{S}(\gamma_i^H - \gamma_i^L)\right)\left\{\lambda\left(1 - \sum_{i=1}^{s-1}\gamma_i^L\right) + (2-\lambda)\sum_{i=1}^{s-1}\gamma_i^L\right\} &> 0 \end{aligned} \tag{A.21}$$

The term in curly brackets is positive, since $\lambda \leq 2$ and $\sum_{i=1}^{s-1} \gamma_i^L < 1$. Furthermore, note that $\sum_{i=s}^{S} (\gamma_i^H - \gamma_i^L) > 0$ since $\beta_s > 0$ for all $b_s > 0$. This completes the first step of the proof.

Step 2: Consider a contract with $b_s > 0$ and $\beta_s \leq 0$ for at least one signal $s \in \{2, \ldots, S\}$ that implements $\hat{a} \in (0, 1)$. Then, under this contract, (IC′) is satisfied and there exists at least one signal t with $\beta_t > 0$ and $b_t > 0$. Obviously, the principal can reduce both b_s and b_t without violating (IC′). This reasoning goes through up to the point where (IC′) is satisfied and $b_s = 0$ for all signals s with $\beta_s \leq 0$. From the first step of the proof we know that the resulting contract implements $\hat{a}$ incentive compatibly. Next, we show that reducing any spread, say b_k, always reduces the principal's cost of implementation.

$$C(\boldsymbol{b}) = \sum_{s=1}^{S} \gamma_s(\hat{a}) h \left(u_1(\boldsymbol{b}) + \sum_{t=2}^{s} b_s \right) , \tag{A.22}$$

$$\text{where} \quad u_1(\boldsymbol{b}) = \bar{u} + c(\hat{a}) - \sum_{t=2}^{S} b_s \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) - (\lambda - 1) \left(\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{s-1} \gamma_t(\hat{a}) \right) \right] .$$

The partial derivative of the cost function with respect to an arbitrary b_k is

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} = \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h' \left(u_1(\boldsymbol{b}) + \sum_{t=2}^{s} b_s \right) \left[\frac{\partial u_1}{\partial b_k} \right] + \sum_{s=k}^{S} \gamma_s(\hat{a}) h' \left(u_1(\boldsymbol{b}) + \sum_{t=2}^{s} b_s \right) \left[\frac{\partial u_1}{\partial b_k} + 1 \right] .$$

Rearranging yields

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} = \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h'(u_s) \underbrace{\left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right]}_{<0} + \sum_{s=k}^{S} \gamma_s(\hat{a}) h'(u_s) \underbrace{\left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) + 1 \right]}_{>0} . \tag{A.23}$$

Note $u_s \leq u_{s+1}$ which implies that $h'(u_s) \leq h'(u_{s+1})$. Thus, the following inequality holds

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} \geq \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right] + \sum_{s=k}^{S} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) + 1 \right] . \tag{A.24}$$

The above inequality can be rewritten as follows

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} \geq h'(u_k)\left[(\lambda-1)\left(\sum_{\tau=k}^{S}\gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1}\gamma_t(\hat{a})\right)\right] > 0 \; .$$

Since reducing any bonus lowers the principal's cost of implementation, it cannot be optimal to set $b_s > 0$ for $\beta_s \leq 0$. This completes the second step of the proof. In combination with step 1, this establishes the desired result.

Q.E.D.

2. Appendix to Chapter II

2.1. Proof of Propositions and Lemmas

Proof of Proposition II.1: As mentioned in Footnote 14, in order to establish the proposition, we follow a different but nevertheless equivalent way than proposed in the paper. In period 1, a naive agent believes that he is time-consistent in period 2. Thus, we first analyze what effort a TC chooses in period 2, given an arbitrary effort level of the first period, $\hat{e}_1$. This effort choice, which maximizes $U_2^{TC} = -c(e_2) + g(\hat{e}_1 + e_2)$, obviously is a function of the first-period effort. Thus, $e_2^{TC}(\hat{e}_1)$ is characterized by the corresponding first-order condition,

$$g'(\hat{e}_1 + e_2^{TC}(\hat{e}_1)) = c'(e_2^{TC}(\hat{e}_1)) \; . \tag{A.25}$$

Differentiating (A.25) with respect to e_1 yields $de_2^{TC}(e_1)/de_1 \in (-1, 0)$. With $U_1^N = -c(e_1) - \beta c(e_2^{TC}(e_1)) + \beta g(e_1 + e_2^{TC}(e_1))$ being a strictly concave function of e_1, the effort level that a naive agent invests in the first period, e_1^N, is implicitly characterized by the following first-order condition:

$$\beta g'\left(e_1^N + e_2^{TC}(e_1^N)\right) = c'(e_1^N) \; . \tag{A.26}$$

The actual problem of a naive agent in period 2 is to maximize $U_2^N = -c(e_2) + \beta g(e_1^N + e_2)$ over his second-period effort choice. The optimal second-period effort, e_2^N, satisfies

$$\beta g'\left(e_1^N + e_2^N\right) = c'(e_2^N) \; . \tag{A.27}$$

Comparison of (A.25)-(A.27) allows to establish the proposition. We prove each part of the proposition in turn.

(iii) Comparison of (A.25) and (A.27) immediately yields $e_2^N < e_2^{TC}(e_1^N) = e_2^{TC}$.

(i) Suppose, in contradiction, that $e_1^N \geq e_2^N$. Then $c'(e_1^N) \geq c'(e_2^N)$, which in turn implies $\beta g'(e_1^N + e_2^{TC}(e_1^N)) \geq \beta g'(e_1^N + e_2^N)$. But since $e_2^N < e_2^{TC}(e_1^N)$ and $g''(\cdot) < 0$ we have $\beta g'(e_1^N + e_2^{TC}(e_1^N)) < \beta g'(e_1^N + e_2^N)$, a contradiction.

(ii) From our considerations of the TC we know that $g'(\hat{e}_1 + e_2^{TC}(\hat{e}_1)) = c'(e_2^{TC}(\hat{e}_1))$ for all $\hat{e}_1$. Hence, $c'(e^{TC}) = c'(e_2^{TC}(e^{TC})) = g'(e^{TC} + e_2^{TC}(e^{TC})) > \beta g'(e^{TC} + e_2^{TC}(e^{TC}))$. For e_1^N we must have $c'(e_1^N) = \beta g'(e_1^N + e_2^{TC}(e_1^N))$. Since $de_2^{TC}/de_1 \in (-1, 0)$, $g''(\cdot) < 0$ and $c''(\cdot) > 0$, we immediately obtain that $e_1^N < e^{TC}$. Now it immediately follows that $e_1^N + e_2^N < e_1^N + e_2^{TC}(e_1^N) < e^{TC} + e_2^{TC}(e^{TC}) = 2e^{TC}$, where the first inequality holds by (i) and the second inequality holds because $e_1^N < e^{TC}$ and $de_2^{TC}(e_1)/de_1 \in (-1, 0)$.

This concludes the proof.

Q.E.D.

Proof of Proposition II.2: First we prove that the effort choice in the first period of a sophisticated agent is characterized by a first-order condition. We can rule out corner solutions to be optimal: With $c(e) \longrightarrow \infty$ as $e \longrightarrow \infty$, $e_1 = \infty$ is not a candidate for the agent's first-period effort. Next we show that $e_1 = 0$ also is not optimal. The derivative of U_1^S with respect to e_1 can be rewritten as follows:

$$\frac{dU_1^S}{de_1} = \left[\frac{de_2^S(e_1)}{de_1}(1-\beta) + 1\right] c'(e_2^S(e_1)) - c'(e_1),$$

where we used twice the fact that $\beta g'\left(e_1 + e_2^S(e_1)\right) = c'(e_2^S(e_1))$. Since $e_2^S(0) > 0$ and $de_2^S(e_1)/de_1 \in (-1, 0)$, we have $dU_1^S/de_1|_{e_1=0} > 0$. Note that U_1^S is a differentiable and hence continuous function, which establishes the desired result.

Next, we prove each part of the proposition in turn.

(i) From (II.5) and (II.6) it follows immediately that $\beta g'\left(e_1^S + e_2^S(e_1^S)\right) - c'(e_1^S) > 0$, which in turn implies that $c'(e_2^S(e_1^S)) = \beta g'(e_1^S + e_2^S(e_1^S)) > c'(e_1^S)$. Thus, $e_2^S(e_1^S) > e_1^S$.

(ii) Suppose, in contradiction, that $e_1^S + e_2^S(e_1^S) \geq 2e^{TC}$. We know that $\beta g'(e_1^S + e_2^S(e_1^S)) - c'(e_1^S) > 0 = g'(e^{TC} + e^{TC}) - c'(e^{TC})$. With $g''(\cdot) < 0$ and $c''(\cdot) > 0$, $e_1^S + e_2^S(e_1^S) \geq 2e^{TC}$ immediately implies $e_1^S < e^{TC}$. Furthermore, $\beta g'(e_1^S + e_2^S(e_1^S)) - c'(e_2^S(e_1^S)) = 0 = g'(e^{TC} + e^{TC}) - c'(e^{TC})$, which under the above functional assumptions implies that $c'(e_2^S(e_1^S)) < c'(e^{TC})$. But this means that $e_2^S(e_1^S) < e^{TC}$, which leads to a contradiction to the assumption that $e_1^S + e_2^S(e_1^S) \geq 2e^{TC}$.

This concludes the proof.

Q.E.D.

Proof of Lemma II.1: For a given first-period effort e_1, both the naive agent and the sophisticated agent face the same maximization problem in period 2. This allows us to write $e_2^N = e_2^S(e_1^N)$. For $i,j \in \{S,N\}$ and $i \neq j$, together with $de_2^S(e_1)/de_1 \in (-1,0)$, this observation immediately yields that $e_1^i > e_1^j$ implies $e_2^i = e_2^S(e_1^i) < e_2^S(e_1^j) = e_2^j$ and $e_1^i + e_2^i > e_1^j + e_2^j$. It remains to show that $e_1^i > e_1^j$ implies $U_0^i = -c(e_1^i) - c(e_2^S(e_1^i)) + g(e_1^i + e_2^S(e_1^i)) \geq -c(e_1^j) - c(e_2^S(e_1^j)) + g(e_1^j + e_2^S(e_1^j)) = U_0^j$. Define $H(e_1) \equiv -c(e_1) - c(e_2^S(e_1)) + g(e_1 + e_2^S(e_1))$. In order to establish the desired result, it suffices to show that

$$\frac{dH(e_1)}{de_1} = g'(e_1 + e_2^S(e_1)) - c'(e_1) + \frac{de_2^S(e_1)}{de_1}\left[g'(e_1 + e_2^S(e_1)) - c'(e_2^S(e_1))\right] > 0$$

for all $e_1 \in [0, e_1^i]$. Since, by Propositions 1 and 2, $e_1^i < e_2^i = e_2^S(e_1^i)$ for $i \in \{S,N\}$, and moreover $de_2^S(e_1)/de_1 < 0$, we have $e_1 < e_2^S(e_1)$ for all $e_1 < e_1^i$. This in turn implies $g'(e_1 + e_2^S(e_1)) - c'(e_1) > g'(e_1 + e_2^S(e_1)) - c'(e_2^S(e_1)) > 0$, where the last inequality follows from (II.5). Together with $de_2^S(e_1)/de_1 \in (-1,0)$, the desired result follows.

Q.E.D.

Proof of Lemma II.2: By the revealed preference argument, for the first-period effort choices of a naive and a sophisticated agent, e_1^N and e_1^S, the following two inequalities have to hold:

$$\begin{aligned}&-c(e_1^N) - \beta c(e_2^{TC}(e_1^N)) + \beta g(e_1^N + e_2^{TC}(e_1^N))\\ &\qquad \geq -c(e_1^S) - \beta c(e_2^{TC}(e_1^S)) + \beta g(e_1^S + e_2^{TC}(e_1^S))\end{aligned}$$

and

$$\begin{aligned}&-c(e_1^S) - \beta c(e_2^{S}(e_1^S)) + \beta g(e_1^S + e_2^{S}(e_1^S))\\ &\qquad \geq -c(e_1^N) - \beta c(e_2^{S}(e_1^N)) + \beta g(e_1^N + e_2^{S}(e_1^N))\end{aligned}$$

Taken together these two inequalities imply

$$\begin{aligned}&\left[g(e_1^N + e_2^{TC}(e_1^N)) - c(e_2^{TC}(e_1^N))\right] - \left[g(e_1^N + e_2^{S}(e_1^N)) - c(e_2^{S}(e_1^N))\right]\\ &\qquad \geq \left[g(e_1^S + e_2^{TC}(e_1^S)) - c(e_2^{TC}(e_1^S))\right] - \left[g(e_1^S + e_2^{S}(e_1^S)) - c(e_2^{S}(e_1^S))\right] . \qquad \text{(A.28)}\end{aligned}$$

Define $F(e_1) \equiv \left[g(e_1 + e_2^{TC}(e_1)) - c(e_2^{TC}(e_1))\right] - \left[g(e_1 + e_2^S(e_1)) - c(e_2^S(e_1))\right]$. Since both sides of (A.28) have the same structure, a sufficient condition for $e_1^S \geq e_1^N$ to hold is $dF(e_1)/de_1 < 0$. From (A.25) and (II.5) we know that $g'(e_1 + e_2^{TC}(e_1)) = c'(e_2^{TC}(e_1))$ and $\beta g'(e_1 + e_2^S(e_1)) = c'(e_2^S(e_1))$. Hence,

$$\frac{dF(e_1)}{de_1} = \left[g'(e_1 + e_2^{TC}(e_1)) - g'(e_1 + e_2^S(e_1))\right] - (1-\beta) g'(e_1 + e_2^S(e_1)) \frac{de_2^S(e_1)}{de_1} . \qquad \text{(A.29)}$$

For $\beta = 0$ we have $dF(e_1)/de_1 = \left[g'(e_1 + e_2^{TC}(e_1)) - g'(e_1 + e_2^S(e_1))\right] < 0$ since $de_2^S(e_1)/de_1 = 0$ in this case. For $\beta = 1$ we have $e_2^{TC}(e_1) = e_2^S(e_1)$ for all e_1, and hence $dF(e_1)/de_1 = 0$. Thus, $\frac{d}{d\beta}(dF(e_1)/de_1) > 0$ is a sufficient condition for $dF(e_1)/de_1 < 0$ for all $\beta \in (0,1)$. Tackling this derivative by brute force yields

$$\begin{aligned}\frac{d}{d\beta}\left[\frac{dF(e_1)}{de_1}\right] &= -g''(\cdot)\frac{de_2^S}{d\beta} \\ &\quad - \left[-g'(\cdot)\frac{de_2^S}{de_1} + (1-\beta)g''(\cdot)\frac{de_2^S}{d\beta}\frac{de_2^S}{de_1} + (1-\beta)g'(\cdot)\frac{d(de_2^S/de_1)}{d\beta}\right] \\ &= (1-\beta)\frac{-2g'(\cdot)g''(\cdot)c''(e_2^S) + \frac{de_2^S}{d\beta}\beta g'(\cdot)\left[g''(\cdot)c'''(e_2^S) - g'''(\cdot)c''(e_2^S)\right]}{\left[\beta g''(\cdot) - c''(e_2^S)\right]^2},\end{aligned}$$

where we made use of the fact that

$$\frac{de_2^S}{d\beta} = -\frac{g'(\cdot)}{\beta g''(\cdot) - c''(e_2^S)} .$$

and

$$\frac{d\{de_2^S/de_1\}}{d\beta} = \frac{c''(e_2^S)[g''(\cdot) + \beta g'''(\cdot)\{de_2^S/d\beta\}] - \beta g''(\cdot)c'''(e_2^S)}{[\beta g''(\cdot) - c''(e_2^S)]^2} .$$

Under the imposed functional assumptions, a sufficient condition for $\frac{d}{d\beta}(dF(e_1)/de_1) > 0$ for all $\beta \in (0,1)$ is $c'''(\cdot) \leq 0$ and $g'''(\cdot) \leq 0$. Together with the above observation that $dF(e_1)/de_1 < 0$ for $\beta = 0$ and $dF(e_1)/de_1 = 0$ for $\beta = 1$, this implies that $dF(e_1)/de_1 < 0$ for all $\beta \in [0,1)$. This allows us to conclude that $e_1^N \leq e_1^S$ when $c'''(\cdot) \leq 0$ and $g'''(\cdot) \leq 0$.

Next, we will show that $e_1^N \neq e_1^S$ for $c'''(\cdot) \leq 0$ and $g'''(\cdot) \leq 0$, which completes the proof. Suppose in contradiction that $e_1^N = e_1^S$. The first-order condition of the utility maximization problem of the first-period sophisticate can be written as follows:

$$\begin{aligned}&\beta g'(e_1^S + e_2^{TC}(e_1^S)) - c'(e_1^S) + \beta\left[g'(e_1^S + e_2^S(e_1^S)) - g'(e_1^S + e_2^{TC}(e_1^S))\right] \\ &\qquad + \frac{de_2^S(e_1)}{de_1}\beta\left[g'(e_1^S + e_2^S(e_1^S)) - c'(e_2^S(e_1^S))\right] = 0 .\end{aligned}$$

Setting $e_1^N = e_1^S$ in the above equation yields

$$\left[g'(e_1^N + e_2^S(e_1^N)) - g'(e_1^N + e_2^{TC}(e_1^N))\right] - (1-\beta)g'(e_1^N + e_2^S(e_1^N))\frac{de_2^S(e_1)}{de_1} = 0. \quad \text{(A.30)}$$

Note that the left-hand side of (A.30) is $dF(e_1)/de_1|_{e_1=e_1^N}$. For $c'''(\cdot) \leq 0$ and $g'''(\cdot) \leq 0$, however, we have just shown that $dF(e_1)/de_1 < 0$ for $\beta \in [0,1)$, a contradiction.

Q.E.D.

Proof of Proposition II.3: Follows immediately from Lemmas 1 and 2.

Q.E.D.

Proof of Proposition II.4: The proof consists of three major parts. First, we formally derive the behavior of a sophisticated agent when facing no deadline. Next we show that when facing a deadline, the utility maximization problem of a sophisticated agent in the first period indeed is solved by a first-period effort pair (e_A, e_{B1}) with $e_A > 0$ and $e_{B1} = 0$. Last, we prove each of the results explicitly stated in the proposition.

PART 1: Consider a sophisticated agent who faces no deadline. With the reward functions for the two tasks being strictly increasing and strictly concave, given any first-period efforts $\hat{e}_{A1}$ and $\hat{e}_{B1}$, a sophisticate will allocate second-period effort in a way such that overall effort is allocated as evenly as possible among the two tasks. Thus, there is the following fundamental distinction to draw for second-period behavior: for a given first-period effort and allocation choice, effort smoothing over tasks in the second period is either optimal or not optimal. Effort smoothing over tasks is not optimal for the second period self, if the the total second-period effort needed to achieve this is too costly. Starting out from this observation, we proceed in two steps: First, we show that it is never optimal for a sophisticate in period 1 to choose effort levels e_{A1} and e_{B1} such that effort smoothing over tasks is not optimal in period 2. Second, given that effort smoothing over tasks is optimal in period 2, we show that when facing no deadline, a sophisticated agent increases effort over time.

Step 1: Let $\alpha_1 \in [0, 1]$ denote the share of the overall first-period effort e_1 which is dedicated to task A, i.e., $e_{A1} = \alpha_1 e_1$ and $e_{B1} = (1 - \alpha_1)e_1$. Further, let $e_2^S(e_1, \alpha_1)$ denote the optimal overall effort for the second-period self of a sophisticate given e_1 and α_1. To prove that it is never optimal for a sophisticate to choose an allocation of first-period effort such that effort smoothing over tasks is not optimal in period 2, assume the opposite: Suppose in period 1 the sophisticate chooses e_1 and α_1 such that (w.l.o.g.) $e_A < e_B$. First, note that for given e_1 and α_1, a necessary condition for effort smoothing over tasks not being optimal in period 2 is that overall second-period effort is lower than overall first-period effort, $e_1 > e_2^S(e_1, \alpha_1)$. Moreover, as argued above, the second-period self will allocate all his effort e_2 to task A in order to make the overall effort allocation over tasks as even as possible, i.e.,

$$\max_{e_2} -c(e_2) + \beta g(\alpha_1 e_1 + e_2) + \beta g((1 - \alpha_1)e_1),$$

with the optimal effort choice $e_2^S(e_1, \alpha_1)$ being characterized by

$$\beta g'(\alpha_1 e_1 + e_2^S(e_1, \alpha_1)) = c'(e_2^S(e_1, \alpha_1)) \, . \tag{A.31}$$

Differentiation with respect to e_1 yields

$$\frac{de_2^S}{de_1} = -\alpha_1 \frac{\beta g''(\alpha_1 e_1 + e_2^S(e_1,\alpha_1))}{\beta g''(\alpha_1 e_1 + e_2^S(e_1,\alpha_1)) - c''(e_2^S(e_1,\alpha_1))} \in (-\alpha_1, 0).$$

In the first period, the sophisticate chooses an effort allocation (e_1, α_1) in order to maximize

$$U_1(e_1,\alpha_1) = -c(e_1) - \beta c(e_2^S(e_1,\alpha_1)) + \beta g(\alpha_1 e_1 + e_2^S(e_1,\alpha_1)) + \beta g((1-\alpha_1)e_1).$$

If the optimal first-period effort allocation is an interior solution, i.e., $e_1 \in (0,\infty)$ and $\alpha_1 \in (0,1)$, then it has to satisfy the necessary first-order conditions for optimality, $\partial U_1(e_1,\alpha_1)/\partial e_1 = 0$ and $\partial U_1(e_1,\alpha_1)/\partial \alpha_1 = 0$. Together with (A.31) and $de_2^S(e_1,\alpha_1)/de_1 < 0$,

$$\frac{\partial U_1(e_1,\alpha_1)}{\partial e_1} = 0 \iff -c'(e_1) + \alpha_1 \beta g'(\alpha_1 e_1 + e_2^S(e_1,\alpha_1))$$
$$+(1-\alpha_1)\beta g'((1-\alpha_1)e_1) + \frac{de_2^S(e_1,\alpha_1)}{de_1}\left[\beta g'(\alpha_1 e_1 + e_2^S(e_1,\alpha_1)) - \beta c'(e_2^S(e_1,\alpha_1))\right] = 0$$

implies

$$-c'(e_1) + \alpha_1 \beta g'(\alpha_1 e_1 + e_2^S(e_1,\alpha_1)) + (1-\alpha_1)\beta g'((1-\alpha_1)e_1) > 0\ . \tag{A.32}$$

Combining (A.31) and (A.32) yields

$$c'(e_2^S(e_1,\alpha_1)) - c'(e_1) - (1-\alpha_1)\beta\left[g'(\alpha_1 e_1 + e_2^S(e_1,\alpha_1)) - g'((1-\alpha_1)e_1)\right] > 0\ . \tag{A.33}$$

As noted above, with effort smoothing over tasks not being optimal we have $e_1 > e_2^S(e_1,\alpha_1)$. Moreover, $\alpha_1 e_1 + e_2^S(e_1,\alpha_1) = e_A < e_B = (1-\alpha_1)e_1$ by assumption. But then $c''(\cdot) > 0$ and $g''(\cdot) < 0$ imply that (A.33) cannot be satisfied, which in turn implies that the optimal first-period allocation (e_1,α_1) cannot be interior. Note that $\alpha_1 = 1$ is not possible because $e_A < e_B$ and effort smoothing over tasks is not possible by assumption. Since $e_1 = \infty$ and $e_1 = 0$ can be ruled out as optimal (note that $\partial U_1(e_1,\alpha_1)/\partial e_1|_{e_1=0} > 0$), we are left with $e_1 \in (0,\infty)$ and $\alpha_1 = 0$. For $\alpha_1 = 0$, second-period behavior is characterized by

$$\beta g'(e_2) = c'(e_2). \tag{A.34}$$

Thus, for $\alpha_1 = 0$, second-period effort does not depend on first-period effort, $e_2^S(e_1,0) = e_2^S$. In period 1, e_1 then is chosen to maximize

$$U_1(e_1,0) = -c(e_1) - \beta c(e_2^S) + \beta g(e_2^S) + \beta g(e_1),$$

and thus is characterized by

$$\beta g'(e_1) = c'(e_1)\ . \tag{A.35}$$

Taken together, (A.34) and (A.35) contradict the assumption that effort smoothing over tasks is not optimal in period two, which requires that first-period effort strictly exceeds second-period effort. This concludes Step 1.
Step 2: From Step 1, we know that in period 2 the agent will allocate effort in a way such that overall effort is spread evenly among the two tasks, i.e., $e_A = e_B = (1/2)(e_1 + e_2)$. Thus, given $\hat{e}_1$, e_2 is chosen in order to maximize

$$U_2^{SND} = -c(e_2) + 2\beta g((1/2)(\hat{e}_1 + e_2)).$$

The optimal second-period effort as a function of the first-period effort, $e_2^{SND}(\hat{e}_1)$, satisfies

$$c'(e_2^{SND}(\hat{e}_1)) = \beta g'((1/2)(\hat{e}_1 + e_2^{SND}(\hat{e}_1))). \tag{II.8}$$

Differentiation of (II.8) yields

$$\frac{de_2^{SND}(e_1)}{de_1} = -\frac{\frac{1}{2}\beta g''(\frac{1}{2}(e_1 + e_2^{SND}(e_1)))}{\frac{1}{2}\beta g''(\frac{1}{2}(e_1 + e_2^{SND}(e_1))) - c''(e_2^{SND}(e_1))} \in (-1, 0).$$

In period 1 a sophisticated agent then chooses his effort level in order to maximize the intertemporal utility of his first-period self,

$$U_1^{SND} = -c(e_1) - \beta c(e_2^{SND}(e_1)) + 2\beta g((1/2)(e_1 + e_2^{SND}(e_1))) .$$

According to the same reasoning as in the single-task case, the optimal first-period effort, e_1^{SND}, is characterized by the following first-order condition:

$$\begin{aligned} &\beta g'((1/2)(e_1^{SND} + e_2^{SND}(e_1^{SND}))) - c'(e_1^{SND}) \\ &\quad + \frac{de_2^{SND}(e_1)}{de_1}\beta \left[g'((1/2)(e_1^{SND} + e_2^{SND}(e_1^{SND}))) - c'(e_2^{SND}(e_1^{SND}))\right] = 0 . \end{aligned} \tag{II.9}$$

From (II.8) we know that $\beta g'(\frac{1}{2}(\hat{e}_1 + e_2^{SND}(\hat{e}_1))) - c'(e_2^{SND}(\hat{e}_1)) = 0$ for all $\hat{e}_1$, and in particular for $\hat{e}_1 = e_1^{SND}$. Since $de_2^{SND}(e_1)/de_1 < 0$, in combination with (II.9) this implies that $\beta g'(\frac{1}{2}(e_1^{SND} + e_2^{SND}(e_1^{SND})) - c'(e_1^{SND}) > 0$. Taken together these two observations yield $c'(e_2^{SND}(e_1^{SND})) = \beta g'(\frac{1}{2}(e_1^{SND} + e_2^{SND}(e_1^{SND}))) > c'(e_1^{SND})$. Since $c''(\cdot) > 0$, it follows that when facing no deadline, a sophisticated agent increases effort over time, that is, $e_1^{SND} < e_2^{SND}(e_1^{SND})$.

PART 2: Next, we provide the proof that when facing a deadline, the utility maximization problem of a sophisticated agent in the first period is solved by a first-period effort pair (e_A, e_{B1}) with $e_A > 0$ and $e_{B1} = 0$. To prove this result, we proceed in three steps. First, we show that we cannot have an interior solution $0 < e_A, e_{B1} < \infty$. Second, we rule out solutions in which the agent chooses an infinite amount of effort for at least one task, and also the solution that the agent does not exhibit any effort at all in the first period. Third, we show that an effort pair (e_A, e_{B1}) with $e_{B1} > 0 = e_A$ is not a solution.

Step 1: Suppose, in contradiction, that there is an interior solution. This solution then would be characterized by the following first-order conditions:

$$\frac{\partial U_1^{S^D}}{\partial e_A} = 0 \iff \beta g'(e_A) - c'(e_A + e_{B1}) = 0 \; , \tag{A.36}$$

$$\frac{\partial U_1^{S^D}}{\partial e_{B1}} = 0 \iff \beta g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - c'(e_A + e_{B1}) + \frac{de_{B2}^{S^D}(e_{B1})}{de_{B1}} \beta \left[g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - c'(e_{B2}^{S^D}(e_{B1})) \right] = 0 \; . \tag{A.37}$$

Combining (A.36) and (A.37) yields

$$\beta g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - \beta g'(e_A) = -\frac{de_{B2}^{S^D}(e_{B1})}{de_{B1}} \beta \left[g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - c'(e_{B2}^{S^D}(e_{B1})) \right] > 0,$$

where the inequality follows from (II.10). This last inequality implies that $e_{B1} + e_{B2}^{S^D}(e_{B1}) < e_A$. From (A.36) it follows that e_A decreases as e_{B1} increases. Comparing (10) and (A.36) yields that for $e_{B1} = 0$ we have $e_A = e_{B2}^{S^D}(0)$. Since $d(e_{B1} + e_{B2}^{S^D}(e_{B1}))/de_{B1} > 0$ it follows that $e_{B1} + e_{B2}^{S^D}(e_{B1}) \geq e_A$ for all $e_{B1} \geq 0$, a contradiction. Hence, the utility maximization problem of a sophisticated agent in the first period cannot have an interior solution.

Step 2: Obviously we can rule out effort choices where the agent invests an infinite high effort in one or both tasks since this would lead to an intertemporal utility of minus infinity. To see that it is not optimal to exert no positive effort at all in the first period, let $\alpha_1 \in [0, 1]$ denote the fraction of e_1 which is dedicated to task B, that is, $e_{A1} = (1 - \alpha_1)e_1$ and $e_{B1} = \alpha_1 e_1$. For each α_1, by (II.10) the optimal second-period effort satisfies $\beta g'(\alpha_1 e_1 + e_{B2}^{S^D}(\alpha_1 e_1)) = c'(e_{B2}^{S^D}(\alpha_1 e_1))$. With this notation, the intertemporal utility in the first period is given by $U_1^{S^D} = -c(e_1) - \beta c(e_{B2}^{S^D}(\alpha_1 e_1)) + \beta g((1 - \alpha_1)e_1) + \beta g(\alpha_1 e_1 + e_{B2}^{S^D}(\alpha_1 e_1))$. Differentiating with respect to e_1, taking into account that $\beta g'(\alpha_1 e_1 + e_{B2}^{S^D}(\alpha_1 e_1)) = c'(e_{B2}^{S^D}(\alpha_1 e_1))$, and rearranging yields

$$\frac{dU_1^{S^D}}{de_1} = \beta(1 - \alpha_1) g'((1 - \alpha_1)e_1) - c'(e_1) + \alpha_1 c'(e_{B2}^{S^D}(\alpha_1 e_1)) \left[1 + (1 - \beta) \frac{de_{B2}^{S^D}(e_{B1})}{de_{B1}} \right] .$$

Evaluated at $e_1 = 0$ we have $dU_1^{S^D}/de_1|_{e_1=0} = \beta(1 - \alpha_1)g'(0) + \alpha_1 c'(e_{B2}^{S^D}(0))[1 + (1 - \beta)(de_{B2}^{S^D}(e_{B1})/de_{B1})] > 0$, for all $\alpha_1 \in [0, 1]$.

Step 3: We are left with two possible candidates for the corner solution: (i) $e_A = 0$ and $e_{B1} > 0$, or (ii) $e_A > 0$ and $e_{B1} = 0$. To see that (i) can be ruled out, suppose that $e_A = 0$ and $e_{B1} > 0$. For $e_A = 0$ to be optimal it must hold that

$$\beta g'(0) - c'(e_{B1}) \leq 0, \tag{A.38}$$

otherwise it would be optimal to invest some positive effort in task A. Since e_{B1} is assumed to be strictly positive, the following first-order condition has to hold:

$$\beta g'(e_{B1} + e_{B2}^{SD}(e_{B1})) - c'(e_{B1}) + \frac{de_{B2}^{SD}(e_{B1})}{e_{B1}} \beta \left[g'(e_{B1} + e_{B2}^{SD}(e_{B1})) - c'(e_{B2}^{SD}(e_{B1})) \right] = 0.$$

The last term of the left-hand side of the above equation is negative, which implies that $\beta g'(e_{B1} + e_{B2}^{SD}(e_{B1})) - c'(e_{B1}) > 0$. Taken together with (A.38) this yields $\beta g'(e_{B1} + e_{B2}^{SD}(e_{B1})) > g'(0)$. This in turn implies $e_{B1} + e_{B2}^{SD}(e_{B1}) < 0$, which is not possible. This establishes the desired result.

PART 3: Having shown that an effort pair (e_A, e_{B1}) with $e_A > 0$ and $e_{B1} = 0$ solves the utility maximization problem of a sophisticated agent in the first period, we now prove each part of the proposition. First we show that a sophisticate exhibits a higher first-period effort when facing deadlines. Suppose, in contradiction, that $e_1^{SND} \geq e^{SD}$. From (II.8) and (II.12) we know, respectively, that $c'(e_2^{SND}(e_1^{SND})) = \beta g'(\frac{1}{2}(e_1^{SND} + e_2^{SND}(e_1^{SND})))$ and $c'(e^{SD}) = \beta g'(\frac{1}{2}(e^{SD} + e^{SD}))$. Since $de_2^{SND}(e_1^{SND})/de_1 \in (-1, 0)$, $e_1^{SND} \geq e^{SD}$ implies that $e_2^{SND}(e_1^{SND}) \leq e^{SD}$, which in turn in implies $e_1^{SND} + e_2^{SND}(e_1^{SND}) \geq 2e^{SD}$. From (II.9), however, we know that $\beta g'(\frac{1}{2}(e_1^{SND} + e_2^{SND}(e_1^{SND}))) - c'(e_1^{SND}) > 0$. Together with (II.12) this implies that $\beta g'(\frac{1}{2}(e_1^{SND} + e_2^{SND}(e_1^{SND}))) - \beta g'(\frac{1}{2}(e^{SD} + e^{SD})) > c'(e_1^{SND}) - c'(e^{SD}) \geq 0$, where the last inequality holds by our initial assumption that $e_1^{SND} \geq e^{SD}$. With $g'(\cdot)$ being strictly decreasing, this last expression implies $e_1^{SND} + e_2^{SND} < 2e^{SD}$, a contradiction. Therefore we must have $e_1^{SND} < e^{SD}$. Together with $e_2^{SND}(e^{SD}) = e^{SD}$, which follows from (II.8) in combination with (II.11) or (II.12), and $de_2^{SND}(e_1)/de_1 \in (-1, 0)$, $e_1^{SND} < e^{SD}$ immediately implies $e_1^{SND} + e_2^{SND}(e_1^{SND}) < 2e^{SD}$. It remains to show that a sophisticate indeed is better off under a deadline from a long-run perspective, i.e., $U_0^{SD} > U_0^{SND}$. Let α and γ denote the allocation of some level of total effort e^{Total} over tasks and time, respectively. Since time-consistent agents and sophisticated agents, both under a deadline and under no deadline, divide effort evenly among tasks, fix $\alpha = \frac{1}{2}$. Long-run utility then is given by $U_0(e^{\text{Total}}, \gamma) = -c(\gamma e^{\text{Total}}) - c((1-\gamma)e^{\text{Total}}) + 2g(\frac{1}{2}e^{\text{Total}})$. Fixing $\gamma = \frac{1}{2}$, it is readily verified that $U_0(e^{\text{Total}}, \frac{1}{2})$ is a strictly concave function of e^{Total} which obtains its maximum for $e^{\text{Total}} = 2e^{TC}$. Hence, with $e_1^{SND} + e_2^{SND}(e_1^{SND}) < 2e^{SD} < 2e^{TC^{ND}}$ we have $U_0^{SD} = U_0(2e^{SD}, \frac{1}{2}) > U_0(e_1^{SND} + e_2^{SND}(e_1^{SND}), \frac{1}{2})$. Next, fixing an arbitrary level of total effort $e^{\text{Total}} > 0$, $U_0(e^{\text{Total}}, \gamma)$ is a strictly concave function with its maximum obtained at $\gamma = \frac{1}{2}$. Hence, $U_0^{SND} < U_0(e_1^{SND} + e_2^{SND}(e_1^{SND}), \frac{1}{2})$, which establishes the desired result.

Q.E.D.

Proof of Proposition II.5: First consider a naive agent who faces no deadline. Since he predicts his own future behavior to be time-consistent, a first-period naif makes a plan that he believes to follow through in period 2. In the first period, he chooses e_1 and plans to choose e_2 tomorrow. Moreover, he plans to allocate $\alpha(e_1 + e_2)$ to task A and $(1-\alpha)(e_1 + e_2)$ to task B. It is important to note, that the allocation of first-period effort respectively second-period effort to a specific task is not important from the perspective of period 1. First-period utility of a naif is

$$U_1^{NND}(e_1, e_2, \alpha) = -c(e_1) - \beta c(e_2) + \beta g(\alpha(e_1 + e_2)) + \beta g((1-\alpha)(e_1 + e_2)).$$

Obviously, $U_1^{NND}(e_1, e_2, \alpha)$ is maximized by an interior solution, $(e_1^{NND}, e_2^{TCND}, \alpha^{NND})$, which is characterized by the following first-order conditions:

$$\frac{\partial U_1^{NND}(e_1, e_2, \alpha)}{\partial e_1} = 0 \iff -c'(e_1^{NND}) + \beta g'(\alpha^{NND}(e_1^{NND} + e_2^{TCND}))\alpha^{NND} + \beta g'((1-\alpha^{NND})(e_1^{NND} + e_2^{TCND}))(1-\alpha^{NND}) = 0\,, \quad \text{(A.39)}$$

$$\frac{\partial U_1^{NND}(e_1, e_2, \alpha)}{\partial e_2} = 0 \iff -\beta c'(e_2^{TCND}) + \beta g'(\alpha^{NND}(e_1^{NND} + e_2^{TCND}))\alpha^{NND} + \beta g'((1-\alpha^{NND})(e_1^{NND} + e_2^{TCND}))(1-\alpha^{NND}) = 0\,, \quad \text{(A.40)}$$

$$\frac{\partial U_1^{NND}(e_1, e_2, \alpha)}{\partial \alpha} = 0 \iff \beta g'(\alpha^{NND}(e_1^{NND} + e_2^{TCND})) = \beta g'((1-\alpha^{NND})(e_1^{NND} + e_2^{TCND})). \quad \text{(A.41)}$$

From (A.41) it follows that $\alpha^{NND} = 1/2$. With $\alpha^{NND} = 1/2$, e_1^{NND} and e_2^{TCND} are characterized by

$$\beta g'((1/2)(e_1^{NND} + e_2^{TCND})) = c'(e_1^{NND})\,, \quad \text{(A.42)}$$

$$g'((1/2)(e_1^{NND} + e_2^{TCND})) = c'(e_2^{TCND})\,, \quad \text{(A.43)}$$

which immediately implies that $e_1^{NND} < e_2^{TCND}$.

Next, we show that given e_1^{NND} the naif will indeed achieve effort smoothing over tasks in the second period. Suppose the opposite, i.e., assume (w.l.o.g.) that in period

1 the naif invested that much more effort in task B than in task A such that even if he invested all second-period effort in task A, effort smoothing is not achieved, $e_{A1}^{NND} + e_2 < e_{B1}$. Since in period 2 the agent prefers an effort allocation as even as possible, all second-period effort is invested in task A and therefore is characterized by

$$\beta g'(e_{A1}^{NND} + e_2^{NND}) = c'(e_2^{NND}) \ . \tag{A.44}$$

From (A.42), we know that

$$\beta g'((1/2)(e_{A1}^{NND} + e_{B1}^{NND} + e_2^{TCND})) = c'(e_{A1}^{NND} + e_{B1}^{NND}) \ . \tag{A.45}$$

Equations (A.44) and (A.45) together with the assumption that effort smoothing is not optimal, which implies $e_2^{NND} < e_{A1}^{NND} + e_{B1}^{NND}$, yields $\beta g'(e_{A1}^{NND} + e_2^{NND}) < \beta g'((1/2)(e_{A1}^{NND} + e_{B1}^{NND} + e_2^{TCND}))$. This last inequality implies $2e_2^{NND} > e_{B1}^{NND} - e_{A1}^{NND} + e_2^{TCND}$, which cannot hold since $e_2^{TCND} > e_1^{NND} > e_2^{NND}$ and $e_{B1}^{NND} - e_{A1}^{NND} > e_2^{NND}$ by the initial assumption that effort smoothing over tasks is not achieved. Thus, first-period effort will always be chosen such that effort smoothing over tasks is achieved in period 2.

Taking into account that effort will be split evenly among tasks, the utility of a second-period naif is

$$U_2^{NND} = -c(e_2) + \beta 2g((1/2)(e_1^{NND} + e^2)) \ .$$

The optimal second-period effort, e_2^{NND}, is characterized by the following first-order condition:

$$\beta g'((1/2)(e_1^{NND} + e_2^{NND}) = c'(e_2^{NND}), \tag{A.46}$$

Comparing (A.43) and (A.46) yields $e_2^{NND} < e_2^{TCND}$, which in combination with (A.42) and (A.46) implies $e_1^{NND} < e_2^{NND}$.

Next, consider the case where a naive agent faces a deadline, formally, $e_{A1} = e_A$, $e_{A2} = 0$, and $e_{B2} = e_2$. The utility of a naive agent in the first period is given by

$$U_1^{ND} = -c(e_A + e_{B1}) - \beta c(e_2^{TC}(e_{B1})) + \beta g(e_A) + \beta g(e_{B1} + e_2^{TC}(e_{B1})) \ ,$$

where $e_2^{TC}(e_{B1})$ is characterized by

$$g'(e_{B1} + e_2^{TC}(e_{B1})) = c'(e_2^{TC}(e_{B1})). \tag{A.47}$$

The first-order conditions of utility maximization take the following form:

$$\frac{\partial U_1^{ND}}{\partial e_A} = 0 \iff \beta g'(e_A) - c'(e_A + e_{B1}) = 0 \tag{A.48}$$

$$\frac{\partial U_1^{ND}}{\partial e_{B1}} = 0 \iff \beta g'(e_{B1} + e_2^{TC}(e_{B1})) - c'(e_A + e_{B1}) = 0 \tag{A.49}$$

If the above maximization problem has interior solutions, $e_A > 0$ and $e_{B1} > 0$, then these solutions are characterized by (A.48) and (A.49). When both first-order conditions are satisfied, we have $g'(e_A) = g'(e_{B1} + e_2^{TC}(e_{B1}))$, that is, at an interior solution we must have $e_A = e_{B1} + e_2^{TC}(e_{B1})$. By (A.48), however, it is immediate that e_A is decreasing in e_{B1}. Moreover, comparing (A.47) and (A.48) reveals that $e_2^{TC}(e_{B1}) > e_A$ for $e_{B1} = 0$. Together with $de_2^{TC}(e_{B1})/de_{B1} \in (-1, 0)$, these last two observations imply that $e_A < e_{B1} + e_2^{TC}(e_{B1})$ for all $e_{B1} \geq 0$, a contradiction. Hence, the naive agent's first-period utility maximization problem has a corner solution. Similar reasoning as in the case of the sophisticate allows us to restrict attention to the following two candidates for this corner solution: (i) $e_A^{N^D} > 0$ and $e_{B1}^{N^D} = 0$ or (ii) $e_A^{N^D} = 0$ and $e_{B1}^{N^D} > 0$. For (ii) to be the solution to the naive agent's first-period problem, the following conditions have to hold:

$$\beta g'(0) - c'(e_{B1}^{N^D}) \leq 0 \tag{A.50}$$

$$\beta g'(e_{B1}^{N^D} + e_{B2}^{TC}(e_{B1}^{N^D})) - c'(e_{B1}^{N^D}) = 0 \tag{A.51}$$

Obviously, for (A.50) and (A.51) to hold simultaneously it is required that $e_{B1}^{N^D} + e_2^{TC}(e_{B1}^{N^D}) \leq 0$, which can never be the case. Therefore we are left with $e_A^{N^D} > 0$ and $e_{B1}^{N^D} = 0$, that is, in the first period the agents invests only in task A. This first-period effort is characterized by

$$\beta g'(e_A^{N^D}) = c'(e_A^{N^D}). \tag{A.52}$$

The second-period utility of a naive agent under a regime with a deadline takes the following form:

$$U_2^{N^D} = -c(e_{B2}) + \beta g(e^{N^D}) + \beta g(e_{B2}) \; .$$

The optimal second-period effort then satisfies

$$\beta g'(e_{B2}^{N^D}) = c'(e_{B2}^{N^D}). \tag{A.53}$$

Comparing (A.52) and (A.53) yields $e_A^{N^D} = e_{B2}^{N^D}$, that is, when facing a deadline a naive agent equates effort over tasks and smoothes effort over time. Let the effort level that is chosen by a naive agent under a regime of deadlines per period and per task be denoted by e^{N^D}.

To show that a naive agent chooses a higher effort level in the first period when facing a deadline, suppose, in contradiction, that $e_1^{N^{ND}} \geq e^{N^D}$. Then $\beta g'(e^{N^D}) = c'(e^{N^D}) \leq c'(e_1^{N^{ND}}) = \beta g'(\frac{1}{2}(e_1^{N^{ND}} + e_2^{TC}(e_1^{N^{ND}}))$, where the first equality holds by

(A.52) and the second equality holds by (A.42). But with $g''(\cdot) < 0$, this implies $e^{ND} \geq \frac{1}{2}(e_1^{NND} + e_2^{TC}(e_1^{NND})) > e_1^{NND}$, a contradiction. Hence we must have $e_1^{NND} < e^{ND}$. Let $e_2^{NND}(e_1)$ be characterized by $\beta g'((1/2)(e_1 + e_2^{NND}(e_1))) = c'(e_2^{NND}(e_1))$. Note that $e_2^{NND}(e^{ND}) = e^{ND}$, which in combination with (A.52) and $de_2^{NND}(e_1)/de_1 \in (-1, 0)$, $e_1^{NND} < e^{ND}$ implies $e_1^{NND} + e_2^{NND} < 2e^{ND}$. That is, when facing a deadline, a naive agent exhibits a higher total effort level than under regime without a deadline.

To see that $U_0^{ND} > U_0^{NND}$ the same reasoning applies as in the case of the sophisticate. For a formal argument we refer to the proof of Proposition 4. Intuitively, under deadlines a naive chooses a more desirable total effort level than under no deadlines, which moreover is allocated more efficiently over the two periods.

Q.E.D.

2.2. Partial Naiveté

In this section of the appendix we analyze the behavior of a partially naive agent in the sense of O'Donoghue and Rabin (2001b). A partially naive person is aware that he has future self-control problems, but he underestimates their magnitude. Formally, let $\hat{\beta} \in (\beta, 1)$ be the person's belief about what his taste for immediate gratification will be in the future. Thus, in the single task model a partially naive agent believes in period 1 that his future self will maximize $-c(e_2) + \hat{\beta} g(\hat{e}_1 + e_2)$, whereas he will actually choose e_2 to maximize $-c(e_2) + \beta g(\hat{e}_1 + e_2)$. Note that the extreme cases $\hat{\beta} = 1$ and $\hat{\beta} = \beta$ correspond to the cases analyzed in the main body of the paper, (total) naiveté and (full) sophistication, respectively.

Single Task Model

Here, we investigate the behavior of a partially naive agent in the single task model with two periods for working on that task.

Definition A.1: *A perception-perfect strategy for a partially naive agent is given by* $(e_1^P, e_2^P(\hat{e}_1))$ *such that (i)* $\forall\ \hat{e}_1 \geq 0$, $e_2^P(\hat{e}_1) \in \arg\max_{e_2}\{-c(e_2) + \beta g(\hat{e}_1 + e_2)\}$, *and (ii)* $e_1^P \in \arg\max_{e_1}\{-c(e_1) - c(e_2^B(e_1)) + \beta g(e_1 + e_2^B(e_1))\}$ *where* $e_2^B(\hat{e}_1) \in \arg\max_{e_2}\{-c(e_2) + \hat{\beta} g(\hat{e}_1 + e_2)\}$. *Let* $e_2^P = e_2^P(e_1^P)$.

In the first period the partially naive agent believes that his second-period effort, $e_2^B(\hat{e}_1)$, is characterized by the following first-order condition

$$\hat{\beta} g'(\hat{e}_1 + e_2^B(\hat{e}_1)) = c'(e_2^B(\hat{e}_1)) \ . \tag{A.54}$$

Differentiating (A.54) with respect to e_1 and rearranging yields

$$\frac{de_2^B(e_1)}{de_1} = -\frac{\hat{\beta}g''(e_1 + e_2^B(e_1))}{\hat{\beta}g''(e_1 + e_B^S(e_1)) - c''(e_2^B(e_1))} \in (-1, 0) \ .$$

The utility of a partially naive agent in the the first period, taking his believed second-period reaction into account, is

$$-c(e_1) - \beta c(e_2^B(e_1)) + \beta g(e_1 + e_2^B(e_1)) \ . \tag{A.55}$$

The corresponding first-order condition for optimality is given by

$$-c'(e_1^P) + \beta g'\left(e_1^P + e_2^B(e_1^P)\right) + \frac{de_2^B(e_1^P)}{de_1}\beta\big[g'\left(e_1^P + e_2^B(e_1^P)\right) - c'(e_2^B(e_1^P))\big] = 0. \tag{A.56}$$

The first-order condition is necessary and sufficient for optimality by similar reasonings as in the case of a sophisticated agent. Finally, actual second-period effort, e_2^P, is characterized by

$$\beta g'(e_1^P + e_2^P) = c'(e_2^P) \ . \tag{A.57}$$

A comparison of (A.54) and (A.57) directly reveals that $e_2^B(e_1^P) > e_2^P$. A partially naive agent is overly optimistic when predicting his future self's willingness to work. From equations (A.54) and (A.56) it follows that $-c'(e_1^P)+\beta g'(e_1^P+e_2^B(e_1^P)) > 0$, which implies that $c'(e_1^P) < \beta g'(e_1^P + e_2^B(e_1^P)) < \beta g'(e_1^P + e_2^P) = c'(e_2^P)$. Thus, $e_1^P < e_2^P$. Put verbally, a partially naive agent increases his effort over time. As last result for the single task case we show that a partially naive agent works less in total than a time-consistent agent, i.e., $e_1^P + e_2^P < 2e^{TC}$. Suppose, in contradiction, that $e_1^P + e_2^P \geq 2e^{TC}$. We know that $\beta g'(e_1^P+e_2^B(e_1^P))-c'(e_1^P) > 0 = g'(e^{TC}+e^{TC})-c'(e^{TC})$. Note that $e_1^P+e_2^P \geq 2e^{TC}$ implies that $e_1^P + e_2^B(e_1^P) \geq 2e^{TC}$. Since $g''(\cdot) < 0$ and $c''(\cdot) > 0$ the above inequality implies that $e_1^P < e^{TC}$. Furthermore, $\beta g'(e_1^P+e_2^P)-c'(e_2^P) = 0 = g'(e^{TC}+e^{TC})-c'(e^{TC})$. But this means that $e_2^P < e^{TC}$ which leads to a contradiction to our a priori assumption that $e_1^P+e_2^P \geq 2e^{TC}$. Now we have established the following result, which is the analog to Propositions 1 and 2 for the case of a partially naive agent.

Proposition A.1: *(i) A partially naive agent invests more effort in period 2 than in period 1, i.e.,* $e_1^P < e_2^P$. *(ii) The total effort a partially naive agent invests is lower than the total effort of a time-consistent person, i.e.,* $e_1^P + e_2^P < 2e^{TC}$. *(iii) A partially naive agent overestimates his future effort, i.e.,* $e_2^P < e_2^B(e_1^P)$.

Two Task Model

This subsection analyzes the two task model introduced in Section 5 for the case of a partially naive agent. Let us first consider the no deadline regime, that is, the agent can work in periods 1 and 2 on both tasks, A and B. In principle, the agent chooses in each period $t \in \{1,2\}$ an effort level e_t and an allocation α_t, with $e_{At} = \alpha_t e_t$ and $e_{Bt} = (1-\alpha_t)e_t$. Obviously, ex post it is optimal that the agent invests the same amount in task A as in task B. Ex ante, however, it is not clear that the agent will choose in the first period (e_1, α_1) such that his second-period self considers it optimal to choose (e_2, α_2) such that $e_A = e_B$. Fortunately, by applying the same line of arguments as in the proof of Proposition 4 one can show that it is never optimal for a first-period partially naive agent to choose (e_1, α_1) such that neither his believed second-period behavior nor his actual second-period behavior will not lead to effort smoothing over tasks. Roughly, the intuition is that e_1 has to be very high and α_1 has to be close to zero or one such that the total effort needed in the second period to set $e_A = e_B$ is too costly for the second-period self. We know from the single task case that the first-period self prefers to work less today and more tomorrow. This can only be achieved by a tupel (e_1, α_1) such that effort smoothing over tasks is a best response for both the actual and the believed second-period self. This observation allows us to focus on the agent's effort choice over time. With effort being spread out evenly among the two tasks, the believed second-period effort as a function of first-period effort, $e_2^{BND}(\hat{e}_1)$, is characterized by

$$c'(e_2^{BND}(\hat{e}_1)) = \hat{\beta} g'((1/2)(\hat{e}_1 + e_2^{BND}(\hat{e}_1))). \tag{A.58}$$

The effort level chosen by a partially naive agent in the first period is determined by the following first-order condition,[1]

$$\begin{aligned} &\beta g'\Big((1/2)(e_1^{BND} + e_2^{BND}(e_1^{BND}))\Big) - c'(e_1^{BND}) \\ &\quad + \frac{de_2^{BND}(e_1)}{de_1}\beta\Big[g'\Big((1/2)(e_1^{BND} + e_2^{BND}(e_1^{BND}))\Big) - c'(e_2^{BND}(e_1^{BND}))\Big] = 0\,. \end{aligned} \tag{A.59}$$

The actual second-period effort, e_2^{PND}, is characterized by

$$c'(e_2^{PND}) = \hat{\beta} g'((1/2)(e_1^{PND} + e_2^{PND})). \tag{A.60}$$

Comparing the above equations reveals that $e_1^{PND} < e_2^{PND}$ and $e_A^{PND} = e_B^{PND} = (1/2)(e_1^{PND} + e_2^{PND})$. Put verbally, when not facing a deadline, a partially naive agent equates effort over tasks but does not achieve effort smoothing over time.

Next, the situation where the partially naive agent faces a deadline after the first period for task A is analyzed. Thus, the agent chooses (e_A, e_{B1}) in the first period and e_{B2} in the second period. When facing this interim deadline, a partially naive agent considers it optimal to work exclusively on task A in the first period, i.e., $e_{B1}^{PD} = 0$.

[1]The first-order approach is valid according to the same reasoning as in the single-task case.

This statement can be verified by the same line of arguments as used to show the corresponding result for the sophisticated agent. Hence, the effort levels which are chosen strictly positive, e_A^{PD} and e_{B2}^{PD}, are characterized as follows:

$$c'(e_A^{PD}) = \beta g'(e_A^{PD}) \tag{A.61}$$
$$c'(e_{B2}^{PD}) = \beta g'(e_{B2}^{PD}) \tag{A.62}$$

When facing a deadline, a partially naive agent smoothes effort over time and equates effort over tasks. Let $e^{PD} = e_A^{PD} = e_1^{PD}$ and $e^{PD} = e_{B2}^{PD} = e_B^{PD} = e_2^{PD}$. Now, we can state the analog result to Proposition 4 (respectively 5) for the case of a partially naive agent.

Proposition A.2: *When facing a deadline, a partially naive agent chooses a higher effort level in the first period and a higher total effort level than under a regime without a deadline, i.e.,* $e_1^{PND} < e^{PD}$ *and* $e_1^{PND} + e_2^{PND} < 2e^{PD}$*. Moreover, the partially naive agent is strictly better off from a long-run perspective when facing a deadline, i.e.,* $U_0^{PD} > U_0^{PND}$.

The statements of the proposition that do not follow from the above analysis can be shown by applying the corresponding parts of the proof of Proposition 4.

3. Appendix to Chapter III

3.1. Proofs of Propositions and Lemmas

Proof of Lemma III.1: Suppose $s_1 \leq s_2 \leq 1$, then consumer θ's net utility is: $\frac{1}{2}\theta(s_i - p_i)^2$ for $i = 1, 2$, if he buys from firm i. Hence, firm 1's profit is,

$$\pi_1 = \begin{cases} p_1(s_1 - p_1)\int_0^1 \theta \, d\theta & \text{, if } s_1 - p_1 > s_2 - p_2 \\ \frac{1}{2} \cdot p_1(s_1 - p_1)\int_0^1 \theta \, d\theta & \text{, if } s_1 - p_1 = s_2 - p_2 \text{ and } s_1 = s_2 \\ 0 & \text{, otherwise} \end{cases} . \tag{A.63}$$

Consequently, firm 1 has an incentive to choose $p_1 \geq 0$ as high as possible such that $s_1 - p_1 > s_2 - p_2$ and it can serve the whole market. Clearly firm 1 will not set p_1 higher than the monopoly price, that is $p_1 \leq \frac{1}{2}s_1$. Note that the problem for firm 2 is similar. Thus for the price game equilibrium one obtains:

$$\begin{array}{llll} \text{if } s_1 = s_2 & \Longrightarrow & p_1^* = 0,\ p_2^* = 0 & \text{and} \quad \pi_1^* = 0,\ \pi_2^* = 0 \\ \text{if } s_1 < s_2 & \Longrightarrow & p_1^* = 0,\ p_2^* = s_2 - s_1 & \text{and} \quad \pi_1^* = 0,\ \pi_2^* > 0\ . \end{array}$$

If firm 1 is aware that $s_2 = 1$, then the profit of firm 1 is always zero, independent of the quality level s_1. Hence, all quality-levels $s_1 \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$ are possible in equilibrium of the two-stage game (quality game and tariff game).

Q.E.D.

Proof of Lemma III.2: When both firms choose the same quality level, their products are perfect substitutes. For the proof, I distinguish two cases.

Case 1: Suppose $\pi_i > \pi_j \geq 0$ and that the corresponding tariffs are (T_i, T_j). Firm j can increase its profit when it offers the tariff

$$T_j^* = \begin{cases} T_i - \varepsilon & , \quad \text{if } A_i > 0 \\ (p_i - \varepsilon)q \ , & \text{if } A_i = 0 \end{cases} ,$$

where $\varepsilon > 0$ is sufficiently small. The profit of firm j is then arbitrarily close to $\pi_i > \pi_j$.

Case 2: Suppose that $\pi_i(T_i, T_j) = \pi_j(T_j, T_i) > 0$, where T_i is the tariff of firm i and T_j the tariff of firm j. Again, firm j can increase its profit by slightly undercutting its rival's tariff. That is, firm j uses the tariff T_j^* defined in Case 1. The rise in firm j's profit is then

$$\lim_{\varepsilon \to 0} \left[\pi_j(T_j^*, T_i) - \pi_j(T_j, T_i) \right] = \lim_{\varepsilon \to 0} \left[\pi_j(T_j^*, T_i) - \pi_i(T_i, T_j) \right] \geq \\ \left[1 - \hat{\theta} \right] A_i + p_i \int_{\hat{\theta}}^{1} q_i(p_i, \theta)\, d\theta - \left[\lambda \cdot A_i + p_i \int_{1-\lambda}^{1} q_i(p_i, \theta)\, d\theta \right] > 0 , \quad \text{(A.64)}$$

where λ is the fraction of customers that purchase from firm i under tariffs (T_i, T_j) and $\hat{\theta}$ is the marginal consumer who is indifferent between buying or not when firm j would not be present.[2] Note that when firm j slightly undercuts firm i's tariff, firm j obtains all customers of firm i, λ, and additionally keeps some of its former customers. Consequently, $\lambda < 1 - \hat{\theta}$.

Hence, when both firms produce the same quality level, there exists no equilibrium where at least one firm earns strictly positive profits.

Q.E.D.

Proof of Lemma III.3: Follows immediately from Lemma 2.

Q.E.D.

[2] If the marginal consumer who is indifferent between purchasing or not when only firm i would be present does not exist, then $\hat{\theta}$ is equal to zero.

Proof of Lemma III.4: Obviously, the high-quality firm can always match the offer of the low-quality firm which yields a profit for the high-quality firm greater than that of the low-quality firm. Thus, if there exists an equilibrium where the low-quality firm serves the consumers with types $\theta \in [\tilde{\theta}, 1] \neq \emptyset$, then both firms share the market in the sense that both firms have a positive market share.

Note that for the complete analysis of the tariff game it was not used that $s_2 > s_1$. Put differently, the label "firm 2" referred to the firm serving the market segment with high valuations whereas "firm 1" referred to the firm serving the market segment of moderate or low valuations. Now, suppose that $s_2 < s_1$ and that firm 2 still serves the consumers with high valuations while firm 1 serves the consumers with low valuations. In any pure strategy equilibrium, in which the two firms share the market, prices and fixed fees have to satisfy the necessary conditions for profit maximization outlined in Section 3. Let $(p_2^*, A_2^*, p_1^*, A_1^*)$ be the equilibrium prices and fixed fees. With firm 2 selling to the high-type consumers and both firms being active in the market it holds that

$$s_2 - p_2^* > s_1 - p_1^* \ . \tag{A.65}$$

Since firm 2's market share is $(1 - \tilde{\theta}^*)$, from (III.11) we obtain

$$-A_2^* + (1 - \tilde{\theta}^*)\frac{1}{2}[(s_2 - p_2^*)^2 - (s_1 - p_1^*)^2] - p_2^*(s_2 - p_2^*)\tilde{\theta}^* = 0 \ .$$

Inserting A_2^* defined by (III.6) into the above equality yields

$$(1 - 2\tilde{\theta}^*)\frac{1}{2}[(s_2 - p_2^*)^2 - (s_1 - p_1^*)^2] - A_1^* - p_2^*(s_2 - p_2^*)\tilde{\theta}^* = 0 \ ,$$

which implies that $\tilde{\theta}^* < 1/2$. Thus, in any pure strategy equilibrium in which both firms share the market, the firm that serves the high-type consumers has a market share greater than 1/2. In the assumed equilibrium, the market share of firm 2 is $(1 - \tilde{\theta}^*)$ and the market share of firm 1 is $(\tilde{\theta}^* - \frac{1}{3}\tilde{\theta}^*)$. Note that in equilibrium $\hat{\theta} = \frac{1}{3}\tilde{\theta}$. The profit of firm 1 in equilibrium, using (III.13), amounts to

$$\pi_1^* = \left(\tilde{\theta}^* - \frac{1}{3}\tilde{\theta}^*\right) A_1^* + \frac{1}{2}p_1^*(s_1 - p_1^*)\left[(\tilde{\theta}^*)^2 - \frac{1}{9}(\tilde{\theta}^*)^2\right] \ . \tag{A.66}$$

The fixed fee of firm 1 is given by (4), i.e., $A_1^* = (1/3)\tilde{\theta}^*(1/2)(s_1 - p_1^*)^2$. Thus,

$$\pi_1^* = \frac{1}{9}(\tilde{\theta}^*)^2(s_1 - p_1^*)^2 + \frac{4}{9}(\tilde{\theta}^*)^2 p_1^*(s_1 - p_1^*) \tag{A.67}$$

$$= \frac{1}{9}(\tilde{\theta}^*)^2(s_1 - p_1^*)[s_1 + 3p_1^*] \ . \tag{A.68}$$

The above profit is maximized for $p_1 = (1/3)s_1$. Since $\tilde{\theta}^* < 1/2$, π_1^* is lower than

$$\pi_1^{MAX} = \frac{1}{4} \cdot \frac{1}{9}\left(\frac{2}{3}s_1\right)(2s_1) = \frac{1}{27}s_1^2 \ . \tag{A.69}$$

In the following, I will show that if firm 1 is the high-quality firm, then there exists a tariff $\hat{T}_1$ that yields a profit $\hat{\pi}_1$ for firm 1 that strictly exceeds π_1^{MAX}. Suppose firm 1 offers the "linear-undercutting" tariff with $\hat{A}_1 = 0$ and $\hat{p}_1 = \min\{(1/2)s_1, s_1 - s_2 + p_2^*\}$ to consumers. By construction, all consumers $\theta \in [0,1]$ now purchase from firm 1, since $s_1 - \hat{p}_1 \geq s_2 - p_2^*$ and $\hat{A}_H = 0 < A_L^*$. The profit of firm 1 when offering $\hat{T}_1$ is

$$\hat{\pi}_1 = \frac{1}{2}\hat{p}_1(s_1 - \hat{p}_1) . \tag{A.70}$$

Suppose $\hat{p}_1 = (1/2)s_1$, then $\hat{\pi}_1 = (1/8)s_1^2$. If, on the other hand, $\hat{p}_1 = s_1 - s_2 + p_2^*$, then $\hat{\pi}_1 = (1/2)(s_1 - s_2 + p_2^*)(s_2 - p_2^*)$. Note that $s_1 - s_2 \geq 1/3$ and $s_2 - p_2^* \geq (1/2)s_1$, since $\min\{(1/2)s_1, s_1 - s_2 + p_2^*\} = s_1 - s_2 + p_2^*$. Thus, in this case $\hat{\pi}_1 \geq (1/12)s_1$. While in both cases $\hat{\pi}_1 \geq \min\{(1/8)s_1^2, (1/12)s_1\} > (1/27)s_1^2 = \pi_1^{MAX}$, offering $\hat{T}_1$ is profitable for firm 1 (high-quality firm), which concludes the proof.

Q.E.D.

Proof of Proposition III.1: In any pure strategy Nash equilibrium of the tariff game either the two firms share the market, or only one firm is active in the market. First, I analyze the equilibrium where only one firm is active in the market. I will show that this is an equilibrium, if the firms' degree of product differentiation is not high. Thereafter, it is shown that there exists an equilibrium in which both firms share the market, if the firms' degree of quality differentiation is not low.

STEP 1 (only one firm has a positive market share): If only one firm has a positive market share in equilibrium, then this must be the high-quality firm (firm 2). For any given tariff of firm 1, if firm 2 offers $T_2(q) = q(s_2 - s_1)$ to consumers, firm 2 serves the whole market and makes strictly positive profits. Thus, if only one firm has a positive market share then $\pi_1^* = 0$. With $\pi_1^* = 0$ firm 2 serves the whole market which implies that $A_2^* = 0$. If firm 2 does not serve the whole market, then there is an unsatisfied residual demand which can profitably be served by firm 1. Next, I will show that $\pi_1^* = 0$ implies that $p_1^* = A_1^* = 0$. Suppose, in contradiction, that firm 1 makes a zero profit but does not set $p_1^* = A_1^* = 0$.

Case 1: $A_1^* > 0$ and $p_1^* \geq 0$. In this case, the best response for firm 2 is to set $A_2^* > 0$. To see this, suppose $A_1^* > 0$, $\pi_1^* = 0$ and $A_2^* = 0$. Since firm 2 serves whole market, if firm 2 increases A_2 slightly above zero, firm 2 looses some consumers with types θ close to zero. Note that these consumers do not switch to firm 1 for A_2 sufficiently small but to the outside option. Nevertheless, firm 2 might loose some consumers with high valuations to firm 1 when slightly increasing A_2 without adjusting p_2. Suppose that – if necessary – firm 2 adjusts p_2 such that all consumers with high valuations still purchase from firm 2. Formally, if the price adjustment is necessary firm 2 decreases p_2

such that $(1/2)(s_2 - p_2)^2 - A_2 = (1/2)(s_1 - p_1^*) - A_1^*$. Thus, the necessary decrease in p_2 due to an increase in A_2 either is zero (if the adjustment is not necessary), or given by

$$\frac{dp_2(A_2)}{dA_2} = \frac{-1}{s_2 - p_2(A_2)} \ . \tag{A.71}$$

Firm 2's profit as a function of $A_2 < A_1^*$ is then given by

$$\pi_2(A_2) = [1 - \bar{\theta}(A_2)]A_2 + p_2(A_2)[s_2 - p_2(A_2)](1/2)[1 - \bar{\theta}^2(A_2)] \ , \tag{A.72}$$

where $\bar{\theta}$ is the marginal consumer who is indifferent between purchasing from firm 2 and not purchasing the good. Taking the derivative of $\pi_2(A_2)$ with respect to A_2 and evaluating this derivative at $A_2 = 0$ yields

$$\left.\frac{d\pi_2(A_2)}{dA_2}\right|_{A_2=0} = 1 + \frac{dp_2(0)}{dA_2}\frac{1}{2}[s_2 - 2p_2^*] \ . \tag{A.73}$$

For $A_2 = 0$ it holds that $p_2(A_2) = p_2^*$. Since p_2^* is an equilibrium price, $p_2^* \leq (1/2)s_2 = p_M(s_2)$, where $p_M(s_2)$ is the optimal linear price of a monopolist with quality s_2. This implies that $2p_2(A_2) \leq s_2$, and hence

$$\left.\frac{d\pi_2(A_2)}{dA_2}\right|_{A_2=0} \geq 1 - \frac{s_2 - 2p_2^*}{2(s_2 - p_2^*)} > 0. \tag{A.74}$$

Thus, given $A_1^* > 0$, we must have $A_2^* > 0$, which cannot happen in an equilibrium with only firm 2 being active.

Case 2: $A_1^* = 0$ and $p_1^* > 0$. For $s_2 = 1$ and $s_1 = 2/3$, if firm 2 serves the whole market it is profitable to set $p_2^* > s_2 - s_1$, since in this case, $s_2 - s_1 < (1/2)s_2 = p_M(s_2)$. With $p_2^* > s_2 - s_1$, firm 1 can realize strictly positive profits by setting $A_1 = 0$ and $0 < p_1 < p_2 - (s_2 - s_1)$, a contradiction to $\pi_1^* = 0$. For $s_1 = 1/3$ and $s_2 \in \{2/3, 1\}$ it is optimal for firm 2 to set $p_2^* \leq p_M(s_2) \leq s_2 - s_1$ and $A_2^* > 0$. Note that for $p_2^* \leq s_2 - s_1$ and $p_1 > 0$ firm 2 offers the product with the strictly higher net value $(s - p)$. Thus, by (III.11), and since $\tilde{\theta} = 0$ implies that $A_2 = 0$,

$$\left.\frac{\partial \pi_2}{\partial \tilde{\theta}}\right|_{\tilde{\theta}=0} > 0 \ .$$

Hence, it is optimal for firm 2 not to serve the whole market, which implies $A_2^* > 0$.

Thus, if only one firm has a positive market share in equilibrium, then $p_1^* = 0$ and $A_1^* = 0$.

Next, given $p_1 = A_1 = 0$, firm 2's best response is calculated. It is shown that the unique best response for firm 2 is to offer the tariff $T_2(q) = 0 + (s_2 - s_1)q$, if the degree of quality differentiation is not high, i.e., $(s_2 = 1, s_1 = 2/3)$ and $(s_2 = 2/3, s_1 = 1/3)$. If firm 2's profit maximization problem has an interior solution, then T_2 is characterized by firm 2's FOCs. According to (III.12), the optimal fixed fee is given by

$$A_2^* = \frac{1}{2}(1-\tilde{\theta})\left[\frac{1}{4}s_2^2(1-\tilde{\theta}^2) - s_1^2\right] . \tag{A.75}$$

The fixed fee is strictly positive if the following condition is satisfied:

$$\frac{1}{4}s_2^2(1-\tilde{\theta}^2) - s_1^2 > 0 . \tag{A.76}$$

For $2s_1 \geq s_2$ and $\tilde{\theta} > 0$ condition (A.76) is violated and thus the optimal tariff is a corner solution. That is, T_2 is either a linear tariff ($A_2 = 0$) or a flat tariff ($p_2 = 0$). Note that $\tilde{\theta} = 0$ implies $A_2 = 0$.

I) Flat tariff ($A_2 > 0, p_2 = 0$): By the definition of the marginal consumer $\tilde{\theta}$, for $p_1 = A_1 = 0$, I obtain

$$A_2 = \frac{1}{2}\tilde{\theta}(s_2^2 - s_1^2) . \tag{A.77}$$

Firm 2's profit is given by

$$\pi_2 = (1-\tilde{\theta})\tilde{\theta}\frac{1}{2}(s_2^2 - s_1^2) .$$

The profit maximizing marginal consumer is $\tilde{\theta} = 1/2$ and the profit π^{flat} is then

$$\pi^{flat} = \frac{1}{8}(s_2^2 - s_1^2) . \tag{A.78}$$

II) Linear tariff ($A_2 = 0, p_2 > 0$): Given that the degree of quality differentiation is not high, for the case with linear tariffs it is clear that the optimal marginal price is $p_2^* = s_2 - s_1$. Then each consumer purchases from firm 2. If firm 2 sets a higher price it has no market share, a lower price is not optimal because $p_2^* = s_2 - s_1 \leq \frac{1}{2} = p_M(s_2)$, where p_M is the price of a monopolist with linear tariff. The profit of firm 2 with linear pricing, π^{lin}, is

$$\pi^{lin} = \frac{1}{2}s_1(s_2 - s_1) . \tag{A.79}$$

A comparison of (A.78) and (A.79) reveals that a linear tariff is optimal for firm 2.

To conclude Step 1 it is shown that firm 2 always leaves an unsatisfied residual demand if $s_2 = 1$ and $s_1 = 1/3$. From equation (III.11) it follows that $\frac{\partial \pi_2}{\partial \tilde{\theta}}\big|_{\tilde{\theta}=0} > 0$, given that

$$(s_2 - p_2)^2 - (s_1 - p_1)^2 > 0 \tag{A.80}$$

Note that $p_2 \leq (1/2)s_2 = p_M(s_2)$. Thus, condition (A.80) holds for $s_2 > 2s_1$, since $(s_2 - p_2) \geq (1/2)s_2 > s_1 \geq (s_1 - p_1)$. Hence, for $s_2 > 2s_1$ in any pure strategy equilibrium both firms share the market.

STEP 2 (both firms are active in the market): Suppose both firms share the market in equilibrium. From Lemma 4 it follows that one can focus on the case where the high-quality firm (firm 2) serves the consumers with high θ's. If there exists a pure strategy Nash equilibrium in which both firms share the market, then this equilibrium is characterized by the equations (III.5), (III.6), (III.10), (III.12), (III.14), and (III.15). This system of equations can be simplified to a single equation in $\tilde{\theta}$, such that $\tilde{\theta}^*$ is characterized by $P(\tilde{\theta}^*) = 0$, where

$$\begin{aligned} P(\tilde{\theta}) = \quad & 27s_2^4(9 - 30\tilde{\theta} - 29\tilde{\theta}^2 + 108\tilde{\theta}^3 + 63\tilde{\theta}^4 - 110\tilde{\theta}^5 - 75\tilde{\theta}^6) \\ + \quad & 108s_2^2(3 - 17\tilde{\theta} + 14\tilde{\theta}^2 + 46\tilde{\theta}^3 - 33\tilde{\theta}^4 - 45\tilde{\theta}^5) \\ + \quad & 108(1 - 8\tilde{\theta} + 18\tilde{\theta}^2 - 27\tilde{\theta}^4) \\ - \quad & 576s_1^2(3 - 23\tilde{\theta} + 57\tilde{\theta}^2 - 45\tilde{\theta}^3) \\ - \quad & 192s_1^2s_2^2(9 - 48\tilde{\theta} + 58\tilde{\theta}^2 + 40\tilde{\theta}^3 - 75\tilde{\theta}^4) \\ - \quad & 16s_1^2s_2^4(27 - 81\tilde{\theta} - 18\tilde{\theta}^2 + 190\tilde{\theta}^3 - 25\tilde{\theta}^4 - 125\tilde{\theta}^5)\ . \end{aligned}$$

Unfortunately, $P(\tilde{\theta})$ is a polynomial of sixth order in $\tilde{\theta}$ that cannot be solved analytically. A plot of the polynomial reveals that $P(\tilde{\theta})$ has exactly one root in $[0, 1]$ for $(s_2 = 1, s_1 = 1/3)$ and $(s_2 = 2/3, s_1 = 1/3)$ and no root in $[0, 1]$ for $(s_2 = 1, s_1 = 2/3)$. A plot of the polynomial for $s_2 = 1$ and $s_1 = 1/3$ is given below. If $P(\tilde{\theta}) = 0$ for

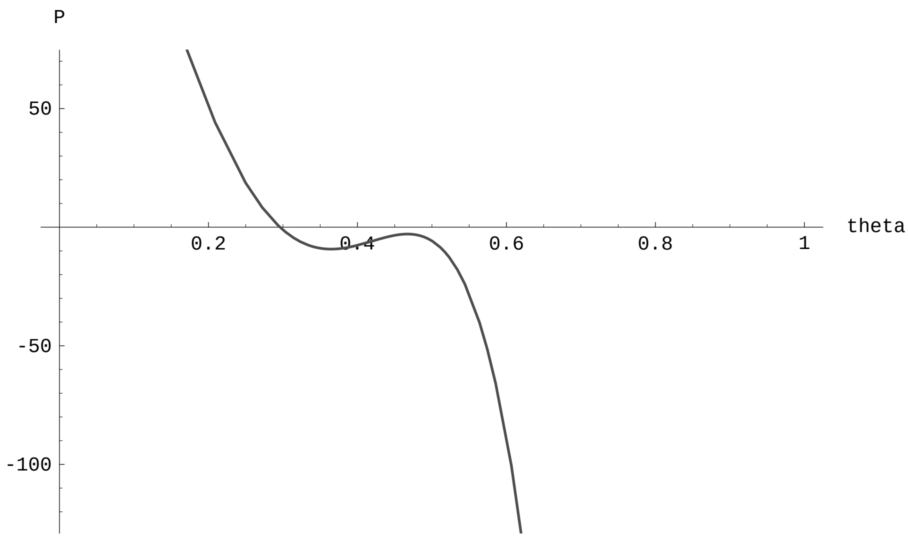

Figure A.1.: Solution of the equilibrium marginal consumer.

$\tilde{\theta} \in [0, 1]$, then the root of $P(\tilde{\theta})$ characterizes a pure strategy Nash equilibrium. By using numerical methods one can solve for the unique $\tilde{\theta}^*$ in the cases $(s_2 = 1, s_1 = 1/3)$ and $(s_2 = 2/3, s_1 = 1/3)$. Having obtained $\tilde{\theta}^*$ one can compute the corresponding equilibrium tariffs, profits, and market shares. For $s_2 = 1$ and $s_1 = 1/3$ the equilibrium values are given in the text, for $s_2 = 2/3$ and $s_1 = 1/3$ the equilibrium values are:

$$\pi_1 = .000885066 \qquad \pi_2 = .0608325$$
$$p_1 = .0753264 \qquad p_2 = .255018$$
$$A_1 = .00260662 \qquad A_2 = .014693$$
$$\hat{\theta} = .078315 \qquad \tilde{\theta} = .234945$$

Q.E.D.

Proof of Proposition III.2: The tariff-game equilibria are characterized in Proposition 1. Thus, it remains to show that $s_i = 1$, $s_j = 1/3$ with $i, j = 1, 2$ and $i \neq j$ is the unique pure strategy equilibrium of the quality game. Clearly, $s_i = 1$ and $s_j = 2/3$ is not an equilibrium, since with these quality levels $\pi_j^* = 0$, but with $s_j = 1/3$ firm j makes strictly positive profits. Similarly, $s_i = 2/3$ and $s_j = 1/3$ is not an equilibrium irrespective of which equilibrium is played at the tariff stage, because it is profitable for firm j to deviate to $s_j = 1$. To see this note that for $s_i = 2/3$ and $s_j = 1/3$ firm j's profit is either zero or .000885066. When deviating to $s_j = 1$, however, firm j serves the whole market and makes profits of: $\pi_j = (s_j - s_i)s_i(1/2) = 1/9 > .000885066$. Obviously, for $s_i = 1$ and $s_j = 1/3$ no firm has an incentive to deviate, hence this is the unique pure strategy equilibrium of the quality game.

Q.E.D.

Proof of Proposition III.3: Follows directly from Proposition 2 in combination with Lemma 1.

Q.E.D.

Proof of Proposition III.4: If price discrimination is permitted, the equilibrium tariffs, profits, and quality-levels are given in the main body of the paper. The social welfare is then:

$$W_D := \pi_1 + \pi_2 + \int_{\tilde{\theta}}^{1} v_2(p_2, \theta)\, d\theta - (1 - \tilde{\theta})A_2 + \int_{\hat{\theta}}^{\tilde{\theta}} v_1(p_1, \theta)\, d\theta - (\tilde{\theta} - \hat{\theta})A_1 \,. \quad \text{(A.81)}$$

Substituting the equilibrium values into (A.81), one obtains

$$W_D^* = 0.201913 \,. \quad \text{(A.82)}$$

Furthermore, standard calculations show that industry profits and consumers' surplus are

$$\Pi_D^* := \pi_1^* + \pi_2^* = 0.144215 \quad \text{(A.83)}$$
$$CS_D^* = 0.0576981 \,. \quad \text{(A.84)}$$

On the other hand, if price discrimination is banned, only one firm (M) enters the market. The monopolist chooses $s_M = 1$. The profit maximization problem for the monopolist is:

$$\pi_M = p_M \int_0^1 \theta(1-p_M)\, d\theta \;\rightarrow\; \max_{p_M} \; . \tag{A.85}$$

The optimal price is $p_M^* = \frac{1}{2}$ and the corresponding equilibrium values for profit, welfare and consumers' surplus are:

$$\pi_M^* = 0.125 \tag{A.86}$$

$$W_M^* = 0.1875 \tag{A.87}$$

$$CS_M^* = 0.0625 \; . \tag{A.88}$$

Q.E.D.

3.2. Examination of the Second-Order Necessary Conditions

First I check the second-order condition (SOC) for firm 2 (high-quality). The partial derivative of π_2 with respect to p_2 is:

$$\frac{\partial \pi_2}{\partial p_2} = \frac{1}{2} s_2 (1-\tilde{\theta})^2 - (1-\tilde{\theta}) p_2 \; . \tag{A.89}$$

From (A.89) it is easy to take the second derivative of π_2 with respect to p_2 and the cross-partial:

$$\frac{\partial^2 \pi_2}{\partial p_2^2} = -(1-\tilde{\theta}) < 0 \tag{A.90}$$

$$\frac{\partial^2 \pi_2}{\partial \tilde{\theta} \partial p_2} = p_2 - s_2(1-\tilde{\theta}) \; . \tag{A.91}$$

Taking the second-order partial derivative of π_2 with respect to $\tilde{\theta}$ yields

$$\frac{\partial^2 \pi_2}{\partial \tilde{\theta}^2} = -\left[(s_2-p_2)^2 - (s_1-p_1)^2\right] - p_2(s_2-p_2) < 0 \; . \tag{A.92}$$

Since the zero points of the FOCs are unique, the FOCs describe a global maximum point if the profit function is concave in the neighborhood of the stationary point. Consequently, it is sufficient to check the sign of the determinant of the Hessian matrix. The determinant of the Hessian is:

$$\det(H) \equiv \frac{\partial^2 \pi_2}{\partial \tilde{\theta}^2} \cdot \frac{\partial^2 \pi_2}{\partial p_2^2} - \left(\frac{\partial^2 \pi_2}{\partial \tilde{\theta} \partial p_2}\right)^2$$

$$\Rightarrow \det(H) = (1-\tilde{\theta})\left[(s_2-p_2)^2 - (s_1-p_1)^2 + p_2(s_2-p_2)\right] - \left(p_2^2 - 2p_2 s_2(1-\tilde{\theta}) + s_2^2(1-\tilde{\theta})^2\right) \; . \tag{A.93}$$

Since a corner solution ($p_2 = 0$ or $p_2 = s_2$) cannot be optimal, the optimal marginal price is characterized by the FOC. Inserting the optimal marginal price $p_2^* = \frac{1}{2}(1-\tilde{\theta})s_2$ into (A.93) yields

$$\det(H) = (1-\tilde{\theta})\Bigg[\underbrace{(s_2-p_2)^2-(s_1-p_1)^2}_{\geq 0,\ \text{for } A_2 \geq 0} + \underbrace{\frac{1}{4}s_2^2(1-\tilde{\theta})(1+\tilde{\theta}) - \frac{1}{4}s_2^2(1-\tilde{\theta})}_{>0}\Bigg] > 0\,. \quad \text{(A.94)}$$

Q.E.D.

Next, I check the SOC for the low-quality firm. To prove that the SOC holds for the low quality firm using the determinant of the Hessian matrix is very tedious. It is easier to show that the profit maximization problem of firm 1 is not solved by a corner solution. Suppose firm 1 sets the marginal price as high as possible, that is, $p_1 = s_1$. Then consumers obtain a non positive surplus if they purchase from firm 1 and consequently $\pi_1 = 0$. Similarly, if firm 1 sets $\tilde{\theta} = 0$ then it has no market share and thus zero profits. On the other hand, if $\tilde{\theta} = 1$ the high-quality firm has no market share and realizes zero profits. This cannot happen in equilibrium: if only one firm is active in the market it is clearly the high-quality firm. To summarize, all corner solutions can be ruled out except that firm 1 sets $p_1 = 0$ and offers a flat tariff.

Suppose $p_1 = 0$ is optimal. The profit maximization problem of firm 1 then is given by

$$\pi_1^{flat}(\tilde{\theta}) := \left(\tilde{\theta} - \hat{\theta}(\tilde{\theta})\right) A_1(\tilde{\theta}) \ \to \max_{\tilde{\theta}}$$

with

$$A_1 = \frac{1}{2}\tilde{\theta}[s_1^2 - (s_2-p_2)^2] + A_2 \ \text{ and } \ \hat{\theta} = \frac{2A_1}{s_1^2}\,.$$

Taking the derivative of π_1^{flat} with respect to $\tilde{\theta}$ yields

$$\frac{d\pi_1^{flat}}{d\tilde{\theta}} = 0 \iff \frac{(s_2-p_2)^2}{s_1^2}\left(\tilde{\theta}\frac{1}{2}\left[(s_1^2-(s_2-p_2)^2\right]+A_2\right) + \left(\frac{\tilde{\theta}(s_2-p_2)^2-2A_2}{s_1^2}\right)\frac{1}{2}\left[s_1^2-(s_2-p_2)^2\right] = 0\,. \quad \text{(A.95)}$$

Solving the above equation for $\tilde{\theta}$ yields

$$\tilde{\theta} = \frac{2(s_2-p_2)^2 - s_1^2}{(s_2-p_2)^2\left[(s_2-p_2)^2 - s_1^2\right]}A_2\,. \quad \text{(A.96)}$$

Evaluating the profit function, π_1^{flat}, at the equilibrium tariff of firm 2 and for $s_1 = \frac{1}{3}$ and $s_2 = 1$ (equivalent calculations for $s_2 = 2/3$, $s_1 = 1/3$) yields

$$\pi_1^{flat} = 0.00129999 < 0.0138222 = \pi_1^*\,.$$

Hence, a flat tariff is not optimal for firm 1. All potential corner solutions have been ruled out as profit maximizing solutions, thus the profit maximization problem of firm 1 has an interior solution.

Q.E.D.

4. Appendix to Chapter IV

4.1. Proofs of Propositions and Lemmas

Proof of Lemma IV.1: For the proof we omit that demand depends on the marginal price p. Suppose, in contradiction, that $\phi_1 < \phi_2$ but $\hat{q}(\phi_1) > \hat{q}(\phi_2)$. By revealed preferences the two inequalities below follow immediately,

$$\begin{aligned} u(\hat{q}(\phi_1);\phi_1) - T(\hat{q}(\phi_1)) &+ \int_{X(\hat{q}(\phi_1))} [T(\hat{q}(\theta)) - T(\hat{q}(\phi_1))] f(\theta)\, d\theta \\ &- \lambda \int_{X^c(\hat{q}(\phi_1))} [T(\hat{q}(\phi_1)) - T(\hat{q}(\theta))] f(\theta)\, d\theta \;\geq\; u(\hat{q}(\phi_2);\phi_1) - T(\hat{q}(\phi_2)) \\ &+ \int_{X(\hat{q}(\phi_2))} [T(\hat{q}(\theta)) - T(\hat{q}(\phi_2))] f(\theta)\, d\theta - \lambda \int_{X^c(\hat{q}(\phi_2))} [T(\hat{q}(\phi_2)) - T(\hat{q}(\theta))] f(\theta)\, d\theta, \end{aligned} \tag{A.97}$$

and

$$\begin{aligned} u(\hat{q}(\phi_2);\phi_2) - T(\hat{q}(\phi_2)) &+ \int_{X(\hat{q}(\phi_2))} [T(\hat{q}(\theta)) - T(\hat{q}(\phi_2))] f(\theta)\, d\theta \\ &- \lambda \int_{X^c(\hat{q}(\phi_2))} [T(\hat{q}(\phi_2)) - T(\hat{q}(\theta))] f(\theta)\, d\theta \;\geq\; u(\hat{q}(\phi_1);\phi_2) - T(\hat{q}(\phi_1)) \\ &+ \int_{X(\hat{q}(\phi_1))} [T(\hat{q}(\theta)) - T(\hat{q}(\phi_1))] f(\theta)\, d\theta - \lambda \int_{X^c(\hat{q}(\phi_1))} [T(\hat{q}(\phi_1)) - T(\hat{q}(\theta))] f(\theta)\, d\theta\,. \end{aligned} \tag{A.98}$$

Subtracting (A.97) from (A.98) and rearranging yields

$$\begin{aligned} [u(\hat{q}(\phi_1);\phi_1) - u(\hat{q}(\phi_2);\phi_1)] - [u(\hat{q}(\phi_1);\phi_2) - u(\hat{q}(\phi_2);\phi_2)] &\geq 0 \\ \Longleftrightarrow \int_{\hat{q}(\phi_2)}^{\hat{q}(\phi_1)} \frac{\partial u(q,\phi_1)}{\partial q}\, dq - \int_{\hat{q}(\phi_2)}^{\hat{q}(\phi_1)} \frac{\partial u(q,\phi_2)}{\partial q}\, dq &\geq 0 \\ \Longleftrightarrow \int_{\hat{q}(\phi_2)}^{\hat{q}(\phi_1)} \int_{\phi_1}^{\phi_2} \frac{\partial^2 u(q,\theta)}{\partial q \partial \theta}\, d\theta dq &\leq 0\,. \end{aligned}$$

The last inequality cannot hold, since $\partial^2 u(q,\theta)/\partial q\partial\theta > 0$ for $q \leq q^S(\theta)$ and by assumption $\phi_1 < \phi_2$ and $\hat{q}(\phi_1) > \hat{q}(\phi_2)$ by hypothesis.

Q.E.D.

Proof of Lemma IV.3: Suppose, in contradiction, there is a personal equilibrium that is at least at one point $\phi \in \Theta$ discontinuous. If the personal equilibrium is discontinuous at ϕ then either $\hat{q}(\phi;p) < \lim_{\varepsilon\to 0}\hat{q}(\phi+|\varepsilon|;p)$ or $\lim_{\varepsilon\to 0}\hat{q}(\phi-|\varepsilon|;p) < \hat{q}(\phi;p)$. While we proof explicitly only the former case, the latter one proceeds by analogous steps. Let $\hat{q}(\phi;p) =: q_1$ and $\lim_{\varepsilon\to 0}\hat{q}(\phi+|\varepsilon|;p) =: q_2$ with $q_1 < q_2$ by discontinuity and monotonicity. First, consider a type $\theta \leq \phi$ who deviates from $\hat{q}(\theta;p) \leq q_1$ to a higher quantity $q \in (q_1, q_2)$. The utility of this type is then given by

$$U(q|\theta,\cdot) = u(q,\theta) - pq - L + p\int_{\phi}^{\bar{\theta}} (\hat{q}(z;p) - q)\, f(z)dz - \lambda p \int_{\underline{\theta}}^{\phi} (q - \hat{q}(z;p))\, f(z)dz\ . \quad \text{(A.99)}$$

For $q \in (q_1, q_2)$ the derivative of type θ's utility with respect to his demand is

$$\frac{dU(q|\theta,\cdot)}{dq} = \frac{\partial u(q,\theta)}{\partial q} - p[2 + (\lambda-1)F(\phi)]\ . \quad \text{(A.100)}$$

Note that $U(q|\phi, \langle\hat{q}(\theta;p)\rangle)$ is continuous for all $q \geq 0$. Thus, it has to hold that $dU/dq|_{q=q_1} \leq 0$ since $\langle\hat{q}(z;p)\rangle$ is a personal equilibrium which implies that type θ has no incentive to demand a quantity $q \in (q_1, q_2)$. Hence, the following inequality has to be satisfied

$$\partial u(q_1,\theta)/\partial q - p[2 + (\lambda-1)F(\phi)] \leq 0\ . \quad \text{(A.101)}$$

Since $\partial u(q_1,\theta)/\partial q$ is increasing in θ, the above inequality is satisfied for all types $\theta \in [\underline{\theta}, \phi]$ if it is satisfied for ϕ. Thus, it has to hold that

$$\partial u(q_1,\phi)/\partial q - p[2 + (\lambda-1)F(\phi)] \leq 0\ . \quad \text{(A.102)}$$

Note that (A.102) gives us a lower bound for q_1.

Now, consider a type $\theta > \phi$ who deviates from $\hat{q}(\theta;p) \geq q_2$ to a lower quantity $q \in (q_1, q_2)$. The marginal change in type θ's utility due to an increase in q amounts to

$$\frac{dU(q|\theta,\cdot)}{dq} = \frac{\partial u(q,\theta)}{\partial q} - p[2 + (\lambda-1)F(\phi)]\ . \quad \text{(A.103)}$$

This downward deviation is not profitable if $dU/dq|_{q=q_2} \geq 0$. Note that $\hat{q}(\theta;p) \geq q_2$ for $\theta > \phi$. Thus, the following inequality needs to be satisfied

$$\partial u(q_2,\theta)/\partial q - p[2 + (\lambda-1)F(\phi)] \geq 0\ . \quad \text{(A.104)}$$

The above inequality is satisfied for all types $\theta \in (\phi, \bar{\theta}]$ if it is satisfied for type ϕ. Thus, it has to hold that

$$\partial u(q_2,\phi)/\partial q - p[2 + (\lambda-1)F(\phi)] \geq 0\ . \quad \text{(A.105)}$$

The inequality (A.105) provides an upper bound for q_2. Combining inequalities (A.102) and (A.105) yields

$$\frac{\partial u(q_2,\phi)}{\partial q} \geq \frac{\partial u(q_1,\phi)}{\partial q} , \tag{A.106}$$

which implies that $q_2 \leq q_1$ a contradiction to $q_1 < q_2$. Thus, the demand profile of any personal equilibrium is continuous in the state of the world.

Q.E.D.

Proof of Lemma IV.4: To proof the result it suffices to show that there cannot be an interval $I \subseteq \Theta$ such that for all $\theta \in I$, $\hat{q}(\theta; p) = \bar{q}$ iff $(C1)$ holds. We show that there is at least one type $\hat{\theta} \in I$ who can profitably deviate to a slightly higher or slightly lower quantity than $\bar{q}$. First, the upward deviation is analyzed.

Consider a type $\hat{\theta} \in I$ who consumes $\bar{q} + \varepsilon$, with $\varepsilon > 0$ but close to zero, his utility is given by

$$U(\bar{q}+\varepsilon|\hat{\theta},\cdot) = u(\bar{q}+\varepsilon,\hat{\theta}) - p(\bar{q}+\varepsilon) - L$$
$$+p\int_{\{\theta\in\Theta|\hat{q}(\theta;p)>\bar{q}+\varepsilon\}}(\hat{q}(\theta;p)-\bar{q}-\varepsilon)f(\theta)\,d\theta - \lambda p\int_{\{\theta\in\Theta|\hat{q}(\theta;p)<\bar{q}+\varepsilon\}}(\bar{q}+\varepsilon-\hat{q}(\theta;p))f(\theta)\,d\theta .$$

Let $\Theta_H \equiv \{\theta \in \Theta|\theta < \inf\{I\}\}$ and $\Theta_L \equiv \{\theta \in \Theta|\theta > \sup\{I\}\}$. Thus, since demand is (weakly) increasing, it follows that for $\varepsilon \to 0$ it holds that $\{\theta \in \Theta|\hat{q}(\theta;p) > \bar{q}+\varepsilon\} = \Theta_H$ and $\{\theta \in \Theta|\hat{q}(\theta;p) < \bar{q}+\varepsilon\} = \Theta_L \cup I$. The increase in utility from consuming slightly more than $\bar{q}$ is

$$\begin{aligned}\left.\frac{dU(\bar{q}+\varepsilon|\hat{\theta},\cdot)}{d\varepsilon}\right|_{\varepsilon=0} &= \frac{du(\bar{q},\hat{\theta})}{dq} - p - p\int_{\theta\in\Theta_H} f(\theta)d\theta - \lambda p\int_{\theta\in\Theta_L\cup I} f(\theta)d\theta \\ &= \partial u(\bar{q},\hat{\theta})/\partial q - p[2+(\lambda-1)F(\theta_H)] ,\end{aligned} \tag{A.107}$$

where $\theta_H := \inf\{\Theta_H\}$.

Next, the case of a downward deviation is considered. Utility of a type $\hat{\theta} \in I$, who consumes $\bar{q} - \varepsilon$ with $\varepsilon > 0$ is

$$U(\bar{q}-\varepsilon|\hat{\theta},\cdot) = u(\bar{q}-\varepsilon,\hat{\theta}) - p(\bar{q}-\varepsilon) - L$$
$$+p\int_{\{\theta\in\Theta|\hat{q}(\theta;p)>\bar{q}-\varepsilon\}}(\hat{q}(\theta;p)-\bar{q}+\varepsilon)f(\theta)\,d\theta - \lambda p\int_{\{\theta\in\Theta|\hat{q}(\theta;p)<\bar{q}-\varepsilon\}}(\bar{q}-\varepsilon-\hat{q}(\theta;p))f(\theta)\,d\theta .$$

The change in utility from an infinitesimal downward deviation is given by

$$\begin{aligned}\left.\frac{dU(\bar{q}-\varepsilon|\hat{\theta},\cdot)}{d\varepsilon}\right|_{\varepsilon=0} &= -\frac{du(\bar{q},\hat{\theta})}{dq} + p + p\int_{\theta\in\Theta_H\cup I} f(\theta)d\theta + \lambda p\int_{\theta\in\Theta_L} f(\theta)d\theta \\ &= -\partial u(\bar{q},\hat{\theta})/\partial q + p[2+(\lambda-1)F(\theta_L)] ,\end{aligned} \tag{A.108}$$

where $\theta_L := \sup\{\Theta_L\}$. A deviation is not profitable if for all $\theta \in I$ it holds that $dU(\bar{q}+\varepsilon|\theta)/d\varepsilon|_{\varepsilon=0} \leq 0$ and $dU(\bar{q}-\varepsilon|\theta)/d\varepsilon|_{\varepsilon=0} \leq 0$.

Thus, a necessary and sufficient condition for the existence of a personal equilibrium where all $\theta \in I$ consume $\bar{q}$ is that (A.107) holds for θ_H and that (A.108) holds for θ_L. Formally, the following to inequalities have to be satisfied:

$$\partial u(\bar{q},\theta_H)/\partial q \leq p[2+(\lambda-1)F(\theta_H)], \tag{A.109}$$

$$\partial u(\bar{q},\theta_L)/\partial q \geq p[2+(\lambda-1)F(\theta_L)] \, . \tag{A.110}$$

Define $\tilde{q}(\theta;p)$ such that $\partial u(\tilde{q}(\theta;p),\theta)/\partial q \equiv p[2+(\lambda-1)F(\theta)]$. Inequalities (A.109) and (A.110) imply that $\tilde{q}(\theta_L;p) \geq \bar{q} \geq \tilde{q}(\theta_H;p)$. By (C1), $d\tilde{q}(\theta;p)/d\theta > 0$. With I being an interval we have $\theta_L < \theta_H$ and thus $\tilde{q}(\theta_L;p) < \tilde{q}(\theta_H;p)$ a contradiction. This completes the proof.

Q.E.D.

Proof of Proposition IV.1: First, note that (IV.11) characterizes the personal equilibrium except for a finite number of kink points. Since the candidate equilibrium is continuously differentiable, we can conclude that there are no kinks in the personal equilibrium if it is strictly increasing in θ.

Remember that local deviations $q \in [\hat{q}(\underline{\theta};p),\hat{q}(\bar{\theta};p)]$ are considered in the main body of the paper. Thus, it remains to show that there is no type who can profitably deviate to a very high or very low quantity, $q < \hat{q}(\underline{\theta};p)$ or $q > \hat{q}(\bar{\theta};p)$. To verify this claim we can focus on the case where $p > 0$.

Suppose the consumer chooses a quantity $q < \hat{q}(\underline{\theta};p)$, then his utility is given by

$$U(q|\theta,\cdot) = u(q,\theta) - pq - L + p\int_{\underline{\theta}}^{\bar{\theta}} [\hat{q}(\hat{\theta};p) - q] f(\hat{\theta}) \, d\hat{\theta} \, .$$

The optimal quantity in this case, q^L is characterized by

$$\frac{\partial u(q^L,\theta)}{\partial q} = 2p \, .$$

Thus, $q^L > \hat{q}(\theta;p)$ for $\theta > \underline{\theta}$ and $q^L = \hat{q}(\theta;p)$ for $\theta = \underline{\theta}$.

Now, consider the case where $q > \hat{q}(\bar{\theta};p)$. Given the state of the world is θ, the consumer's utility is

$$U(q|\theta,\cdot) = u(q,\theta) - pq - L - p\lambda\int_{\underline{\theta}}^{\bar{\theta}} [q - \hat{q}(\hat{\theta};p)] f(\hat{\theta}) \, d\hat{\theta} \, .$$

The optimal quantity in this case, q^H is characterized by

$$\frac{\partial u(q^H,\theta)}{\partial q} = (\lambda+1)p \, .$$

Note that $q^H < \hat{q}(\theta;p)$ for $\theta < \bar{\theta}$ and $q^H = \hat{q}(\theta;p)$ for $\theta = \bar{\theta}$. Hence, no type has an incentive to deviate.

Q.E.D.

Proof of Lemma IV.5: First, note that if the personal equilibrium is strictly increasing in θ in some interval, then in this interval $\hat{q}(\theta;p) \equiv \tilde{q}(\theta;p)$. The proof of Proposition 1 reveals that there is a unique equilibrium candidate if $d\hat{q}/d\theta > 0$.

Suppose there exists an interval $I \subseteq \Theta$ such that $\hat{q}(\theta;p) = \bar{q}$ for all $\theta \in I$. Let $\theta_A := \inf\{I\}$ and $\theta_B := \sup\{I\}$, with $\theta_A < \theta_B$. Furthermore, assume that if $\theta_A > \underline{\theta}$ ($\theta_B < \bar{\theta}$) then there exists an neighborhood $(\theta_A - \varepsilon, \theta_A)$ (respectively $(\theta_B, \theta_B + \varepsilon)$) for $\varepsilon > 0$ sufficiently small, where $\hat{q}(\cdot)$ is strictly increasing. For $\hat{q}(\cdot)$ being constant for all $\theta \in I$ it has to hold that in all states of the world $\theta \in I$ neither a downward deviation nor an upward deviation does improve the consumer's utility.

Downward Deviation ($q < \bar{q}$): Suppose $\theta_A > \underline{\theta}$. The consumer deviates to a lower quantity than $\bar{q}$. Let $\hat{\theta}(q)$ denote the state of the world for which the consumer expected to choose the same quantity. Thus, the consumer feels a gain compared to types $(\hat{\theta}(q), \bar{\theta})$ and a loss compared to types $[\underline{\theta}, \hat{\theta}(q)]$. Formally, $\hat{q}(\hat{\theta}(q);p) = q$ and $\lim_{\varepsilon \to 0} \hat{\theta}(\bar{q} - |\varepsilon|) = \theta_A$. Note, for a minor downward deviation $\hat{\theta}'(q) > 0$. The consumer's utility from a (minor) downward deviation is

$$U^D = u(q,\theta) - pq - L + p\int_{\hat{\theta}(q)}^{\bar{\theta}} (\hat{q}(\theta;p) - q) f(\theta) d\theta \\ - \lambda p \int_{\underline{\theta}}^{\hat{\theta}(q)} (q - \hat{q}(\theta;p)) f(\theta) d\theta \,. \quad \text{(A.111)}$$

Differentiating the above utility with respect to q yields

$$\frac{dU^D}{dq} = \frac{\partial u(q,\theta)}{\partial q} - p\left[2 + (\lambda - 1)F(\hat{\theta}(q))\right] \,. \quad \text{(A.112)}$$

A downward deviation is not utility enhancing if for all $\theta \in I$ the right-hand side of (A.112) is non-negative even at $q = \bar{q}$. Since we imposed the Spence-Mirrlees condition, this holds for all $\theta \in I$ if it holds for θ_A. Thus, it has to hold that

$$\frac{\partial u(\bar{q},\theta_A)}{\partial q} - p\left[2 + (\lambda - 1)F(\theta_A)\right] \geq 0 \,. \quad \text{(A.113)}$$

Note that $\hat{q}(\theta;p)$ is continuous and defined by $\tilde{q}(\theta;p)$ for θ slightly below θ_A. Hence, for $\theta_A > \underline{\theta}$ condition (A.113) has to hold with equality.

Now suppose $\theta_A = \underline{\theta}$. It is straightforward to show that a downward deviation is not utility improving if the following condition holds

$$\frac{\partial u(\bar{q},\underline{\theta})}{\partial q} - 2p \geq 0 \,. \quad \text{(A.114)}$$

With similar reasonings it can be shown that a non-minor downward deviation is not utility enhancing if the above inequality or (A.113) holds.

UPWARD DEVIATION ($q > \bar{q}$): Suppose $\theta_B < \bar{\theta}$. Let $\hat{\theta}(q)$ still denote the cutoff state, i.e., the consumer feels a gain compared to types $(\hat{\theta}(q), \bar{\theta}]$ and a loss compared to types $[\underline{\theta}, \hat{\theta}(q)]$. Now $\hat{\theta}(q) > \theta_B$ and $\lim_{\varepsilon \to 0} \hat{\theta}(\bar{q} + |\varepsilon|) = \theta_B$. The consumer's utility from a (minor) upward deviation is given by

$$U^U = u(q,\theta) - pq - L + p \int_{\hat{\theta}(q)}^{\bar{\theta}} (\hat{q}(\theta;p) - q) f(\theta) d\theta - \lambda p \int_{\underline{\theta}}^{\hat{\theta}(q)} (q - \hat{q}(\theta;p)) f(\theta) d\theta \,. \quad \text{(A.115)}$$

The derivative of U^U with respect to q is

$$\frac{dU^U}{dq} = \frac{\partial u(q,\theta)}{\partial q} - p \left[2 + (\lambda - 1) F(\hat{\theta}(q)) \right] \,. \quad \text{(A.116)}$$

An upward deviation is not utility enhancing for all $\theta \in I$ if

$$\frac{\partial u(\bar{q}, \theta_B)}{\partial q} - p \left[2 + (\lambda - 1) F(\theta_B) \right] \leq 0 \,. \quad \text{(A.117)}$$

Since the personal equilibrium is continuous and $\theta_B < \bar{\theta}$ the above inequality has to hold with equality.

Suppose $\theta_B = \bar{\theta}$. In this case the consumer has no incentive to choose a quantity $q > \bar{q}$ for all states $\theta \in I$ if

$$\frac{\partial u(\bar{q}, \bar{\theta})}{\partial q} - p \left[1 + \lambda \right] \leq 0 \,. \quad \text{(A.118)}$$

Q.E.D.

Proof of Lemma IV.6: First, note that any personal equilibrium is bounded from above by $q^{MAX}(p)$, which is implicitly defined by $\partial u(q^{MAX}, \bar{\theta})/\partial q = 2p$. Let $q^{FB}(\theta)$ denote the first-best quantities, i.e., $\partial u(q^{FB}(\theta), \theta)/\partial q = c$. In the following, we will show that for $p \geq \bar{p}$ the joint surplus, $S(p)$, is bounded from above and that this bound is lower than $S(0)$. To establish the above claim we define $\check{q}(\theta) := \min\{q^{FB}(\theta), q^{MAX}(\bar{p})\}$. It is important to note that there is a positive mass of types for which $\check{q}(\theta) = q^{MAX}(\bar{p})$ if $\bar{p} > (1/2)c$. The joint surplus generated with a unit price $p \geq \bar{p}$ is strictly lower than

$$\check{S} = \int_{\underline{\theta}}^{\bar{\theta}} \left[u(\check{q}(\theta), \theta) - c\check{q}(\theta) \right] f(\theta) d\theta \,, \quad \text{(A.119)}$$

since with a positive unit price the consumer expects to incur some net losses. A sufficient condition for $S(p)$ being maximized by a unit price $p \in [0, \bar{p})$ is that $S(0) \geq \check{S}$

(This condition is by no means necessary). $S(0) \geq \check{S}$ is equivalent to

$$\int_{\underline{\theta}}^{\bar{\theta}} \Big\{ u(q^S(\theta), \theta) - u(\check{q}(\theta), \theta) - c[q^S(\theta) - \check{q}(\theta)] \Big\} f(\theta) d\theta \geq 0 \ . \tag{A.120}$$

The above condition is satisfied for c being sufficiently small, which completes the proof.

Q.E.D.

Proof of Proposition IV.2: First, it is shown how to derive equation (IV.16). Taking the derivative of (IV.15) with respect to p yields

$$\begin{aligned} S'(p) = \int_{\underline{\theta}}^{\bar{\theta}} \Bigg\{ & \left(\frac{\partial u(\hat{q}(\theta,p),\theta)}{\partial q} - c \right) \frac{d\hat{q}(\theta,p)}{dp} \\ & + \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi,p) - \hat{q}(\theta,p)] f(\phi)\, d\phi + p \int_{\theta}^{\bar{\theta}} \left[\frac{d\hat{q}(\phi,p)}{dp} - \frac{d\hat{q}(\theta,p)}{dp} \right] f(\phi)\, d\phi \\ - \lambda & \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,p) - \hat{q}(\phi,p)] f(\phi)\, d\phi - \lambda p \int_{\underline{\theta}}^{\theta} \left[\frac{d\hat{q}(\theta,p)}{dp} - \frac{d\hat{q}(\phi,p)}{dp} \right] f(\phi)\, d\phi \Bigg\} f(\theta)\, d\theta \ . \end{aligned} \tag{A.121}$$

The above equation can be rearranged to

$$\begin{aligned} S'(p) = \int_{\underline{\theta}}^{\bar{\theta}} \Bigg\{ & \left(\frac{\partial u(\hat{q}(\theta,p),\theta)}{\partial q} - p[1 + (\lambda - 1)F(\theta)] - c \right) \frac{d\hat{q}(\theta,p)}{dp} \\ & + p \int_{\theta}^{\bar{\theta}} \frac{d\hat{q}(\phi,p)}{dp} f(\phi)\, d\phi + \lambda p \int_{\underline{\theta}}^{\theta} \frac{d\hat{q}(\phi,p)}{dp} f(\phi)\, d\phi \\ + & \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi,p) - \hat{q}(\theta,p)] f(\phi)\, d\phi - \lambda \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,p) - \hat{q}(\phi,p)] f(\phi)\, d\phi \Bigg\} f(\theta)\, d\theta \ . \end{aligned} \tag{A.122}$$

Note that the following equality holds

$$\begin{aligned} & \int_{\underline{\theta}}^{\bar{\theta}} \left\{ \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi,p) - \hat{q}(\theta,p)] f(\phi)\, d\phi - \lambda \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta,p) - \hat{q}(\phi,p)] f(\phi)\, d\phi \right\} f(\theta)\, d\theta \\ & \qquad = -(\lambda - 1) \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi,p) - \hat{q}(\theta,p)] f(\phi) f(\theta)\, d\phi d\theta. \end{aligned} \tag{A.123}$$

Inserting (A.123) and (IV.11) into (A.122) yields the equation (IV.16) stated in the text part. By using the definition of $\Psi(\cdot)$ the above derivative can be further simplified to

$$S'(p) = \Psi(p) + p \int_{\underline{\theta}}^{\bar{\theta}} \left\{ \int_{\theta}^{\bar{\theta}} \frac{d\hat{q}(\phi,p)}{dp} f(\phi) d\phi + \lambda \int_{\underline{\theta}}^{\theta} \frac{d\hat{q}(\phi,p)}{dp} f(\phi) d\phi \right\} f(\theta) d\theta \ . \tag{A.124}$$

First observe that $S'(p) < 0$ for $p \geq c$. Since $\Psi(p)$ is non-increasing for $p \in [0, c)$, it holds that $S'(0) > S'(p)$ for $p \in (0, c)$. Hence, if $S'(p) \leq 0$ the joint surplus is maximized at $p = 0$. If, on the other hand, $S'(0) > 0$ then there exists a $\hat{p} \in (0, c)$ at which $S(p)$ is maximized. The price $\hat{p}$ is characterized by the first-order condition $S'(\hat{p}) = 0$, since $S(\cdot)$ is continuously differentiable. Note, however, that the first-order condition may not be sufficient.

Next, we show that $S'(0) \leq 0$ is equivalent to $\Sigma(\lambda) \geq c$. By evaluating (IV.16) at $p = 0$, it is obvious that $S'(0) \leq 0$ iff

$$-c \int_{\underline{\theta}}^{\bar{\theta}} \frac{d\hat{q}(\theta, 0)}{dp} f(\theta) d\theta - (\lambda - 1) \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi, 0) - \hat{q}(\theta, 0)] f(\phi) f(\theta)\, d\phi d\theta \leq 0\,. \quad \text{(A.125)}$$

Rearranging the above inequality and using the definition of $\Sigma(\lambda)$ reveals that $S'(0) \leq 0$ if and only if $\Sigma(\lambda) \geq c$. Finally, we verify the following claim.

Claim: $\Sigma'(\lambda) > 0$.

Proof: To cut down on notation we often write $\hat{q}(\theta)$ instead of $\hat{q}(\theta; p)$. Define $Z(\lambda)$ and $N(\lambda)$ as the numerator and the denominator, respectively, of the fraction of $\Sigma(\cdot)$. Thus,

$$Z(\lambda) \equiv \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi, 0) - \hat{q}(\theta, 0)] f(\phi) f(\theta) d\phi d\theta, \quad \text{(A.126)}$$

$$\text{and} \quad N(\lambda) \equiv -\int_{\underline{\theta}}^{\bar{\theta}} [d\hat{q}(\theta, 0)/dp] f(\theta) d\theta\,. \quad \text{(A.127)}$$

With this notation the derivative of $\Sigma(\cdot)$ with respect to λ can be written as

$$\Sigma'(\lambda) = \frac{Z(\lambda)}{N(\lambda)} + (\lambda - 1) \frac{Z'(\lambda) N(\lambda) - N'(\lambda) Z(\lambda)}{N^2(\lambda)} \quad \text{(A.128)}$$

In order to show that $\Sigma'(\lambda) > 0$, we analyze the itemized parts separately. First, we take the derivative of $\hat{q}(\cdot)$ with respect to λ which leads to

$$\frac{d\hat{q}(\cdot)}{d\lambda} = \frac{pF(\theta)}{\partial^2 u(\hat{q}(\theta), \theta)/\partial q^2} \leq 0\,. \quad \text{(A.129)}$$

Thus,

$$\frac{d}{d\lambda}[\hat{q}(\phi) - \hat{q}(\theta)] = \frac{pF(\phi)}{\partial^2 u(\hat{q}(\phi), \phi)/\partial q^2} - \frac{pF(\theta)}{\partial^2 u(\hat{q}(\theta), \theta)/\partial q^2}\,, \quad \text{(A.130)}$$

which equals zero at $p = 0$. Hence, $Z'(\lambda) = 0$. Taking the derivative of (IV.17) with respect to λ yields

$$\frac{d}{d\lambda}\left[\frac{d\hat{q}(\cdot)}{dp}\right] = F(\theta) \left(\frac{\partial^2 u(\hat{q}(\theta), \theta)}{\partial q^2}\right)^{-1} - [2 + (\lambda - 1)F(\theta)] \left(\frac{\partial^2 u(\hat{q}(\theta), \theta)}{\partial q^2}\right)^{-2} \frac{\partial^3 u(\hat{q}(\theta), \theta)}{\partial q^3} \frac{d\hat{q}(\theta)}{d\lambda}\,. \quad \text{(A.131)}$$

Evaluating the above derivative at $p = 0$, and thus $d\hat{q}/d\lambda|_{p=0} = 0$, leads to

$$\frac{d}{d\lambda}\left[\frac{d\hat{q}(\cdot)}{dp}\right]\bigg|_{p=0} = F(\theta)\left(\frac{\partial^2 u(\hat{q}(\theta),\theta)}{\partial q^2}\right)^{-1} < 0 \, .$$

Thus,

$$N'(\lambda) = -\int_{\underline{\theta}}^{\bar{\theta}} F(\theta)\left(\frac{\partial^2 u(\hat{q}(\theta),\theta)}{\partial q^2}\right)^{-1} f(\theta)d\theta \, .$$

Since $Z'(\lambda) = 0$, equation (A.128) simplifies to

$$\begin{aligned}\Sigma'(\lambda) &= \frac{Z(\lambda)}{N(\lambda)} - (\lambda-1)\frac{Z(\lambda)N'(\lambda)}{N^2(\lambda)} \\ &= \frac{Z(\lambda)}{N^2(\lambda)}\Big[N(\lambda) - (\lambda-1)N'(\lambda)\Big] \, . \end{aligned} \tag{A.132}$$

Since $Z(\lambda) > 0$, it remains to show that $N(\lambda) - (\lambda - 1)N'(\lambda) > 0$, which is equivalent to

$$-\int_{\underline{\theta}}^{\bar{\theta}} [d\hat{q}(\theta,0)/dp] f(\theta)d\theta + (\lambda-1)\int_{\underline{\theta}}^{\bar{\theta}} F(\theta)\left(\frac{\partial^2 u(\hat{q}(\theta),\theta)}{\partial q^2}\right)^{-1} f(\theta)d\theta > 0 \, . \tag{A.133}$$

Inserting the explicit formula for $d\hat{q}(\cdot)/dp$ into the above inequality yields

$$\int_{\underline{\theta}}^{\bar{\theta}} \Bigg\{ -[2 + (\lambda-1)F(\theta)]\left(\frac{\partial^2 u(\hat{q}(\theta),\theta)}{\partial q^2}\right)^{-1} + (\lambda-1)F(\theta)\left(\frac{\partial^2 u(\hat{q}(\theta),\theta)}{\partial q^2}\right)^{-1}\Bigg\} f(\theta)d\theta > 0 \tag{A.134}$$

$$\Longleftrightarrow \int_{\underline{\theta}}^{\bar{\theta}} -2\left(\frac{\partial^2 u(\hat{q}(\theta),\theta)}{\partial q^2}\right)^{-1} f(\theta)d\theta > 0 \, . \tag{A.135}$$

The above inequality is satisfied since $u(\cdot)$ is a strictly concave function in q for $q \leq q^S(\theta)$. Q.E.D.

Q.E.D.

Proof of Proposition IV.3: In order to apply Proposition 1 of Armstrong and Vickers (2001), the following three properties have to be satisfied: (i) $[\partial m(v^A, v^B)/\partial v^A]$ $[m(v^A, v^B)]^{-1}$ is non-decreasing in v^B, (ii) there exists $\tilde{v}_j > -\infty$ that maximizes $m(v, v)\pi_j(v)$ for $j = 1, 2$, and (iii) for $j = 1, 2$ there exists $\bar{v}_j$ defined by $\pi_j(\bar{v}_j) = 0$, $\pi_j(v) < 0$ if $v > \bar{v}_j$. Since we explicitly assumed (i) and (ii) these properties are satisfied. To see that (iii) is also satisfied note that $\bar{v}_j = \max_p\{S_j(p)\}$. Obviously, $\pi_j(\bar{v}_j) = 0$ and $\pi_j(v) < 0$ if $v > \bar{v}_j$. Hence, we can apply Proposition 1 of Armstrong and Vickers. According to this proposition there are no asymmetric equilibria and the equilibrium utility level $\hat{v}_j \in (\tilde{v}_j, \bar{v}_j)$. Since $m(v^A, v^B)\pi_j(v^A)$ is continuously differentiable, the equilibrium utility level satisfies the first-order condition of profit maximization. Thus, $\pi_j(\hat{v}_j) = \Phi(\hat{v}_j)$.

From Proposition 2 it follows that the optimal marginal price $\hat{p}_j$ is greater than zero if and only if $\Sigma(\lambda_j) < c$. If this is the case then $\hat{p}_j$ is characterized by $S'(\hat{p}_j) = 0$, as was shown in the proof of Proposition 2. The per customer profit of a firm is given by

$$\pi_j = L + (p - c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_j(\theta; p) f(\theta) d\theta \, . \tag{A.136}$$

Since, in equilibrium, $\pi_j = \Phi(\hat{v}_j)$ the equilibrium fixed fee is given by

$$L_j = \Phi(\hat{v}_j) - (p_j - c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_j(\theta; p_j) f(\theta) d\theta \, . \tag{A.137}$$

Replacing p_j by $\hat{p}$ and 0, leads to the fixed fees $\hat{L}$ and L^F, respectively.

Q.E.D.

Proof of Lemma IV.7: With slight abuse of notation, we omit for the proof that the demand function, $\hat{q}(\cdot)$, depends on the marginal price p. Define $V(\lambda; \theta)$ as the consumer's surplus for a given state of the world on the equilibrium path. Formally,

$$\begin{aligned} V(\lambda; \theta) &= u(\hat{q}(\theta), \theta) - p\hat{q}(\theta) - L + p \int_{\theta}^{\bar{\theta}} [\hat{q}(\phi) - \hat{q}(\theta)] f(\phi) d\phi \\ &\quad -\lambda p \int_{\underline{\theta}}^{\theta} [\hat{q}(\theta) - \hat{q}(\phi)] f(\phi) d\phi \\ &= u(\hat{q}(\theta), \theta) - L - p\hat{q}(\theta)[2 + (\lambda - 1)F(\theta)] \\ &\quad + p \int_{\theta}^{\bar{\theta}} \hat{q}(\phi) f(\phi) d\phi - \lambda p \int_{\underline{\theta}}^{\theta} \hat{q}(\phi) f(\phi) d\phi \, . \end{aligned} \tag{A.138}$$

Taking the derivative of $V(\cdot; \theta)$ with respect to λ yields

$$\begin{aligned} V'(\lambda; \theta) = \frac{d\hat{q}(\theta)}{d\lambda} & \underbrace{\left[\frac{\partial u(\hat{q}(\theta), \theta)}{\partial q} - p[2 + (\lambda - 1)F(\theta)] \right]}_{=0} - pF(\theta)\hat{q}(\theta) \\ & + p \int_{\theta}^{\bar{\theta}} \frac{d\hat{q}(\phi)}{d\lambda} f(\phi) d\phi + p \int_{\underline{\theta}}^{\theta} \hat{q}(\phi) f(\phi) d\phi - \lambda p \int_{\underline{\theta}}^{\theta} \frac{d\hat{q}(\phi)}{d\lambda} f(\phi) d\phi. \end{aligned} \tag{A.139}$$

Note that

$$\frac{d\hat{q}(\lambda; \theta)}{d\lambda} = \frac{pF(\theta)}{\partial^2 u(\hat{q}(\theta), \theta)/\partial q^2} \leq 0 \, . \tag{A.140}$$

The consumer's expected utility is given by $\mathbb{E}_\theta[V(\lambda; \theta)] = \int_{\underline{\theta}}^{\bar{\theta}} V(\lambda; \theta) f(\theta) d\theta$. Hence, the change in expected utility due to an increase in the consumer's degree of loss aversion is given by

$$\frac{d}{d\lambda}\mathbb{E}_\theta[V(\lambda;\theta)] = \int_{\underline{\theta}}^{\bar{\theta}} V'(\lambda;\theta)f(\theta)d\theta.$$

Hence,

$$\begin{aligned}\mathbb{E}_\theta[V'(\lambda)] = p\Big\{ \int_{\underline{\theta}}^{\bar{\theta}}\int_{\theta}^{\bar{\theta}} \frac{d\hat{q}(\phi)}{d\lambda} f(\phi)f(\theta)d\phi d\theta + \lambda \int_{\underline{\theta}}^{\bar{\theta}}\int_{\underline{\theta}}^{\theta} \frac{d\hat{q}(\phi)}{d\lambda} f(\phi)f(\theta)d\phi d\theta \\ + \int_{\underline{\theta}}^{\bar{\theta}}\int_{\underline{\theta}}^{\theta} \hat{q}(\phi)f(\phi)f(\theta)d\phi d\theta - \int_{\underline{\theta}}^{\bar{\theta}} F(\theta)\hat{q}(\theta)f(\theta)d\theta \Big\} \,. \quad \text{(A.141)}\end{aligned}$$

By using integration by parts for all but the last term, the above expression can be simplified to

$$\mathbb{E}_\theta[V'(\lambda)] = p\int_{\underline{\theta}}^{\bar{\theta}} \left\{ \frac{d\hat{q}(\theta)}{d\lambda}[\lambda - (\lambda-1)F(\theta)] - \hat{q}(\theta)[2F(\theta)-1] \right\} f(\theta)d\theta \,. \quad \text{(A.142)}$$

Thus, the consumer's expected utility is non-increasing in his degree of loss aversion if

$$\int_{\underline{\theta}}^{\bar{\theta}} \hat{q}(\theta)[2F(\theta)-1]f(\theta)d\theta \geq 0 \,. \quad \text{(A.143)}$$

Condition (A.143) is satisfied since $\hat{q}(\theta)$ is non-decreasing in θ. Nevertheless, we prove this claim formally. Let θ^M be the median state, i.e., $F(\theta^M) = 1/2$. By splitting the integral of (A.143) into two parts, the states below the median and the states above the median, we obtain

$$\begin{aligned}\int_{\underline{\theta}}^{\bar{\theta}} \hat{q}(\theta)[2F(\theta)-1]f(\theta)d\theta = \\ \int_{\underline{\theta}}^{\theta^M} \hat{q}(\theta)[2F(\theta)-1]f(\theta)d\theta + \int_{\theta^M}^{\bar{\theta}} \hat{q}(\theta)[2F(\theta)-1]f(\theta)d\theta \,. \quad \text{(A.144)}\end{aligned}$$

Since $\hat{q}(\theta)$ is non-decreasing, the above expression is at least as great as

$$\begin{aligned}\int_{\underline{\theta}}^{\theta^M} \hat{q}(\theta^M)[2F(\theta)-1]f(\theta)d\theta + \int_{\theta^M}^{\bar{\theta}} \hat{q}(\theta^M)[2F(\theta)-1]f(\theta)d\theta \\ = \hat{q}(\theta^M)\left[\int_{\theta^M}^{\bar{\theta}}[2F(\theta)-1]f(\theta)d\theta - \int_{\underline{\theta}}^{\theta^M}[1-2F(\theta)]f(\theta)d\theta\right] \,. \quad \text{(A.145)}\end{aligned}$$

By using integration by parts one can show that

$$\int_{\underline{\theta}}^{\theta^M} [1-2F(\theta)]f(\theta)d\theta = \int_{\underline{\theta}}^{\theta^M} 2F(\theta)f(\theta)d\theta, \quad \text{(A.146)}$$

and

$$\int_{\theta^M}^{\bar{\theta}} [2F(\theta)-1]f(\theta)d\theta = \int_{\underline{\theta}}^{\theta^M} 2[1-F(\theta)]f(\theta)d\theta. \quad \text{(A.147)}$$

Hence, the right-hand side of (A.145) is non-negative if

$$\int_{\theta^M}^{\bar{\theta}} f(\theta)d\theta - \int_{\underline{\theta}}^{\bar{\theta}} F(\theta)f(\theta)d\theta \geq 0, \tag{A.148}$$

which holds true since $1/2 \geq 1/2$.

Q.E.D.

Proof of Proposition IV.4: Irrespectively of the rival's tariff offer, if the sorting constraint is satisfied it is optimal for a firm to choose p_j such that $S_j(p_j)$ is maximized. Put differently, the firm will choose the method of generating v_j that maximizes its (per customer) profits. Thus, if no type $\lambda \in \{\lambda_1, \lambda_2\}$ has an incentive to mimic the other type, it is an equilibrium that the firms offer the same tariffs as in the full information case. Obviously, in case (ii) where $c \leq \Sigma(\lambda_1) < \Sigma(\lambda_2)$, both firms offer a flat-rate tariff to consumers. In this case, a flat-rate tariff maximizes $S_1(p)$ as well as $S_2(p)$. Moreover, the generated joint surplus is the same for both types of loss averse consumers. Since the brand preferences are i.i.d. across the λ_1 and λ_2 types, in any equilibrium each firm offers a single flat-rate tariff to consumers.

In the remaining part of the proof we show hat in the case where $\Sigma(\lambda_1) < c \leq \Sigma(\lambda_2)$ neither type λ_1 has an incentive to choose the tariff $(0, L^F)$ nor does type λ_2 have an incentive to choose the tariff $(0, L^F)$.

Claim: $\hat{v}_1 \geq \hat{v}_2$.

Proof: Let $S_j^* \equiv \max_p\{S_j(p)\}$. Note that $S_1(0) = S_2(0) = S_2^*$. The firm's per customer profit from type $j = 1, 2$ when offering utility v is

$$\pi_j(v) = S_j^* - v \ . \tag{A.149}$$

Thus, for any v it holds that $\pi_1(v) \geq \pi_2(v)$, since $S_1^* - v \geq S_2^* - v$. The equilibrium utilities are characterized by $\pi_j(\hat{v}_j) = \Phi(\hat{v}_j)$. Hence, we obtain the following relations:

$$\Phi(\hat{v}_1) = \quad \pi_1(\hat{v}_1) \geq \pi_2(\hat{v}_1) \tag{A.150}$$

$$\pi_1(\hat{v}_2) \geq \pi_2(\hat{v}_2) \quad = \Phi(\hat{v}_2) \ . \tag{A.151}$$

Suppose, in contradiction, $\hat{v}_1 < \hat{v}_2$. This immediately implies that $\pi_j(\hat{v}_1) > \pi_j(\hat{v}_2)$. Hence,

$$\Phi(\hat{v}_1) = \pi_1(\hat{v}_1) > \pi_1(\hat{v}_2) \geq \pi_2(\hat{v}_2) = \Phi(\hat{v}_2) \ . \tag{A.152}$$

Since $\Phi'(v) \geq 0$ the above formula holds only if $\hat{v}_1 > \hat{v}_2$, a contradiction. Q.E.D.

Since $\hat{v}_1 \geq \hat{v}_2$ and the expected utility from a flat-rate tariff being independent of λ, one can conclude that a consumer of type λ_1 has no incentive to choose the tariff $(0, L^F)$ that is designed for consumers of type λ_2. Finally, we show that type λ_2 has no incentive to mimic type λ_1. Let v_2^{DEV} denote the expected utility of a consumer of type λ_2 who accepts the tariff $(\hat{p}, \hat{L})$ designed for type λ_1.

Claim: $v_2^{DEV} < \hat{v}_2$.

Proof: The expected utility of type λ_2 from the tariff $(\hat{p}, \hat{L})$ equals the generated joint surplus minus the profits of the firm he purchases from. Thus,

$$v_2^{DEV} = S_2(\hat{p}) - \hat{L} - (\hat{p} - c) \int_{\underline{\theta}}^{\bar{\theta}} \hat{q}_2(\theta; \hat{p}) f(\theta) d\theta, \tag{A.153}$$

where $\hat{q}_2(\theta; p)$ denotes the demand of type λ_2 in the personal equilibrium. Inserting the explicit formula of $\hat{L}$ into (A.153) yields

$$v_2^{DEV} = S_2(\hat{p}) - \Phi(\hat{v}_1) - (\hat{c} - p) \int_{\underline{\theta}}^{\bar{\theta}} [\hat{q}_1(\theta; \hat{p}) - \hat{q}_2(\theta; \hat{p})] f(\theta) d\theta. \tag{A.154}$$

Note that $\hat{q}_1(\theta, \hat{p}) > \hat{q}_2(\theta, \hat{p})$ for all $\theta \in \Theta$, since $d\hat{q}/d\lambda < 0$ if $p > 0$. By Proposition 3, $c > \hat{p}$ hence

$$v_2^{DEV} < S_2(\hat{p}) - \Phi(\hat{v}_1). \tag{A.155}$$

The expected utility of a consumer of type λ_2 when choosing the tariff that is designed for him can be expressed as follows,

$$\hat{v}_2 = S_2^* - \Phi(\hat{v}_2). \tag{A.156}$$

Hence, a deviation is not utility improving if

$$S_2^* - \Phi(\hat{v}_2) \geq S_2(\hat{p}) - \Phi(\hat{v}_1) \tag{A.157}$$

$$\iff [S_2^* - S_2(\hat{p})] + [\Phi(\hat{v}_1) - \Phi(\hat{v}_2)] \geq 0 \,. \tag{A.158}$$

The above inequality is satisfied since $\Phi'(\cdot) \geq 0$ and $\hat{v}_1 \geq \hat{v}_2$. Q.E.D.

Thus, if the firms offer the optimal tariffs of the full information case, each type of loss averse consumer selects the tariff that is designed for him, which completes the proof.

Q.E.D.

4.2. Examples of Discrete Choice Models

Hotelling Model with Linear Transport Cost. Suppose consumers' ideal brands are uniformly distributed on the unit interval $[0,1]$. The brands of the two firms, A and B, are located at the two extreme points, brand A at zero and brand B at one. A consumer with ideal brand $x \in [0,1]$ has brand preferences $\zeta = (0, -tx, -t(1-x))$. The parameter $t > 0$ is a consumer's "transport cost" per unit distance between his ideal brand and the brand he purchases from. For the Hotelling specification the market share function takes the following form,

$$m(v^A, v^B) = \min\left\{\frac{1}{2t}(t + v^A - v^B), \frac{v^A}{t}\right\}. \tag{A.159}$$

The market share function has to be modified if v^A and v^B differ by so much that $m(\cdot) \notin [0,1]$ (this never happens in equilibrium). Moreover, the Hotelling model has the well-known drawback that market shares are kinked. If, however, the transport cost is sufficiently low then one can focus on the case where the market share function is given by the first term of the above expression and thus well behaved. Formally, for $t \leq (2/3)S_2^*$ it suffices to analyze firms' profit maximization problem for[3]

$$m(v^A, v^B) = [1/(2t)](t + v^A - v^B). \tag{A.160}$$

Hence, $\partial m(v^A, v^B)/\partial v^A = (2t)^{-1}$ which immediately implies that

$$\Phi(v) \equiv \frac{m(v,v)}{\partial m(v,v)/\partial v^A} = t. \tag{A.161}$$

Obviously, $\Phi(\cdot)$ is non-decreasing. Note that

$$\frac{\partial m(v^A, v^B)/\partial v^A}{m(v^A, v^B)} = (t + v^A - v^B)^{-1}. \tag{A.162}$$

It can easily be seen that the above fraction is increasing in v^B. Thus, the Hotelling model satisfies all imposed assumptions if the transport cost is sufficiently low. One can check that the collusive utility level exists. To calculate the collusive utility level one has to use the market share function given in (A.159).

Logit Demand Model. An obvious drawback of the Hotelling specification is that a firm does not compete with the rival and the outside option at the same time. A model that accounts for this simultaneous competition on two fronts is the logit demand model. Here, a consumer's brand preferences ζ^i for $i = 0, A, B$ are i.i.d. according to

[3]See Lemma 1 of Armstrong and Vickers (2001).

the double exponential distribution with mean zero and variance $\mu^2\pi^2/6$, where π (here) denotes the circular constant. Thus, the cumulative distribution function is

$$G(\zeta^i) = \exp\big\{-\exp[-(\gamma + \zeta^i/\mu)]\big\}, \tag{A.163}$$

where γ is the Euler–Mascheroni constant and μ is a positive constant. With this specification, the market share of firm A is given by (see Anderson et al. 1992)

$$m(v^A, v^B) = \frac{\exp[v^A/\mu]}{\exp[v^A/\mu] + \exp[v^B/\mu] + 1}. \tag{A.164}$$

The parameter μ captures the degree of heterogeneity among consumers with respect to their brand preferences. Put differently, μ measures the degree of product differentiation. A lower value of μ corresponds to a more competitive market. For $\mu \to \infty$ the firms are local monopolists. Taking the partial derivative of (A.164) with respect to v^A yields

$$\frac{\partial m(v^A, v^B)}{\partial v^A} = \frac{\exp[v^A/\mu]\{\exp[v^B/\mu] + 1\}}{\mu\{\exp[v^A/\mu] + \exp[v^B/\mu] + 1\}^2}. \tag{A.165}$$

Thus,

$$\frac{m(v^A, v^B)}{\partial m(v^A, v^B)/\partial v^A} = \frac{\mu\{\exp[v^A/\mu] + \exp[v^B/\mu] + 1\}}{\exp[v^B/\mu] + 1}. \tag{A.166}$$

Evaluating the above expression at $v^A = v^B = v$ leads to

$$\Phi(v) = \mu\frac{2\exp[v/\mu] + 1}{\exp[v/\mu] + 1}. \tag{A.167}$$

Taking the derivative of $\Phi(\cdot)$ with respect to v yields

$$\Phi'(v) = \frac{\exp[v/\mu]}{(\exp[v/\mu] + 1)^2} > 0. \tag{A.168}$$

Moreover, the derivative of $[\partial m(v^A, v^B)/\partial v^A][m(v^A, v^B)]^{-1}$ with respect to v^B amounts to

$$\frac{d}{dv^B}\left[\frac{\partial m(v^A, v^B)/\partial v^A}{m(v^A, v^B)}\right] = \frac{1}{\mu^2}\,\frac{\exp[v^B/\mu]\{\exp[v^B/\mu] + 1\}}{\mu\{\exp[v^B/\mu] + \exp[v^B/\mu] + 1\}^2} > 0. \tag{A.169}$$

The collusive utility level $\tilde{v}$ maximizes $m(v, v)\pi(v)$. Note that $m(v, v) \to 0$ for $v \to -\infty$ and $\pi(v) \leq 0$ if $v \geq \max_p\{S(p)\}$. Thus, the collusive utility exists, since $m(v, v)\pi(v)$ is continuously differentiable.

Bibliography

[1] **Abeler, J.**, **A. Falk**, **L. Götte**, and **D. Huffman (2009):** Reference Points and Effort Provision, *IZA Discussion Paper*, No. 3939, http://ftp.iza.org/dp3939.pdf.

[2] **Akerlof, G.A. (1991):** Procrastination and Obedience, *American Economic Review*, Vol. 81, 1-19.

[3] **Akin, Z. (2007):** Time Inconsistency and Learning in Bargaining Games, *International Journal of Game Theory*, Vol. 36, 275-299.

[4] **Amador, M.**, **I. Werning**, and **G.-M. Angeletos, G.-M. (2006):** Commitment vs. Flexibility, *Econometrica*, Vol. 74, 365-396.

[5] **Anderson, S.P.**, **A. dePalma** and **J.-F. Thisse (1992):** Discrete Choice Theory of Product Differentiation, *MIT Press*, Cambridge.

[6] **Anderson, S.** and **M.Engers (1994):** Spatial competition with price-taking firms, *Economica*, Vol.61, 125-136.

[7] **Angeletos, G.-M.**, **D. Laibson**, **A. Repetto**, **J. Tobacman** and **S.Weinberg (2001):** The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation, *Journal of Economic Perspectives*, Vol. 15, 47-68.

[8] **Ariely, D.** and **K. Wertenbroch (2002):** Procrastination, Deadlines, and Performance: Self-Control by Precommitment, *Psychological Science*, Vol. 13, 219-224.

[9] **Armstrong, M.** and **J.Vickers (2001):** Competitive price discrimination, *RAND Journal of Economics*, Vol. 32(4), 579-605.

[10] **Auriol, E.** and **R. Renault (2008):** Status and incentives, *RAND Journal of Economics*, Vol. 39, 305-326.

[11] **Avenel, E.** and **S. Caprice (2006):** Upstream market power and product line differentiation in retailing, *International Journal of Industrial Organization*, Vol.24, 319-334.

[12] **Barberis, N.**, **M. Huang** and **T. Santos (2001):** Prospect Theory and Asset Prices, *Quarterly Journal of Economics*, Vol. 116, 1-53.

[13] **Becker, G.S.** and **G.J. Stigler (1974):** Law Enforcement, Malfeasance, and Compensation of Enforcers, *Journal of Legal Studies*, Vol. 3, 1-18.

[14] **Bell, D.E. (1985):** Disappointment in Decision Making under Uncertainty, *Operations Research*, Vol. 33, 1-27.

[15] **Bénabou, R.** and **J. Tirole (2002):** Self-Confidence and Personal Motivation, *Quarterly Journal of Economics*, Vol. 117, 871-915.

[16] **Bertrand, J.L.F. (1883):** Théorie des Richesses: revue de Théories mathématiques de la richesse sociale par Léon Walras et Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot, *Journal des Savants*, Vol. 30, 499-508.

[17] **Bertsekas, D. (1974):** Necessary and Sufficient Conditions for Existence of an Optimal Portfolio, *Journal of Economic Theory*, Vol. 8, 235-247.

[18] **Besley, T.** and **M. Ghatak (2008):** Status Incentives, *American Economic Review*, Vol. 98, 206-211.

[19] **Blackwell, D. (1951):** Comparison of Experiments, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, editor, University of California Press, Berkeley.

[20] **Blackwell, D. (1953):** Equivalent Comparison of Experiments, *Annals of Mathematics and Statistics*, Vol. 24, 265-272.

[21] **Bonanno, G.** and **J. Vickers (1988):** Vertical Separation, *The Journal of Industrial Economics*, Vol.36(3), 257-265.

[22] **Borch, K. (1962):** Equilibrium in a Reinsurance Market, *Econometrica*, Vol. 30, 424-444.

[23] **Bowman, D.**, **D. Minehart** and **M. Rabin (1999):** Loss-Aversion in a Consumptions-Savings Model, *Journal of Economic Behavior and Organization*, Vol. 38, 155-178.

[24] **Breiter, H.C.**, **I. Aharon**, **D. Kahneman**, **A. Dale** and **P. Shizgal (2001):** Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains and Losses, *Neuron*, Vol. 30, 619-639.

[25] **Camerer, C.F.** and **U. Malmendier (2007):** Behavioral Economics of Organizations, in *Behavioral Economics and its Applications*, P. Diamond and H. Vartiainen, editors, Princeton University Press, Princeton, 235-281.

[26] **Caron, L.** and **G. Dionne (1997):** Insurance Fraud Estimation: More Evidence from the Québec Automobile Insurance Industry, *Assurances*, Vol. 64, 567-578.

[27] **Carrillo, J.D.**, and **T. Mariotti (2000):** Strategic Ignorance as Self-Disciplining Device, *Review of Economic Studies*, Vol. 67, 529-544.

[28] **Champsaur, P.** and **J-C. Rochet (1989):** Multiproduct Duopolists, *Econometrica*, Vol. 57(3), 533-557.

[29] **Chiappori, P.-A.**, **B. Jullien**, **B. Salanié**, and **F. Salanié (2006):** Asymmetric information in insurance: general testable implications, *RAND Journal of Economics*, Vol. 37, 783-798.

[30] **Choi, C.J.** and **H.S. Shin (1992):** A comment on a Model of vertical Product Differentiation, *The Journal of Industrial Economics*, 229-231.

[31] **Churchill, G.A.**, **N.M. Ford**, and **O.C. Walker (1993):** *Sales Force Management*, Irwin, Homewood.

[32] **Clay K.B.**, **D.S. Sibley** and **P. Srinagesh (1992):** Ex Post vs. Ex Ante Pricing: Optimal Calling Plans and Tapered Tariffs, *Journal of Regulatory Economics*, Vol. 4, 115-138.

[33] **Corts, K.S. (1998):** Third-degree price discrimination in oligopoly: all-out competition and strategic commitment, *RAND Journal of Economics*, Vol. 29(2), 306-323.

[34] **Cournot, A.A. (1838):** Recherches sur les Principes Mathématiques de la Théorie des Richesses, Paris. English translation (1897): N. Bacon, trans., Researches into the Mathematical Principles of the Theory of Wealth, : *Macmillan & Company*, New York.

[35] **Courty, P.** and **H. Li (2000):** Sequential Screening, *Review of Economic Studies*, Vol. 67, 697-717.

[36] **Daido, K.** and **H. Itoh (2007):** The Pygmalion and Galatea Effects: An Agency Model with Reference-Dependent Preferences and Applications to Self-Fulfilling Prophecy, *Working Paper*, Hitotsubashi University.

[37] **Dantzig, G.B. (1957):** Discrete Variable Extremum Problems, *Operations Research*, Vol. 5, 266-277.

[38] **DellaVigna, S.** and **U. Malmendier (2004):** Contract Design and Self-Control: Theory and Evidence, *Quarterly Journal of Economics*, Vol. 119, 353-402.

[39] **DellaVigna, S.** and **U. Malmendier (2006):** Paying Not to Go to the Gym, *American Economic Review*, Vol. 96, 694-719.

[40] **DellaVigna, S.** and **M.D. Paserman (2005):** Job Search and Impatience, *Journal of Labor Economics*, Vol. 23, 527-588.

[41] **De Meza, D.** and **D.C. Webb (2007):** Incentive Design Under Loss Aversion, *Journal of the European Economic Association*, Vol. 5, 66-92.

[42] **Demougin, D.** and **C. Fluet (1998):** Mechanism Sufficient Statistic in the Risk-Neutral Agency Problem, *Journal of Institutional and Theoretical Economics*, Vol. 154, 622-639.

[43] **Desai, P. (2001):** Quality Segmentation in Spatial Markets: When Does Cannibalization Affect Product Line Design?, *Marketing Science*, Vol.20, 265-283.

[44] **Diamond, P.** and **B. Kőszegi (2003):** Quasi-Hyperbolic Discounting and Retirement, *Journal of Public Economics*, Vol. 87, 1839-1872.

[45] **Dionne, G.** and **R. Gagné (2001):** Deductible Contracts Against Fraudulent Claims: Evidence from Automobile Insurance, *Review of Economics and Statistics*, Vol. 83, 290-301.

[46] **Dittmann, I.**, **E. Maug** and **O. Spalt (2007):** Executive Stock Options When Managers Are Loss Averse, *Working Paper*, University of Mannheim.

[47] **Donnenfeld, S.** and **S. Weber (1992):** Vertical Product Differentiation with Entry, *International Journal of Industrial Organization*, Vol.10, 449-472.

[48] **Donnenfeld, S.** and **S. Weber (1995):** Limit Qualities and Entry Deterrence, *RAND Journal of Economics*, Vol.26, 113-130.

[49] **Eliaz, K.** and **R. Spiegler (2006):** Contracting with Diversely Naive Agents, *Review of Economic Studies*, Vol. 73, 689–714.

[50] **Eliaz, K.** and **R. Spiegler (2008):** Consumer Optimism and Price Discrimination, *Theoretical Economics*, Vol. 3(4), 459–497.

[51] **Englmaier, W.** and **A. Wambach (2006):** Optimal Incentive Contracts under Inequity Aversion, *Working Paper*, University of Cologne.

[52] **Esteban, S.** and **E. Miyagawa (2006):** Temptation, Self-Control, and Competitive Nonlinear Pricing, *Economics Letters*, Vol. 90, 348–355.

[53] **Esteban, S.**, **E. Miyagawa** and **M. Shum (2007):** Nonlinear Pricing with Self-Control Preferences, *Journal of Economic Theory*, Vol. 135, 306-338.

[54] **Fang, H.** and **G. Moscarini (2005):** Morale hazard, *Journal of Monetary Economics*, Vol. 52, 749-777.

[55] **Faulhaber, G.R.** and **C. Hogendorn (2000):** The Market Structure of Broadband Telecommunications, *Journal of Industrial Economics*, Vol. 68(3), 305-329.

[56] **Fehr, E.** and **K.M. Schmidt (1999):** A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics*, Vol. 114, 817-868.

[57] **Fischer, C. (1999):** Read this Paper Even Later: Procrastination with Time-Inconsistent Preferences, *Resources for the Future discussion paper 99-20.*

[58] **Fischer, C. (2001):** Read this Paper Later: Procrastination with Time-Consistent Preferences, *Journal of Economic Behavior and Organization*, Vol. 46, 249-269.

[59] **Foppert, D. (1994):** Waging War Against Fraud, *Best's Review: Property-Causality*, Ed. 94.

[60] **Frederick, S.**, **G. Loewenstein** and **T. O'Donoghue (2002):** Time Discounting and Time Preference, *Journal of Economic Literature*, Vol. 40, 351-401.

[61] **Gabszewicz, J.J.** and **J.-F. Thisse (1979):** Price Competition, Quality, and Income Disparities, *Journal of Economic Theory*, Vol.20, 340-359.

[62] **Gerpott T.J. (2009):** Biased Choice of a Mobile Telephony Tariff Type: Exploring Usage Boundary Perceptions as a Cognitive Cause in Choosing Between a Used-Based or a Flat-Rate Plan, *Telematics and Informatics*, Vol. 26, 167-179.

[63] **Gjesdal, F. (1982):** Information and Incentives: The Agency Information Problem, *Review of Economic Studies*, Vol. 49, 373-390.

[64] **Goldman, M.B.**, **H.L. Leland** and **D.S. Sibley (1984):** Optimal Nonuniform Pricing, *Review of Economic Studies*, Vol. 51, 305-319.

[65] **Grossman, S.** and **O. Hart (1983):** An Analysis of the Principal-Agent Problem, *Econometrica*, Vol. 51, 7-45.

[66] **Grubb, M. (forthcoming):** Selling to Overconfident Consumers, *American Economic Review.*

[67] **Gruber, J.** and **B. Kőszegi (2001):** Is Addiction Rational? Theory and Evidence, *Quarterly Journal of Economics*, Vol. 116, 1261-1303.

[68] **Gul, F. (1991):** A Theory of Disappointment Aversion, *Econometrica*, Vol. 59, 667-686.

[69] **Gul, F.** and **W. Pesendorfer (2001):** Temptation and Self Control, *Econometrica*, Vol. 60, 1293-1330.

[70] **Haller, H. (1985):** The Principal-Agent Problem with a Satisficing Agent, *Journal of Economic Behavior and Organization*, Vol. 6, 359-379.

[71] **Heidhues, P.** and **B. Kőszegi (2005):** The Impact of Consumer Loss Aversion on Pricing, *CEPR Discussion Paper*, No. 4849, http://ssrn.com/abstract=717706.

[72] **Heidhues, P.** and **B. Kőszegi (2008a):** Competition and Price Variation when Consumers are Loss Averse, *American Economic Review*, Vol. 98, 1245–1268.

[73] **Heidhues, P.** and **B. Kőszegi (2008b):** Exploiting Naivete about Self-Control in the Credit Market, *Working Paper*, University of Berkeley.

[74] **Herweg, F.**, **D. Müller** and **P. Weinschenk (2008):** The Optimality of Simple Contracts: Moral Hazard and Loss Aversion, *Bonn Econ Discussion Paper*, University of Bonn.

[75] **Hobson, M.** and **R.H. Spady (1988):** The Demand for Local Telephone Service Under Optimal Local Measured Service, *Bellcore Economics Discussion Paper*, No. 50.

[76] **Holmström, B. (1979):** Moral Hazard and Observability, *Bell Journal Of Economics*, Vol. 10, 74-91.

[77] **Holmström, B. (1982):** Moral Hazard in Teams, *Bell Journal Of Economics*, Vol. 13, 324-340.

[78] **Hotelling, H. (1929):** The Stability of Competition, *Economic Journal*, Vol. 39, 41-57.

[79] **Iantchev, E.P. (2005):** Optimal Contracts with Prospect Theory Preferences in the Presence of Moral Hazard, *Working Paper*, University of Chicago.

[80] **Jensen, S. (2008):** Two-part tariffs with quality degradation, *International Journal of Industrial Organization*, Vol.26, 473-489.

[81] **Jewitt, I.**, **O. Kadan**, and **J.M. Swinkels (2008):** Moral Hazard with Bounded Payments, *Journal of Economic Theory*, Vol. 143, 59-82.

[82] **Johnson, J.P.** and **D.P. Myatt (2003):** Multiproduct Quality Competition: Fighting Brands and Product Line Pruning, *The American Economic Review*, Vol.93(3), 748-774.

[83] **Joseph, K.** and **M.U. Kalwani (1998):** The Role of Bonus Pay in Salesforce Compensation Plans, *Industrial Marketing Management*, Vol. 27, 147-159.

[84] **Kahneman, D.** and **A. Tversky (1979):** Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, Vol. 47, 263-291.

[85] **Katz, N.L. (1984):** Price Discrimination and Monopolistic Competition, *Econometrica*, Vol.52, 1453-1471.

[86] **Kim, S.K. (1995):** Efficiency of an Information System in an Agency Model, *Econometrica*, Vol. 63, 89-102.

[87] **Kim, S.K. (1997):** Limited Liability and Bonus Contracts, *Journal of Economics & Management Strategy*, Vol. 6, 899-913.

[88] **Kling, J.P.** and **S.S. van der Ploeg (1990):** Estimating Local Elasticities with a Model of Stochastic Class of Service and Usage, in *Telecommunications Demand Modeling: An Integrated View*, A. de Fontenay, M.H. Shugard and D.S. Sibley, editors, Amsterdam, North Holland, 119-136.

[89] **Kőszegi, B.** and **M. Rabin (2006):** A Model of Reference-Dependent Preferences, *Quarterly Journal of Economics*, Vol. 121, 1133-1165.

[90] **Kőszegi, B.** and **M. Rabin (2007):** Reference-Dependent Risk Preferences, *American Economic Review*, Vol. 97, 1047-1073.

[91] **Kreps, D.M.** and **J.A. Scheinkman (1983):** Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes, *Bell Journal of Economics*, Vol. 14(2), 326-337.

[92] **Kridel, D.J.**, **D.E. Lehman** and **D.L. Weisman (1993):** Option Value, Telecommunication Demand, and Policy, *Information Economics and Policy*, Vol. 5 (2), 125-144.

[93] **Laibson, D. (1996):** Hyperbolic Discount Functions, Undersaving, and Savings Policy, *NBER Working Paper Series*, No.5635, Cambridge, MA.

[94] **Laibson, D. (1997):** Golden Eggs and Hyperbolic Discounting, *Quarterly Journal of Economics*, Vol. 112, 443-477.

[95] **Laibson, D. (1998):** Life-Cycle Consumption and Hyperbolic Discount Functions, *European Economic Review*, Vol. 42, 861-871.

[96] **Laibson, D.**, **A. Repetto** and **J. Tobacman (1998):** Self-Control and Saving for Retirement, *Brookings Papers on Economic Activity*, Vol. 1, 91-196.

[97] **Lambrecht, A.** and **B. Skiera (2006):** Paying Too Much and Being Happy About It: Existence, Causes, and Consequences of Tariff-Choice Biases, *Journal of Marketing Research*, Vol.63 (3), 212-223.

[98] **Lambrecht, A.**, **K. Seim** and **B. Skiera (2007):** Does Uncertainty Matter? Consumer Behavior Under Three-Part Tariffs, *Marketing Science*, Vol.26 (5), 698-710.

[99] **Lancaster, K.J. (1979):** Variety, Equity and Efficiency, *Columbia University Press*, New York.

[100] **Larsen, J.T.**, **A.P. Mc Graw**, **B.A. Mellers**, and **J.T. Cacioppo (2004):** The Agony of Victory and Thrill of Defeat: Mixed Emotional Reactions to Disappointing Wins and Relieving Losses, *Psychological Science*, Vol. 15, 325-330.

[101] **Lazear, E.P.** and **K.L. Shaw (2007):** Personnel Economics: The Economist's View of Human Resources, *Journal of Economic Perspectives*, Vol. 21, 91-114.

[102] **Lazear, E.P.** and **P. Oyer (2007):** Personnel Economics, *NBER Working Paper 13480*, http:// www.nber.org/papers/w13480.

[103] **Leland, H.L.** and **R.A. Meyer (1976):** Monopoly Pricing Structures with Imperfect Discrimination, *Bell Journal of Economics*, Vol. 7, 449-462.

[104] **Loomes, G.** and **R. Sugden (1986):** Disappointment and Dynamic Consistency in Choice under Uncertainty, *Review of Economic Studies*, Vol. 53, 271-282.

[105] **MacLeod, B. (2003):** Optimal Contracting with Subjective Evaluation, *American Economic Review*, Vol. 93, 216-240.

[106] **Maskin, E.** and **J. Riley (1984):** Monopoly with Incomplete Information, *RAND Journal of Economics*, Vol. 15, 171-196.

[107] **Mellers, B.**, **A. Schwartz** and **H. Ritov (1999):** Emotion-Based Choice, *Journal of Experimental Psychology: General*, Vol. 128, 332-345.

[108] **Miravete E.J. (2002):** Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans, *Review of Economic Studies*, Vol. 69, 943-971.

[109] **Miravete E.J. (2003):** Choosing the Wrong Calling Plan? Ignorance and Learning, *American Economic Review*, Vol. 93 (1), 297-310.

[110] **Mitchell, B.M.** and **I. Vogelsang (1991):** Telecommunications Pricing: Theory and Practice, *Cambridge University Press*, Cambridge.

[111] **Mitomo, H.**, **T. Otsuka** and **K. Nakaba (2009):** A Behavioral Economic Interpretation of the Preferences for Flat Rates: A Case of Post-Paid Mobile Phone Services, *mimeo*, Waseda University.

[112] **Moynahan, J.K. (1980):** *Designing an Effective Sales Compensation Program*, AMACOM, New York.

[113] **Mussa, M.** and **S. Rosen (1978):** Monopoly and Product Quality, *Journal of Economic Theory*, Vol.18, 301-317.

[114] **Nocke, V.** and **M. Peitz (2003):** Hyperbolic Discounting and Secondary Markets, *Games and Economic Behavior*, Vol. 44, 77-97.

[115] **Nunes, J.C. (2000):** A Cognitive Model of People's Usage Estimations, *Journal of Marketing Research*, Vol. 37, 397-409.

[116] **O'Donoghue, T.** and **M. Rabin (1999a):** Addiction and Self-Control, in *Addiction: Entry and Exits*, J. Elster, editor, Russel Sage Foundation.

[117] **O'Donoghue, T.** and **M. Rabin (1999b):** Doing It Now Or Later, *American Economic Review*, Vol. 89, 103-124.

[118] **O'Donoghue, T.** and **M. Rabin (1999c):** Incentives for Procrastinators, *Quarterly Journal of Economics*, Vol. 114, 769-816.

[119] **O'Donoghue, T.** and **M. Rabin (2001a):** Addiction and Present-Biased Preferences, *working paper*, University of California, Berkeley.

[120] **O'Donoghue, T.** and **M. Rabin (2001b):** Choice and Procrastination, *Quarterly Journal of Economics*, Vol. 116, 121-160.

[121] **O'Donoghue, T.** and **M. Rabin (2005):** Incentives and Self-Control, *working paper*, University of California, Berkeley.

[122] **O'Donoghue, T.** and **M. Rabin (2008):** Procrastination on Long-Term Projects, *Journal of Economic Behavior and Organization*, Vol. 66, 161-175.

[123] **Oi, W.Y. (1971):** A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly, *Quarterly Journal of Economics*, Vol. 85, 77-96.

[124] **Oyer, P. (1998):** Fiscal Year Ends and Non-Linear Incentive Contracts: The Effect on Business Seasonality, *Quarterly Journal of Economics*, Vol. 113, 149-188.

[125] **Oyer, P. (2000):** A Theory of Sales Quotas with Limited Liability and Rent Sharing, *Journal of Labor Economics*, Vol 18, 405-426.

[126] **Park, E.-S. (1995):** Incentive Contracting under Limited Liability, *Journal of Economics & Management Strategy*, Vol. 4, 477-490.

[127] **Peitz, M.** and **T.M. Valletti (2008):** Content and advertising in the media: Pay-tv versus free-to-air, *International Journal of Industrial Organization*, Vol.26, 949-965.

[128] **Phelps, E.S.** and **R.A. Pollak (1968):** On Second-Best National Saving and Game-Equilibrium Growth, *Review of Economic Studies*, Vol. 35, 185-199.

[129] **Pigou, A.C. (1920):** The Economics of Welfare, 2nd edition (1924), *Macmillian Press*, London.

[130] **Pollak, R.A. (1968):** Consistent Planning, *Review of Economic Studies*, Vol. 35, 201-208.

[131] **Prendergast, C. (1999):** The Provision of Incentives in Firms, *Journal of Economic Literature*, Vol. 37, 7-63.

[132] **Post, T.**, **M.J. Van den Assem**, **G. Baltussen** and **R.H. Thaler (2008):** Deal Or No Deal? Decision Making Under Risk in a Large-Payoff Game Show, *American Economic Review*, Vol. 98, 38-71.

[133] **Puelz, R.** and **A. Snow (1994):** Evidence on Adverse Selection: Equilibrium Signalling and Cross-Subsidization in the Insurance Market, *Journal of Political Economy*, Vol. 102, 236-257.

[134] **Rabin, M. (2000):** Risk Aversion and Expected-Utility Theory: A Calibration Theorem, *Econometrica*, Vol. 68(5), 1281-1292.

[135] **Rayo, L.** and **G.S. Becker (2007):** Evolutionary Efficiency and Happiness, *Journal of Political Economy*, Vol. 115, 302-337.

[136] **Rey, P.** and **J. Stiglitz (1988):** Vertical Restraints and Producers' Competition, *European Economic Review*, Vol.32, 561-568.

[137] **Rochet, J.-C.** and **L.A. Stole (2002):** Nonlinear Pricing with Random Participation, *Review of Economic Studies*, Vol.69, 277-311.

[138] **Rothschild, M.** and **J.E. Stiglitz (1976):** Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics*, Vol. 90, 639-649.

[139] **Salanié, B. (2003):** Testing Contract Theory, *CESifo Economic Studies*, Vol. 49, 461-477.

[140] **Samuelson, P.A. (1937):** A Note on Measurement of Utility, *Review of Economic Studies*, Vol. 4, 155-161.

[141] **Schmidt, U. (1999):** Moral Hazard and First-Order Risk Aversion, *Journal of Economics*, Supplement 8, 167-179.

[142] **Shaked, A.** and **J. Sutton (1982):** Relaxing Price Competition Through Product Differentiation, *Review of Economic Studies*, Vol.49, 3-13.

[143] **Shavell, S. (1979):** Risk Sharing and Incentives in the Principal and Agent Relationship, *Bell Journal Of Economics*, Vol. 10 (1), 55-73.

[144] **P.M. Skiba** and **J. Tobacman (2008):** Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default, *working paper*, University of Pennsylvania.

[145] **Spulber, D.F. (1989):** Product Variety and Competitive Discounts, *Journal of Economic Theory*, Vol.48, 510-525.

[146] **Steenburgh, T.J. (2008):** Effort or Timing: The Effect of Lump-Sum Bonuses, *Quantitative Marketing and Economics*, Vol. 6, 235-256.

[147] **Stole, L.A. (1995):** Nonlinear Pricing and Oligopoly, *Journal of Economics and Management Strategy*, Vol.4, 529-562.

[148] **Strausz, R. (2006):** Deterministic Versus Stochastic Mechanisms in Principal-Agent Models, *Journal of Economic Theory*, Vol. 128, 306-314.

[149] **Strotz, R.H. (1956):** Myopia and Inconsistency in Dynamic Utility Maximization, *Review of Economic Studies*, Vol. 23, 156-180.

[150] **Sundararajan, A. (2004):** Nonlinear Pricing of Information Goods, *Marketing Science*, Vol.50 (12), 1660-1673.

[151] **Thomson, W. (1999):** The Young Person's Guide to Writing Economic Theory, *Journal of Economic Literature*, Vol. 37(1), 157-183.

[152] **Tirole, J. (1988):** The Theory of Industrial Organization, *MIT Press*, Cambridge.

[153] **Tobacman, J. (2007):** The Partial Naivete Euler Equation, *working paper*, University of Pennsylvania.

[154] **Train, K.E.**, **D.L. McFadden** and **M. Ben-Akiva (1987):** The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choice, *RAND Journal of Economics*, Vol.18 (1), 109-123.

[155] **Train, K.E.**, **M. Ben-Akiva** and **T. Atherton (1989):** Consumption Patterns and Self-Selecting Tariffs, *Review of Economics and Statistics*, Vol.1, 62-73.

[156] **Train, K.E. (1991):** Optimal Regulation: The Economic Theory of Natural Monopoly, *MIT Press*, Cambridge.

[157] **Tversky, A.** and **D. Kahneman (1991):** Loss Aversion in Riskless Choice: A Reference-Dependent Model, *Quarterly Journal of Economics*, Vol. 106, 1039-1061.

[158] **Uthemann, A. (2005):** Competitive Screening of Customers with Non-Common Priors, *Working Paper*, University College London.

[159] **Wauthy, X. (1996):** Quality Choice in Models of Vertical Differentiation, *The Journal of Industrial Economics*, Vol.44(3), 345-353.

[160] **Willig, R.D. (1978):** Pareto-Superior Nonlinear Outlay Schedules, *Bell Journal of Economics*, Vol. 9, 56-69.

[161] **Wilson, R.B. (1993):** Nonlinear Pricing, *Oxford University Press*, New York.

[162] **Winter, R.A. (2000):** Optimal Insurance under Moral Hazard, in *Handbook of Insurance*, G. Dionne, editor, Kluwer Academic Publishers, Boston, 155-183.

[163] **Wong, W.-K. (2008):** How Much Time-Inconsistency Is There and Does It Matter? Evidence on Self-Awareness, Size, and Effects, *Journal of Economic Behavior and Organization*, Vol. 68, 645-656.

[164] **Yaari, M.E. (1987):** The Dual Theory of Choice under Risk, *Econometrica*, Vol. 55, 95-115.

[165] **Yin, X. (2004):** Two-part tariff competition in duopoly, *International Journal of Industrial Organization*, Vol.22, 799-820.