Essays in Applied Microeconomics and Management

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
Daniel Müller
aus Bonn

Bonn 2010

Dekan:	Prof. Dr. Christian Hillgruber
Erstreferent:	Prof. Dr. Urs Schweizer
Zweitreferent:	Prof. Dr. Matthias Kräkel

Tag der mündlichen Prüfung: 12.02.2010

Acknowledgements

In preparing this thesis I received support from many people to whom I am grateful. First and foremost, I want to thank Urs Schweizer for being a great supervisor. More a skeptic than an proponent of behavioral economics, it was Urs Schweizer who taught me some of the most important lessons about how to write a behavioral paper. Secondly, I want to thank Matthias Kräkel for acting as a referee on my thesis committee and for countless helpful comments on all chapters gathered in this thesis.

I am particularly grateful to Fabian Herweg. While jointly preparing for the micro final during our first year in the Bonn Graduate School of Economics, we realized that we had a knack for understanding each other—which should become one of the most valuable assets in the ensuing cooperation. Our joint research over the past four years has been a fruitful, usually fun, and inspiring activity. Chapters I, II, and IV of this thesis are testament to this cooperation and friendship.

I am indebted to Philipp Weinschenk, who also contributed to Chapter II. Our views on how to put things verbally often diverged, which gave rise to many heated debates about how the paper should be written or structured. Nevertheless, I believe we managed to strike a balance more often than not and that both of us are happy with the form the paper finally took.

I am grateful to Jörg Budde, whose topics course on the theory of incentives was—in a sense—the place of birth of Chapter II of this thesis.

I owe thanks to Paul Heidhues, whose interest and expertise in the field of behavioral economics had major impact on Chapters I and II of this thesis.

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, in particular Urs Schweizer—he not only does a great job as the BGSE's director/speaker, but also encouraged me to apply for the BGSE in the first place.

I want to thank both former and present members of the Chair of Personnel Economics at the University of Bonn, in particular Matthias Kräkel, Oliver Gürtler, Anja Schöttner, Petra Nieken, Judith Przemeck, Sebastien Kronisch, Simon Dato, and Wei

Ding. They have been great and very supportive colleagues, working with whom has always been fun.

Moreover, I owe thanks to Daniel Göller, Botond Kőszegi, Daniel Krämer, Sebastian "Reinhard" Kranz, Matthias Lang, Susanne Ohlendorf, Thomas Rieck, Andreas Roider, Patrick Schmitz, Nora Szech, Philipp Wichardt, and many other people.

I am very much indebted to my family and friends. In particular my parents were always supportive in every respect, even though it was probably them who bore the brunt of the negative externalities associated with a graduate student's life.

Finally, I want to thank Andrea Bedersdorfer. First, I owe her thanks for once in a while reminding me to think of subscripts i and j not as mere indices but as what they stand for—individuals, real people. Second, I am grateful to her, as well as our cats Luna and Athos, for so much more.

Contents

List of Figures

Introduction

Having spent most of my time and energy over the past four years on writing this doctoral thesis, I find myself somewhat reluctant to spend the pages of this introduction exclusively on emphasizing the importance and relevance of the insights I gained throughout this process for people both inside and outside the economic profession. A minor reason for this reluctance, I guess, is that I hope that my co-authors and I have convincingly done that already in the introductory sections of the papers collected in this thesis. A second more substantial reason, which is not that much rooted in the desire to avoid doing the same work twice, is that—in my opinion—much would be lost when focusing on the final output alone. Therefore, at this point, I want to pause, take stock, and take a look back at how this thesis came to be.

Three of four chapters gathered in this thesis are settled in the realms of "behavioral economics", which I first became aware of during a 10-month stay at the University of Berkeley. Over that time, it was the enthusiasm of a certain Professor Matthew Rabin, when talking about how insights from psychological research may enrich economic modeling, that drew my interest toward this relatively young but nevertheless fascinating area of economic research. So what, in a nutshell, is behavioral economics, also known as "economics and psychology"?

The core model used in orthodox economic theory posits that individuals make choices in order to maximize the expected value of a utility function, using and correctly processing all the available information. Moreover, individuals' preferences are assumed to depend only on own payoffs, to be unaffected by the framing of the decision, and to be time consistent. A growing body of evidence documents, however, that aspects of behavior deviate from the forecasts of standard theory both in laboratory experiments as well as actual market environments, thereby casting serious doubt on the underlying assumptions. For a very recent survey, see DellaVigna (2009). The research in behavioral economics suggests that individuals deviate from the core model in several respects and offers various modifications to the orthodox economic conception of human choice in order to make this model more realistic. Following the classification in Rabin (1999), the first important class of deviations is nonstandard preferences, the most prominent examples of which are reference dependence and social preferences. Roughly spoken, nonstandard preferences advocate a modification of the utility function that allows for an individual's utility to depend not only on the absolute level of the own final payoff but also on changes in outcomes relative to a reference level as well as the payoffs other individuals receive. The second class of deviations from the standard model, which considers biases in judgment under uncertainty, allows for individuals making errors when attempting to maximize their utility function, thereby presenting a more drastic challenge to the familiar economic setup. These inferential errors are acknowledged to come in various forms, with the tendency to infer too much from too little evidence as well as the propensity to misread contradicting evidence as confirming a previously held hypothesis being among the most widely known manifestations of such biases in judgment. An even more radical critique of the economics model is posed by the third class of deviations, which posits that people have tremendous difficulties in evaluating their own preferences and, in particular, in forecasting their own future preferences. These difficulties, which inter alia can be rooted in overprojection of current tastes on future tastes, different evaluation modes or only partial awareness of a taste for immediate gratification, can give rise to phenomena such as habit formation, preference reversals or self-control problems. While careful modeling and incorporation of these deviations into formal economic analysis definitely is a challenging and perhaps even daunting task, the reward seems promising: Over the past decade there have been numerous contributions to a wide variety of fields of economic interest, ranging from industrial organization over labor economics to finance, that adopted the behavioral approach sketched above and thereby helped to reconcile theoretical predictions with actually observed behavior in situations where the standard economics model was hard pressed in explaining the evidence in question. Thus, with one of the proclaimed goals of behavioral economics being the improvement of behavioral realism of economic models, it seems only appropriate that most of the ideas that led to the chapters of this dissertation popped up while I stumbled through life with my eyes open—perhaps a little bit wider open than before encountering Professor Matthew Rabin.

The first chapter of this dissertation, which is based on a joint paper with Fabian Herweg, inquires into the effects of the imposition of binding deadlines on the performance of an individual with time-inconsistent preferences. With time-inconsistency giving rise to self-control problems, this chapter draws on the third class of deviations from the standard model according to the above classification. The idea of thinking about the incentive effect of deadlines came to me while talking to one of my colleagues. It was shortly after the first semester of grim coursework in the Bonn Graduate School of Economics, that I found myself dodging work and instead talking to Daniel Engelage, whose office was next to mine. With all the doctoral students of my class being required to present a first research idea in the brown-bag seminar in the upcoming semester, and at that point being on the look-out for an idea myself, I asked Daniel whether he already knew what he would talk about, which he answered in the affirmative. Begrudging Daniel the opportunity to prepare his brown-bag presentation that much ahead of time, even before the detailed schedule of who was supposed to present when was announced, his answer to my next question left me puzzled. Asking him if he had already started preparing his presentation, he told me the following: "Well, I know I should start working on that presentation right away since I am not stressed right now, but I don't see myself doing that as long as I don't know when it is due." Seeing Daniel, a doctoral student of economics with a background in financial mathematics, whom I always had thought of as a well-organized and straightforward person, putting off an unpleasant task while at the same time being fully aware that the cost of completing the task was as low as it probably ever would be, and moreover making his decision when to start working on that task conditional on a deadline being imposed on him, intrigued me.

When looking for a starting point in order to inquire into the incentive effects of deadlines, my conversations with Fabian Herweg turned out to be very helpful for getting things straight in my miond. More importantly, Fabian finally became hooked up with the topic, which should be the beginning of a mutually inspiring cooperation for the years to come. Before long we realized that everyday life is virtually pervaded with situations in which people who are working on a task over a certain span of time find themselves faced with interim deadlines: next to a final exam, students taking a class for credit often have to hand in mandatory problem sets in order to pass; students writing a thesis meet in regular intervals with their thesis adviser to report on their progress; employees working on a long-term project have to submit memos at different stages of completion of a project; comic book artists, when working for a major comic publisher, have to turn around a five to six page strip in about three weeks when drawing a 24-pages comic book. Moreover, as the following quote of Tim Townsend, a comic book artist inking for major comic companies for over 17 years by now, suggests, people seem to value the presence of deadlines:

> *"It's always the deadlines. I think that every artist would like to be able to put as much time as they see fit into their work. Its tough to have to cut corners for the sake of deadlines. That's the nature of the beast though. I'm not complaining though. That would be like going to the beach and complaining about the heat."*
>
> — *Tim Townsend*

This observation seemed particularly puzzling to us because, through the lense of the standard economics model, a perfectly rational decision maker with time-consistent preferences would not welcome such restrictions on his choice set. Nevertheless, the somewhat subtle hint of the importance of deadlines being the prevention of an artist to get lost in his striving for perfectionism when the top priority is to finish the comic book on time, gave us a first sense that a possible reason for the ubiquity of deadlines might be rooted in people having problems to focus on priorities. What finally allowed us to call a spade a spade was the following statement made by Tim Seeley, another renowned artist in the comic industry who also authored the book "How to be a Comic Book Artist":

"The biggest pitfalls facing any comic book artist are video games and pornography. Seriously, most artists work from their homes and they can't structure their day without the atmosphere of an office. Video games are pretty much responsible for about 85% of late comics. I think 'Grand Theft Auto' alone nearly destroyed the industry."

— *Tim Seeley*

Thus, having identified self-control problems—like giving into the temptation of video gaming or watching porn instead of drawing the final pages of a comic book—as a possible candidate for the driving forces underlying the widespread phenomenon of deadlines, we delved into the literature of self-control problems and procrastination in order to prepare the stage for a more thorough investigation.

The origin of the standard economics model of intertemporal choice dates back to Paul Samuelson, who proposed the discounted utility model in 1937 as a generalized model of intertemporal choice that was applicable to multiple time periods. While Samuelson himself never made any claims on the model's descriptive validity, due to its high tractability the discounted utility model was almost instantly adopted by the economic profession and thereupon became the standard framework for analyzing intertemporal decision making. A key assumption of this model is that all psychological motivations possibly underlying intertemporal choice can be squeezed in a single parameter, a discount rate between any two periods which is independent of when the utility is evaluated. The assumption of constant discounting implies that a decision maker's intertemporal preferences are time consistent in the sense that a person's relative preference for well-being at an earlier over a later date is the same no matter when that person is asked to state this preference. Over the years, however, ample evidence has been compiled which gives testimony to various inadequacies of the discounted utility framework as a descriptive model of behavior, the best documented of which is hyperbolic discounting. The term hyperbolic discounting refers to a person discounting events in the near future at a higher discount rate than events in the distant future. With discount rates declining, while patient when evaluating outcomes in the distant future, people become increasingly impatient as the future draws near. Preferences with these features induce time inconsistency and thus are capable of capturing self-control problems. During a surge of interest among economists in the implications of hyperbolic discounting in a variety of contexts, in the late 1990s a tractable two-parameter model was established which allowed to capture the essence of hyperbolic discounting by only slightly modifying Samuelson's discounted utility model. This model of "quasi-hyperbolic discounting", which was originally introduced in Phelps and Pollak (1968) to study intergenerational altruism, was reapplied by David Laibson on the one hand, and Ted O'Donoghue and Matthew Rabin on the other hand, to rigorously study intra-personal decision making with self-control problems in the realms of consumption-saving decisions and procrastination. With the concept of quasi-hyperbolic discounting having become the framework of choice for analyzing time-inconsistent decision making in the applied economics literature over the past decade, we also apply this model to inquire into the effects of interim deadlines on performance and well-being of agents with self-control problems.

We develop a model of continuous effort choice over time that shifts the focus away from completion of to performance on a task, where the delayed reward for a task depends on the total effort devoted to that task. Besides showing that procrastination induced by time-inconsistent preferences in general hampers performance, we mainly ask two questions: First, does awareness of own self-control problems increase performance and overall well-being? Secondly, do interim deadlines enhance performance? With regard to the first question, we find that being aware of one's own self-control problems can reduce a person's performance as well as his overall well-being. This finding is in stark contrast to the extant literature on time-inconsistent procrastination, where sophistication was found to be rather boon than bane when costs are immediate and rewards are delayed. Regarding the second question, we show that being exposed to an interim deadline increases performance as well as overall well-being of a hyperbolic discounter, irrespectively of his awareness of own self-control problems. The forces at work have a clear intuition: an interim deadline, which limits the time available to work on a particular of several tasks, helps a time-inconsistent person combat procrastination on that task and thereby to achieve a more efficient allocation of effort over time. This more favorable allocation reduces effort cost, which in turn leads to the individual devoting overall more effort, which results in a better performance. Besides providing an explanation for recent empirical evidence in the psychological literature, these findings suggest that there is also scope for the employer of a time-inconsistent agent to benefit from imposing interim deadlines. Therefore, Chapter I provides a theoretical underpinning for the frequent observation of interim deadlines in working environments.

The second chapter of this thesis, which is based on joint work with Fabian Herweg and Philipp Weinschenk, enriches a principal-agent framework with moral hazard by incorporating reference-dependent preferences, i.e., a notion of nonstandard preferences in the sense of the first class of deviations from the standard model. We were inspired to inquire into the incentive effects of loss aversion when watching the motion picture "Dodgeball — Grab Life by the Balls". When asked whether he is just apathetic or if he really has no goal in life, Vince Vaughn, who plays the owner of a run-down gym in dire financial distress, answers:

"I found that if you have a goal you might not reach it. But if you don't have one, then you are never disappointed—and I gotta tell you it feels phenomenal."

— *Vince Vaughn*

Though a hilarious funny movie otherwise, there certainly is a grain of truth in that almost philosophical quote. People suffer from disappointment when they fail to reach goals that they wanted to achieve and, moreover, people may prefer not to try at all in order to avoid that disappointment. But then, if the prospect of not obtaining a desired reward may discourage an individual from trying to reap that reward, one question is immediately at hand: Why do so many incentive schemes put forth in the economic literature propose to reward an agent for good performance which is more likely—but not certain—to be observed the higher the agent's effort? The idea that performance-dependent incentive schemes might possibly backfire in this sense piqued our curiosity, and we decided to pursue this idea more rigorously.

One formal way proposed in the economic literature to capture the phenomenon of disappointment is loss aversion, first proposed by Kahneman and Tversky (1979) in their seminal contribution that introduced prospect theory into the economic discourse: when an outcome falls short compared to some reference point, people incur a loss which looms larger than a gain of equal size. With reference-dependent preferences being at the heart of loss aversion, each modeling approach that makes use of this concept faces the task of specifying how this reference point is determined. With no unifying approach provided how to determine a decision maker's reference point, however, the choice of the reference point basically is subject to the full arbitrariness of the modeler. While most of the literature that emerged in the wake of the original formulation of prospect theory assumed that an individual's reference point should be equated with the status quo, over time serious doubts have been raised about the adequacy of this approach. One criticism that has been propelled particularly forceful is that in many situations of economic interest, a person's reference point should not only be backward-looking, but also incorporate expectations about future events. A very recent concept of reference-dependent preferences and reference-point determination proposed by Botond Kőszegi and Matthew Rabin seems capable of accounting for both afore-mentioned difficulties. In this approach, the reference point is completely determined by a decision maker's rational expectations about outcomes and thus allows for the reference point to be forward-looking on the one hand, while at the same time removing modeling arbitrariness by deriving the reference point coherently from the economic environment. Beside this methodological advantages, we opted for this concept of reference-dependent preferences because our gut feeling regarding the statement of Vince Vaughn—that the fear of being disappointed might be demotivating—was directly challenged by a conjecture found in Kőszegi and Rabin (2006). Here it is argued that if agents are 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." So we set out to answer the the question who was right, Hollywood-based philosophy or economic intuition? The answer we should find—which, by the way, turned out to be the favorite answer of many economists: "It depends."—before long was pushed into the background of our research efforts when we suddenly found ourselves on the trail of a puzzle of by far greater economic interest.

In the classical principal-agent model with moral hazard, a principal seeks to contract with an agent who is supposed to exert costly effort which stochastically influences the principal's profits. With effort being costly and unobservable, in order to provide the agent with incentives to exert the desired level of effort, the principal has to make the agent's compensation payment depend on a verifiable performance measure which is imperfectly correlated with the agent's choice of action. Besides creating incentives, such a contractual arrangement also imposes some income risk on the agent since the performance measure is only a noisy signal of his effort choice. With most individuals being risk averse at least to some degree in the sense that they dislike being exposed to risky situations, this constitutes the classical tradeoff faced by the principal between providing incentives and optimal risk sharing. Orthodox economic theory models risk aversion by assuming a strictly concave utility function, which implies that an individual exhibits local risk neutrality. In consequence, as was shown, for example, in the pioneering contributions by Holmström (1979), the optimal compensation scheme for a risk averse agent in the afore-mentioned sense is based on the agent's performance in a fairly complicated way. In particular, with paying slightly different wages for different signals improving incentives at negligible costs due to local risk neutrality of the agent, the optimal performance-based compensation scheme was found to result in a fully differentiated wage scheme which rewards pieces of information that differ in their informative content about the agent's effort differently. This theoretically predicted complexity of contractual form, however, often is at odds with the form contracts take in observed practice. While already Prendergast (1999) referred to the discrepancy between theoretically predicted and actually observed contractual form about a decade ago, the same question was raised somewhat more pointedly by Bernard Salanié in his paper "Testing Contract Theory" from the year 2003:

"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*

With the standard notion of risk aversion implying a fully differentiated wage profile, a particular prevalent form of contractual arrangement observed in practice is hard to reconcile with standard economic theory: (lump-sum) bonus contracts. This contractual form pays out a base wage if the employees performance falls below a prespecified threshold, whereas for performances above this threshold the employee receives the base wage plus a fixed—and otherwise performance-independent—bonus payment.

As recently was demonstrated in Rabin (2000), the orthodox way of modeling risk preferences describes actually observed risk preferences only inadequately. With the tradeoff between incentive provision and risk sharing being at the core 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. Following the model of reference-dependent preferences according to Kőszegi and Rabin (2006, 2007), we assume that the agent is expectation-based loss averse, with his reference point being determined by his rational expectations: the agent compares his actual wage pairwise which each other wage that he could have received instead, where each comparison is weighted by the occurrence probability of the alternative outcome. Our main finding is that, no matter how rich the performance measure the principal can deploy, a simple (lump-sum) bonus scheme is optimal when loss aversion is the driving force of the agent's risk preferences. Though hard to grasp in only a few words, the basic intuition for this finding, which is in stark contrast to the predictions made by the standard economics model, is as follows: specification of many different wage payments induces uncertainty for the agent as to what he will receive. If he earns a relatively low wage, he compares this to higher wages he could have received, and experiences the sensation of a loss from this comparison. The anticipation of these losses reduces the agent's expected utility and thus he demands a higher average payment. While the principal has a classic rationale for rewarding signals strictly higher if they are stronger indicators of good performance, this negative "comparison effect" dominates this consideration if standard risk aversion plays a minor role. This makes a bonus contract, i.e., the most basic form an incentive contract can possibly take, the optimal contractual arrangement. In this sense, reference-dependent preferences according to Kőszegi and Rabin introduce an endogenous complexity cost into contracting based on psychological foundations.

After working on this project for about eighteen months and presenting it at about ten conferences or workshops, Fabian, Philipp, and I myself are very happy to have seen the paper "Binary Payment Schemes—Moral Hazard and Loss Aversion", on which this second chapter is based, being accepted for publication in the *American Economic Review*.

In the third chapter of this dissertation, I discuss a particular form of workplace design, job rotation, as a promising approach for organizations to cope with problems arising from their members being subject to confirmation bias. With confirmation bias being recognized as one of the most important forms of inferential errors, this third chapter incorporates nonstandard beliefs in the sense of the second class of deviations from the orthodox economic model into the context of organizational design. The idea for this third chapter presented itself shortly after accepting Professor Matthias Kräkel's offer to work as a scientific assistant at his Chair for Personnel and Organizational Economics, with my main field of activity being to act as supervisor for students writing their diploma, master or bachelor theses. Before long, I was contacted by a bachelor student, Meike Ahrends, who asked me if I was willing to be her thesis adviser. After finishing her bachelor degree at the University of Bonn, Meike planned to go for a master degree in business administration. With regard to her future applications for a slot in the master program of whatever university she fancied, we agreed on looking for a thesis topic with immediate relevance for business administration. Starting out from the survey on relational contracting in MacLeod (2007), and after taking a detour over dissolution of partnerships, we finally arrived at subjective performance evaluation and performance appraisal. Somehow, not really knowing why, I found myself going astray, visiting various internet platforms for topical debate, exchange of experience, and informed opinion on management, leadership and human resources issues. And then, without me being on the lookout for it, there it was again, behavioral economics all over the place, with raters' biases in performance evaluations seeming a particular "hot topic".

The practice of performance appraisal is an important cornerstone for many decisions made in organizations, in particular personnel decisions like promotion, training, and firing decisions. Performance appraisal itself, however, is a process by which humans judge other humans. With many performance measures regarding a firm's employees being subjective rather than objective in nature, the doors are wide open for behavioral biases and inferential errors to enter and—more importantly—to distort this process. According to Brian Davis, executive vice president of Personnel Decisions International, a Minneapolis-based consultancy firm, raters' bias in performance appraisal by now is considered a severe problem in organizational practice:

> "The problem with rater-bias is that it takes away the organization's ability to objectively use data from performance evaluations with any validity. [In consequence,] you can't count on the objectivity or accuracy of a performance assessment, and you have no differentiating data that allows you to make confident decisions about promotions, training or leadership development."
>
> — *Brian Davis*

The reason for organizations to worry about their decisions being undermined by raters' bias seems obvious: According to Right Management, a globally operating career transition and organizational consulting firm, the average cost of coping with an employee who does not work out is 2.5 times his salary, with lower employee morale, decreased productivity, and lost customer share being the most dire consequences of bad promotion decisions. So it hardly seems surprising that organizations strive to find a way for dealing with the problem of raters' bias, as for example stated by Rick Smith, Senior Vice President of Right Management:

> "There is a smaller margin for error today in selection and promoting people into key positions, and a greater need to target development efforts to ensure that they really make a difference."
>
> — *Rick Smith*

What puzzled me about the suggestions posted on the various internet platforms how to cope with raters' bias is that—at least to my eyes—they seemed somewhat half-heartedly. For example, a call for an objectification of the performance management process seems somewhat off-color because if there was a sufficiently rich set of objective performance measures, firms would not have to rely on subjective performance measures in the first-place. Likewise, awareness training, which informs supervisors about the types of subtle bias that can interfere with their performance as appraisers, surely is a reasonable first line of defense against raters' bias. But with raters' bias often taking the form of inferential errors which enter the human judgment process most often unknowingly and unwittingly, it seems questionable that awareness training can get the problem under control. So I started pondering whether organizational design provides some other, perhaps more reliable means to overcome the problems associated with raters' bias.

Raters' bias is documented to manifest itself in all sorts of ways. One of the most prominent notions is the so-called "horns-and-halo effect", which refers to supervisors' tendency to judge employees as either good or bad, thus putting on them an angelic halo in the first case and fiendish horns in the latter, and then to seek evidence that supports their opinion. In the psychological literature, the horns-and-halo effect is widely known as one of the many notions of a phenomenon called confirmation bias. Confirmation bias in general refers to unintentional and unknowing selectivity in the acquisition and use of information. Abundant empirical evidence supports the view that once one has come to believe in a position on an issue, one's primary purpose becomes that of justifying or defending that position. The conception that information is treated partially once an opinion has been established clearly represents a deviation from the assumption of the traditional economics model that, when faced with uncertainty, people correctly form their subjective probabilistic assessment according to the laws of probability. The first model to formalize this idea was proposed in Rabin and Schrag (1999). This model captures a manifestation of confirmation bias known as "primacy effect", which is intimately connected with the horns-and-halo effect. The primacy effect refers to the finding that, when information is gathered and integrated over time, evidence acquired in early stages is likely to carry more weight than evidence acquired later on. In consequence, opinions are formed early in the process and subsequently acquired information evaluated and interpreted in a way that is partial to that opinion: people tend to see ambiguous evidence more likely as supporting rather than disconfirming an established opinion, to question conflicting information more willingly than information supportive of preexisting beliefs, to explain away events that are inconsistent with a held position, and even to interpret evidence that should count against a hypothesis as counting in favor of it.

Applying the model by Rabin and Schrag (1999), I develop a simple framework which allows me to argue that organizational design provides a tool which is capable of thwarting confirmation bias: job rotation. In this model, with different types of jobs being available, the efficient allocation of a worker within an organization depends on his ability, which is assumed to be commonly unknown. If the firm wants to base this decision on a more solid informational footing by gathering additional information, it has to rely on its divisions' supervisors to do so. Supervisors are subject to confirmation bias, which implies that a supervisor's opinion about the worker's ability established early in the appraisal process will color all information this supervisor subsequently receives. Job rotation refers to a job practice which assigns an employee not to a single specific task but to a set of several tasks, associated with a meaningful change in job content, among which he rotates with some frequency. Thus, by implementing job rotation, the firm regularly breaks up the matches of supervisors and their subordinates, thereby creating multiple unbiased evaluations of many supervisors regarding one particular employee. While rotation of an employee between several divisions is acknowledged to cause costs by disrupting work flows or forgoing productivity gains of specialization, I show that preventing confirmation bias from affecting supervisors' judgment can indeed outweigh this cost, making job rotation the optimal form of job design. Next to advocating job rotation as a potential remedy for the horns-and-halo effect which harmonically blends in into many organizational forms in practice, these findings also complement the extant theory on job rotation by providing a rationale for this particular form of job design based upon psychological foundations.

Let me wrap up things by saying that I suggested to Meike Ahrends to write her thesis about Rabin and Schrag's (1999) model of confirmation bias and the phenomenon's implications in situations of economic relevance. On the one hand, being at least partially sophisticated about my own self-control problems, this presented a useful commitment device for myself, forcing me to read that paper in detail. On the other hand, it provided Meike with the opportunity to write her thesis based on a topic of importance for economics as well as business administration. She did extremely well and by now participates in the master program in business administration at the University of Mannheim.

The fourth chapter of this doctoral thesis, which is based on joint work with Fabian Herweg, examines the welfare effects of third-degree price discrimination in intermediate-good markets when the structure of the intermediate industry is not exogenously given but allows for costly entry. With all parties having standard preferences, processing information according to the laws of probability, and maximizing stable and correctly perceived preferences, this fourth chapter incorporates none of the behavioral deviations classified above but sticks to the orthodox economics model. So why, after delivering a pleading in the defense of economics and psychology, does this last chapter of my thesis build on the standard economics model? I don't think that there is a way to convincingly argue why my interest in behavioral economics recently should have abated—in particular, because it hasn't in the slightest. On the other hand, while perhaps lying dormant for some time, my interest in economic analysis based on the orthodox economics model has never abated either. Therefore, I was not really surprised seeing this interest rekindle when Fabian Herweg suggested to do some research in classical industrial organization, which—I guess—will always be where his heart belongs.

While looking for a starting point, a paper that caught our attention was by Inderst and Valetti (2009), who inquire into the effects of a credible threat of demand-side substitution in a model of third-degree price discrimination in input markets. In the introduction, Inderst and Valetti often refer to "the extant theory" or "the extant literature" on price discrimination in intermediate-good markets, without being very specific about what contributions constitute the extant literature. The insights gained by a somewhat more thorough literature research conducted by Fabian and myself were twofold: First, the extant theory on price discrimination in input markets turned out to be a manageable field, beginning in 1987 with the seminal contribution by Michael L. Katz. With the earlier papers by and large drawing the conclusion that allowing discriminatory wholesale pricing results in adverse welfare effects, most recently there has been a renewed surge of interest in the topic among both economists and practitioners. These more recent contributions in particular point to situations in which price discrimination may be welfare improving. Second, some evidently relevant questions have not yet been addressed in the extant literature, even though these questions by far are not far-fetched. For example, in all contributions on the topic so far, the upstream supplier of the input is assumed to be perfectly informed about the downstream firms' production technologies and associated efficiency in production, based upon which he subsequently charges possibly discriminatory wholesale prices. While we inquire into the implications of private information about production costs in another paper, in the paper upon which this fourth chapter is based, we relax the ubiquitous assumption of an exogenously given market structure of the downstream industry.

The common setup applied in the extant theory of third-degree price discrimination in input markets considers a monopolistic upstream firm supplying some input to firms in an intermediate industry which use this input to produce a final good subsequently sold to final consumers. An assumption shared by all those contributions is that the structure of the intermediate industry is exogenously given. As was already noted by Katz in his chapter on vertical restraints in the first volume of the Handbook of Industrial Organization,

"[t]here are several ways in which the manufacturer may influence the number of retailers. The number of dealers may be chosen directly by the manufacturer [...]. Alternatively, the manufacturer may indirectly control the number of dealers through his pricing policy or the nonterritorial resale restraint provisions of his contract offer."

– Michael L. Katz (1989)

Thus, abstracting from entry into the intermediate industry ignores the fact that pricing decisions of the upstream supplier are a major determinant of the resulting industry structure and market outcome. With these pricing decisions in turn being determined by the pricing instruments available to the upstream supplier, inquiring into the implications of banning discriminatory wholesale pricing for structure of the intermediate industry and market outcome seems a natural question in this context.

In this chapter, we present a formal model of third-degree price discrimination in input markets which allows for costly entry into the downstream industry. First we consider the case where downstream firms operate in separate markets, which allows us to isolate a first important effect: Price discrimination is found to foster entry in the sense that whenever entry occurs under uniform pricing, entry also occurs under a discriminatory pricing regime. Intuitively, if entry is very costly, the maximum wholesale price the upstream supplier can possibly charge to the entrant such that entry occurs becomes low. Under a discriminatory pricing regime the wholesale price charged to the incumbent firm, and thus upstream profits from the incumbent downstream market are unaffected by this constraint possibly imposed by the entry fee. Under uniform pricing, on the other hand, the upstream supplier has to pass this low wholesale price on to the incumbent firm as well, thereby reducing profits from the incumbent market. Therefore, in contrast to a discriminatory pricing regime, under a uniform pricing regime the input supplier may be willing to forgo profits from opening a new downstream market in order to maintain high profits from the incumbent downstream market. Whenever entry occurs under price discrimination but not under uniform pricing, social welfare is strictly higher under the discriminatory pricing regime because opening of a new market increases consumers' surplus and upstream profits, and also gives rise to non-negative profits for the entering downstream firm. We then consider the case where downstream firms operate in the same market and entry into the intermediate industry leads to downstream competition. While the entry-promoting effect of price discrimination is operative also in the case of downstream competition, entry occurring under a discriminatory but not under a uniform pricing regime no longer is a sufficient condition for an increase in welfare. The reason is that, if entry into the intermediate industry occurs under price discrimination but not under uniform pricing, a part of the output produced by the monopolistic incumbent under uniform pricing is shifted to the less efficient entrant under a discriminatory pricing regime due to competition. While this shift in production shares benefits the upstream supplier and consumers because the associated increase in aggregate quantity attenuates the double-marginalization problem with a downstream monopolist, it may not be sufficient to make society better off if the cost of entry or the increase in the cost of production is high. Thus, in the light of our findings which identify the welfare effects of third-degree price discrimination in input market as highly ambiguous, calls for a determinate legislature which either bans or permits discriminatory pricing practices seem inadequate.

Having reached the end of this introspective retrospect, there is nothing left to do except for wishing you an enjoyable read.

I. 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 advisor 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-hyperbollically, 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 aquisition.

[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

[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.

agents and discuss the impact of awareness on performance and overall satisfaction. 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 the appendix.

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. This delayed reward, which is assumed to be a function of total effort invested, is denoted by $g(\sum_{t=1}^{2} e_t)$.[8]

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

$$c'(x) > 0, \qquad c''(x) > 0, \qquad c(0) = 0, \qquad c'(0) = 0$$
$$g'(x) > 0, \qquad g''(x) < 0, \qquad g(0) = 0, \qquad g'(0) > 0$$

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 $U_1^i = -c(e_1) - \beta c(e_2) + \beta g\left(e_1 + e_2\right)$. Accordingly, given first-period effort $\hat{e}_1$, the second-period intertemporal utility takes the form $U_2^i = -c(e_2) + \beta g\left(\hat{e}_1 + e_2\right)$. 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

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

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 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

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

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

[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).

shifting effort from the high-effort period to the low-effort period, thereby reducing total effort costs.

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 I.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:

$$g'(e_1^N + e_2^{TC}) = c'(e_2^{TC}) \tag{I.2}$$

$$\beta g'(e_1^N + e_2^{TC}) = c'(e_1^N)\ . \tag{I.3}$$

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 caracterizes the second-period effort, e_2^N, is given by

$$\beta g'(e_1^N + e_2^N) = c'(e_2^N)\ . \tag{I.4}$$

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

Proposition I.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 I.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

[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.

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.

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 I.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,

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

Differentiating (I.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

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

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

[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).

Proposition I.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 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 I.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 (fictitiuos) 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 I.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

[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).

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. 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

[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 nenu 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.

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. 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 shortcommings. 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 sohisticate 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

[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.

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

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 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

[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".

[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-biasedness, $\beta = 1/4$, 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.

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 I.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 I.2 states a very intuitive result: given the pessimism effect outweighs the incentive effect, then sophisticates choose higher first-period efforts than naifs.

Proposition I.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]

[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.

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 particular 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

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

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

[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.

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 spent on a specific task. 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, $e_2^{S^{ND}}(\hat{e}_1)$, is characterized by

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

The effort level choosen by a sophisticate in the first period, $e_1^{S^{ND}}$, is determined by the following first-order condition,[31]

$$\begin{aligned}\beta g'((1/2)(e_1^{S^{ND}} + e_2^{S^{ND}}(e_1^{S^{ND}}))) - c'(e_1^{S^{ND}}) \\ + \frac{de_2^{S^{ND}}(e_1)}{de_1}\beta\left[g'((1/2)(e_1^{S^{ND}} + e_2^{S^{ND}}(e_1^{S^{ND}}))) - c'(e_2^{S^{ND}}(e_1^{S^{ND}}))\right] = 0\ .\end{aligned} \tag{I.9}$$

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. $e_1^{S^{ND}} < e_2^{S^{ND}}$ and $e_A = e_B = (1/2)(e_1^{S^{ND}} + e_2^{S^{ND}})$, where $e_2^{S^{ND}} = e_2^{S^{ND}}(e_1^{S^{ND}})$.

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, $e_{A2} = 0$, $e_A = e_{A1}$

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

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

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

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

$$c'(e_{B2}^{S^D}(\hat{e}_{B1})) = \beta g'(\hat{e}_{B1} + e_{B2}^{S^D}(\hat{e}_{B1})) . \tag{I.10}$$

Differentiation of (I.10) yields

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

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,

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

This utility maximization problem, however, does not have an interior solution.[32] When facing a deadline, a sophisticated agent considers it optimal to work exclusively on task A in the first period, that is, $e_{B1}^{S^D} = 0$. 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 plausible, the formal proof of this statement is somewhat elaborate and therefore deferred to the appendix. The effort levels which are chosen strictly positive, $e_A^{S^D}$ and $e_{B2}^{S^D}$, are characterized as follows:

$$c'(e_A^{S^D}) = \beta g'(e_A^{S^D}) \tag{I.11}$$

$$c'(e_{B2}^{S^D}) = \beta g'(e_{B2}^{S^D}) \tag{I.12}$$

From (I.11) and (I.12) it follows immediately that $e_A^{S^D} = e_{B2}^{S^D}$. 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^{S^D} denote the effort level that is chosen under a regime of deadlines in each period and per task. Formally we have $e_1^{S^D} = e_A^{S^D} = e^{S^D}$ and $e_B^{S^D} = e_2^{S^D} = e^{S^D}$.

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.

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

Proposition I.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 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.[33]

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 I.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,*

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

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, $e^{SD} = e^{ND}$.[34] 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 undetermided which type of hyperbolic discounter benefits more from 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.[35] 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.[36] 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, particpants 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

[34] 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 I.5.

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

[36] 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.

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, performace 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.[37] 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, $\Delta U_1^S < 0 < \Delta U_0^S$.[38] 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

[37] 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.

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

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 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, finishig 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 B, 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.

II. Binary Payment Schemes: Moral Hazard and Loss Aversion

In this chapter, we extend the principal-agent model with moral hazard by assuming that the agent is expectation-based loss averse according to Kőszegi and Rabin (2006, 2007). The optimal contract is a binary payment scheme even for a rich performance measure, where standard preferences predict a fully contingent contract. The logic is that, due to the stochastic reference point, increasing the number of different wages reduces the agent's expected utility without providing strong additional incentives. Moreover, for diminutive occurrence probabilities for all signals the agent is rewarded with the fixed bonus if his performance exceeds a certain threshold.

1. Introduction

A lump-sum bonus contract, with the bonus being a payment for achieving a certain level of performance, is probably one of the most simple incentive schemes for employees one can think of. According to Thomas J. Steenburgh (2008), salesforce compensation plans provide incentives mainly via a lump-sum bonus for meeting or exceeding the annual sales quota.[1] The observed plainness of contractual arrangements, however, is at odds with predictions made by economic theory, as nicely stated in the above quote by Bernard Salanié (2003). While Canice 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 Edward P. Lazear and Oyer (2007), and the answer still is not fully understood.

Beside this gap between theoretical prediction and observed practice, both theoretical and 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. This observation raises "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, 16).

To give one possible explanation for the widespread use of binary payment schemes, we modify the principal-agent model with moral hazard by assuming that the agent is expectation-based loss averse according to Botond Kőszegi and Matthew Rabin (2006, 2007).[2] 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. Following Kőszegi and Rabin, we posit that the agent—next to standard consumption utility—derives gain-loss utility from comparing the actual outcome with his lagged expectations. Specifically, the agent compares his actual

[1] Incentives for salespeople in the food manufacturing industry often are solely created by a lump-sum bonus, see Paul Oyer (2000). Moreover, in his book about designing effective sales compensation plans, John K. Moynahan (1980) argues that for a wide range of industries lump-sum bonus contracts are optimal. For a survey on salesforce compensation plans, see Kissan Joseph and Manohar U. Kalwani (1998). Simple binary contracts are commonly found not only in labor contexts, but also in insurance markets, where straight-deductible contracts are prevalent.

[2] We will use the terms bonus contract, bonus scheme, and binary payment scheme interchangeably to refer to a contract that specifies exactly two distinct wage payments, a base wage and a lump-sum bonus.

wage pairwise with each other wage that he could have received instead, where each comparison is weighted by the occurrence probability of the alternative outcome. 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. We derive this finding in a model where the principal can make use of a rich performance measure and where the standard notion of risk aversion would predict fully contingent contracts. Intuitively, specifying many different payments induces uncertainty for the agent as to what he will receive. If he earns a relatively low wage, he compares this to higher wages he could have received, and experiences the sensation of a loss from this comparison. The anticipation of these losses reduces the agent's expected utility and thus he demands a higher average payment. While the principal has a classic rationale for rewarding signals strictly higher if they are stronger indicators of good performance, the negative "comparison effect" dominates this consideration if standard risk aversion plays a minor role.[3] In this sense, reference-dependent preferences according to Kőszegi and Rabin introduce an endogenous complexity cost into contracting based on psychological foundations.

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 is rewarded with the bonus payment contains either only good signals, or all good signals and possibly a few bad signals as well.[4] By paying the bonus very often or very rarely, the principal can minimize the weight the agent puts ex ante on the disappointing event of feeling a loss when not obtaining the bonus. When abstracting from integer-programming problems, it is optimal for the principal to order the signals according to their relative informativeness (likelihood ratio). Put differently, the agent receives the bonus for all signals that are more indicative of high effort than a cutoff signal, e.g., a salesperson receiving a bonus for meeting or exceeding the annual sales quota.

In addition, we show that an increase in the agent's degree of loss aversion may allow the principal to use a lower-powered incentive scheme. 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 scope for incurring a loss. The overall cost of implementation, however, increases in the agent's degree of loss aversion.

[3]The term "comparison effect" was first introduced by Kőszegi and Rabin (2006).

[4]The theoretical prediction that inferior performance may also well be rewarded with a bonus is in line with both Joseph and Kalwani's (1998) 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 Gilbert A. Churchill, Neil M. Ford and Orville C. Walker's (1993) prescription that bonuses should be based on objectives that can be achieved with reasonable rather than Herculean efforts.

While assuming for most of the paper that the agent is not too loss averse, which guarantees that the first-order approach is valid, we also briefly investigate the principal's problem for higher degrees of loss aversion. Here, to keep the analysis tractable, we focus on binary measures of performance. We show that 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 performance measure. Moreover, for high degrees of loss aversion, stochastic ignorance of the performance measure also lowers the cost of implementing the desired level of effort. The logic of this result is that stochastic ignorance allows the principal to pay the bonus to the agent even if she observes the signal that indicates low effort. By doing this, the agent considers it ex ante less likely that he will be disappointed at the end of the day, and thus he demands a lower average payment. In this case, with the optimal contract including randomization which would not be optimal under the standard notion of risk aversion, loss aversion leads to more complex contracts than predicted by orthodox theory.

Before launching out into the model description, we briefly review the existing evidence documenting that expectations matter in the determination of the reference point, which is a key feature of the Kőszegi-Rabin concept.[5] While mainly based on findings in the psychological literature,[6] evidence for this assumption is provided also by some recent contributions to the economic literature. Investigating decision making in a large-stake game show, Thierry Post et al. (2008, 62) come to the conclusion that observed behavior is "consistent with the idea that the reference point is based on expectations." Alike, analyzing field data, Vincent P. Crawford and Juanjuan Meng (2009) propose a model of cabdrivers' labor supply that builds on the Kőszegi-Rabin theory of reference-dependent preferences. Their estimates suggest that a reference-dependent model of drivers' labor supply where targets are carefully modeled significantly improves on the neoclassic model. In a real-effort experiment, Johannes Abeler et al. (forthcoming) manipulate the rational expectations of subjects. They find that effort provision is significantly different between treatments in the way predicted by models

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

[6]For instance, Barbara Mellers, Alan Schwartz and Ilana Ritov (1999) and Hans C. Breiter et al. (2001) document that both the actual outcome and unattained possible outcomes affect subjects' satisfaction with their payoff.

of expectation-based loss aversion.

In the following Section II, we formulate the principal-agent relationship. Section III specifies the principal's problem and derives the set of feasible contracts. In Section IV, the principal's problem is solved and properties of the optimal contract are discussed. Section V investigates the implications of high degrees of loss aversion. In Section VI, next to the related literature, alternative notions of loss aversion are discussed. Section VII concludes. All proofs as well as additional technical discussions are relegated to the Appendix.

2. The Model

A principal offers a one-period employment contract to an agent, who has an outside employment opportunity yielding expected utility $\bar{u}$.[7] 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.[8] 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 $\hat{a} \in (0, 1)$.

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. 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 $\gamma_s(a) \equiv a\gamma_s^H + (1 - a)\gamma_s^L$.[9] For technical convenience, we make the following assumption.

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

[8] 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.

[9] The results of Section IV do not rely on the linear structure of the performance measure. The linearity is needed to show validity of the first-order approach and in Section V.

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).*

Part (i) guarantees that any signal s is either a good or a bad signal, i.e., 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 part (iii) signals can unambiguously be ranked according to the relative impact of an increase in effort on the probability of observing a particular signal.

The contract which the principal offers to the agent consists of a payment for each realization of the performance measure, $(w_s)_{s=1}^S \in \mathbb{R}^S$.[10]

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 Amos Tversky and Daniel Kahneman (1991) on their "value function". In our model, the agent's consumption space comprises of two dimensions, money income ($x_1 = W$) and effort ($x_2 = a$).[11] 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 $\lim_{a \to 1} c(a) =$

[10] Restricting the principal to offer nonstochastic wage payments is standard in the principal-agent literature and also in accordance with observed practice. In Section V, we comment on this assumption.

[11] 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.

∞. We assume that the gain-loss function is piece-wise linear,[12]

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

The parameter $\lambda \geq 1$ characterizes the weight put on losses relative to gains.[13] The weight 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.

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 expected utility is obtained by averaging over all these comparisons. 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. The agent's utility from this particular

[12] In their work on asset pricing, Nicholas Barberis, Ming Huang, and Tano Santos (2001) argue that for prospects involving both gains and losses, loss aversion at the kink is more relevant than the degree of curvature away from the kink. Implications of a more general gain-loss function are discussed in Section VII.

[13] Alternatively, one could introduce a weight attached to gain-loss utility relative to intrinsic utility, $\eta \geq 0$. We implicitly normalized $\eta = 1$ which can be done without much loss, since this normalization does not qualitatively affect any of our results.

outcome is given by

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

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 contract and 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. Preliminary Analysis

Let $h(\cdot) := u^{-1}(\cdot)$, i.e., 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 Sanford J. Grossman and Oliver D. 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

$$EU(a) = \sum_{s \in \mathcal{S}} \gamma_s(a) u_s - (\lambda - 1) \sum_{s \in \mathcal{S}} \sum_{\{\tau | u_\tau > u_s\}} \gamma_\tau(a)\gamma_s(a)(u_\tau - u_s) - c(a). \qquad \text{(II.1)}$$

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 that $(\lambda - 1)$ can also be interpreted as 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 is

$$EU'(a) = \sum_{s\in\mathcal{S}}(\gamma_s^H - \gamma_s^L)u_s \\ - (\lambda - 1)\sum_{s\in\mathcal{S}}\sum_{\{\tau|u_\tau>u_s\}}[\gamma_\tau(a)(\gamma_s^H - \gamma_s^L) + \gamma_s(a)(\gamma_\tau^H - \gamma_\tau^L)](u_\tau - u_s) - c'(a). \quad \text{(II.2)}$$

A principal who wants to implement action $\hat{a} \in (0, 1)$ minimizes her expected wage payment subject to the usual individual rationality and incentive compatibility constraints:

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

$$\text{subject to} \quad EU(\hat{a}) \geq \bar{u}\ , \quad \text{(IR)}$$

$$\hat{a} \in \arg\max_{a\in\mathcal{A}} EU(a)\ . \quad \text{(IC)}$$

Suppose the agent's action choice is contractible, i.e., the incentive constraint (IC) is absent. In this first-best situation, the principal pays a risk- or loss-averse agent a fixed wage $u^{FB} = \bar{u} + c(\hat{a})$. 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 fully to 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 $\hat{a}$. Second, the first-order approach is valid, i.e., the incentive constraint to implement action $\hat{a}$ can equivalently be represented as $EU'(\hat{a}) = 0$. 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 $\lambda > 2$. 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 incurring 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 his 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. In Section V, 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 additional assumption on the agent's cost function.

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.[14] In fact, our results only require the validity of the first-order approach, not that Assumption (A3) holds. In Section V, we consider the case in which the first-order approach is invalid.

Lemma II.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. Existence of a second-best optimal contract is shown separately for the three cases analyzed: 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 $u_{s-1} \leq u_s$ 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 $b_\tau = u_\tau - u_{\tau-1} \geq 0$. Let $\boldsymbol{b} = (b_2,\dots,b_S)$. With this notation the individual rationality constraint can be stated as follows:

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

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

[14] The validity of the first-order approach under assumptions (A1)-(A3) is proven in Appendix B. The reader should be aware that the proof requires notation introduced later in this section.

$\boldsymbol{b}'\boldsymbol{\rho}(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})$. 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 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".[15] A crucial point is that the loss premium increases in the contract's degree of wage differentiation. 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 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.[16]

Given that 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 agent's action balances the tradeoff between maximizing intrinsic utility and minimizing the

[15] 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 $[(\lambda - 1)/2]D(\boldsymbol{u}) = \boldsymbol{b}'\boldsymbol{\rho}(\hat{\gamma}, \lambda, \hat{a})$.

[16] 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.

expected net loss. Overall, 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. It is obvious that (IC′) can only be satisfied if there exists at least one $\beta_s(\cdot) > 0$. 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:

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

4. The Optimal Contract

In this part of the paper, we first review the standard case where the agent is only risk averse but not loss averse. Thereafter, the case of a loss-averse agent with a risk-neutral intrinsic utility function is analyzed. Last, we discuss the intermediate case of a risk- and loss-averse agent.

4.1. Pure Risk Aversion

Consider an agent who is risk averse in the usual sense, $h''(\cdot) > 0$, but does not exhibit loss aversion, $\lambda = 1$.

Proposition II.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 II.1 restates the well-known finding by Bengt 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 contingent, with each signal being rewarded differently.

4.2. Pure Loss 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, $\lambda > 1$. 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 aforementioned sense. Nonemptiness of this set follows from Lemma II.1.

The Optimality of Bonus Contracts.—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. 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:

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

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

Intuitively, the principal seeks to minimize the agent's expected net loss. 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 case with a concave intrinsic utility function. Even though the principal's objective will not reduce to minimizing the agent's loss premium alone, this nevertheless remains an important aspect of her problem. Since the solution to Program ML tells us how to minimize the 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. 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; this problem clearly has a solution.

Proposition II.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 II.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 the contract specifies only two different wage payments no matter how rich the signal space. This endeavor to reduce the complexity of the contract is plausible since a high degree of wage differentiation increases the loss premium. A loss-averse agent considers a wage schedule as riskier if the average margin between any two wages is higher. The principal can reduce the riskiness of the contract by setting the spread of as many wage pairs as possible equal to zero.

More intuitively, what are the effects of the principal specifying many different wage payments? With a contract specifying many different wages, receiving a relatively low wage feels like a loss when comparing it to possible higher ones, which in turn decreases the agent's utility. Likewise, in the case of obtaining a high wage most comparisons are drawn to lower wages, with the associated gains increasing the agent's utility. Since losses loom larger than gains, anticipating these comparisons ex ante reduces the agent's expected utility and thus a higher average payment is needed to make him accept the contract. In order to avoid these unfavorable comparisons, the principal has an incentive to lump together wages for different signals.

With effort being unobservable but costly for the agent, however, any incentive-compatible contract has to display at least some degree of wage differentiation. Under the standard notion of risk aversion, creating incentives via increasing the utility margin between two signal realizations becomes more and more costly due to the agent's marginal utility of money being decreasing. Loosely speaking, instead of creating incentives via one big bonus payment, provision of incentives is achieved at lower cost by setting many small wage spreads. When facing a purely loss-averse agent, whose marginal intrinsic utility of money is constant, the principal cannot capitalize on differentiating payments according to performance. In this case, pooling together as many wages as possible is beneficial to the principal and thus the optimal contract is a binary payment scheme.

Features of the Optimal Contract.—Up to now 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 constraints. The problem of choosing the optimal partition of signals, $\mathcal{B}^*$, is an integer programming problem. As is typical for this class of problems, and as is nicely illustrated by the well-known "Knapsack Problem", it is impossible to provide a general characterization of the solution.[17]

Next to these standard intricacies of integer programming, there is an additional difficulty in our model: the principal's objective behaves nonmonotonically when including an additional signal into the "bonus set" $\mathcal{B}$. From Program ML it follows that, for a given bonus set $\mathcal{B}$, the minimum cost of implementing action $\hat{a}$ is

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

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

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

This objective function illustrates the tradeoff that the principal faces. 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 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. Intuitively, 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. These two goals may conflict with each other. Nevertheless, it can be shown that the optimal contract displays the following plausible property.

Proposition II.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.

[17] The 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.

Back to the 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. 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.[18] With this assumption, we can show that it is optimal to order the signals according to their likelihood ratios.

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

If one is prepared to assume that higher sales revenues are associated with higher likelihood ratios, then Proposition II.4 states that the sales agent receives the bonus only if his sales exceed a previously specified sales quota.

Comparative Statics.—Last, we want to point out the following comparative static results.

Proposition II.5: *(i) The minimum cost of implementing action $\hat{a}$ strictly increases in λ. (ii) For a given feasible bonus set $\mathcal{B}$, the wage spread necessary to implement action $\hat{a}$ decreases in λ if and only if $P_{\mathcal{B}} > 1/2$.*

Part (ii) of Proposition II.5 relates to the reasoning by Kőszegi and Rabin (2006, 1156) that if the agent is expectation-based loss averse, 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." The agent's expected utility 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 there is a second component since in expectation the agent suffers from deviations from his reference point. A deviation from the agent's reference point occurs with probability $P_{\mathcal{B}^*}(1 - P_{\mathcal{B}^*})$, which we refer to 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 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, leads 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 II.5, the overall cost of implementation strictly increases in the agent's degree of loss aversion.

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

4.3. The General Case: Loss Aversion and Risk Aversion

While binary wage schemes based on a rich signal space are hard to reconcile with the orthodox notion of risk aversion, it is well-known that bonus contracts may be optimal if both contracting parties are risk (and loss) neutral. This finding, however, immediately collapses when the agent is somewhat risk averse.[19] As we argue in this section, our finding that under loss aversion the optimal contractual arrangement takes the form of a bonus scheme is robust towards introducing a slightly concave intrinsic utility function. The intuition is as follows: with the intrinsic utility function for money being concave the principal has a classic rationale for rewarding signals that are stronger indicators of good performance strictly higher. Due to the agent being loss averse, however, the principal still has an incentive to lump together wages in order to eliminate the negative comparison effect. If risk aversion is relatively unimportant compared to loss aversion, then this motive outweighs the principal's benefit from differentiating payments according to performance, and the optimal contract is a binary payment scheme. More formal, when the agent's intrinsic utility function becomes close to linearity the risk premium goes to zero, whereas due to loss aversion there are still first-order costs of wage differentiation. While we provide a more thorough discussion as well as a formal proof of this intuition in Appendix C, at this point we content ourselves by illustrating this conjecture by means of an example.

Suppose $h(u) = u^r$, with $r > 1$. More precisely, $R = 1 - \frac{1}{r}$ denotes the Arrow-Pratt measure for relative risk aversion of the intrinsic utility function. The agent's effort cost is $c(a) = (1/2)a^2$, the effort level to be implemented is $\hat{a} = 1/2$, and the reservation utility $\bar{u} = 10$. Assume that the agent's performance can take only three values, excellent (E), satisfactory (S) or inadequate (I). Let

$$\gamma_E^H = 5/10 \qquad \gamma_S^H = 4/10 \qquad \gamma_I^H = 1/10$$
$$\gamma_E^L = 1/10 \qquad \gamma_S^L = 3/10 \qquad \gamma_I^L = 6/10.$$

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 $u_3 = u_E$. The structure of the optimal contract for this specification and various values of r and λ is presented in Table 1. Table 1 suggests that the optimal contract typically involves pooling of the two good

[19] With both contracting parties being risk neutral, a broad range of contracts—including bonus schemes—is optimal. If the agent is protected by limited liability, Eun-Soo Park (1995), Son Ku Kim (1997), Oyer (2000), and Dominique Demougin and Claude Fluet (1998) show that the unique optimal contract is a bonus scheme. As demonstrated by Ian Jewitt, Ohad Kadan, and Jeroen M. Swinkels (2008), these findings break down if risk aversion is introduced even to the slightest degree.

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

signals, in particular when the agent's intrinsic utility is not too concave. Table 1 nicely illustrates the tradeoff 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.[20]

5. Implementation Problems and Stochastic Contracts

In order to explore the implications of a higher degree of loss aversion, we relax assumption (A2), which implies that the first-order approach is not necessarily valid. To ease the exposition, we consider a purely loss-averse agent and restrict attention to binary measures of performance, i.e., $\mathcal{S} = \{1, 2\}$. For notational convenience , let γ^H and γ^L denote the probabilities of observing signal $s = 2$ conditional on B being realized and not being realized, respectively.[21] 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 assume that $s = 2$ is the good signal.

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 u if the bad signal is observed, and he is paid the base wage plus a bonus b if the good signal is observed. For now assume that $b \geq 0$.[22] We assume that the agent's intrinsic disutility of effort is a quadratic function, $c(a) = (k/2)a^2$.[23] The first derivative of the agent's expected utility with respect to effort is given by

$$EU'(a) = \underbrace{(\gamma^H - \gamma^L)b\,[2 - \lambda + 2\gamma(a)(\lambda - 1)]}_{MB(a)} - \underbrace{ka}_{MC(a)} . \tag{II.6}$$

[20] For a given r, the degree of pooling actually may decrease in λ. This can happen, however, only locally: at some point, the degree of pooling increases in λ.

[21] 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 $\gamma_2^L = \gamma^L$.

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

[23] Allowing for more general effort cost functions does not qualitatively change the insights that are to be obtained.

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

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

Implementation problems in our sense refer to a situation where there are actions $a \in (0, 1)$ that are not incentive compatible for any bonus payment.

Proposition II.6: *Suppose (A4) holds, then effort level* $\hat{a} \in (0, 1)$ *is implementable if and only if* $a_0 \leq 0$.

Obviously, implementation problems do not arise when (A2) is satisfied. Implementation problems do occur, however, 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 are highly informative. These implementation problems arise because the agent has two possibly conflicting targets: on the one hand, he seeks to maximize net intrinsic utility, $u + b\gamma(a) - (k/2)a^2$, while on the other hand, he wants to minimize the expected loss by choosing an action such that the loss probability, $\gamma(a)(1 - \gamma(a))$, becomes small. For $\gamma^L \geq 1/2$ these targets are perfectly aligned: the loss probability is strictly decreasing in the agent's action, which implies that an increase in the bonus unambiguously increases effort and thus each action $a \in (0, 1)$ is implementable. For $\gamma^L < 1/2$, however, implementation problems do arise when λ is sufficiently large. Roughly speaking, being very loss averse, the agent primarily cares about reducing the loss probability. With the loss probability being inverted U-shaped in this case, the agent achieves this by choosing one of the two extreme actions $a \in \{0, 1\}$, i.e., the principal faces severe implementation problems.

Turning a Blind Eye.—One might wonder if there is a remedy for these implementation problems. The answer is "yes". 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 $p \in [0, 1)$. Thus, the overall probability of receiving the bonus is given by $\gamma(p, a) \equiv p + (1 - p)\gamma(a)$. This strategic ignorance of information gives rise to a transformed performance measure with $\gamma^H(p) = p + (1 - p)\gamma^H$ and $\gamma^L(p) = p + (1 - p)\gamma^L$ denoting the probabilities that the bonus is paid to the agent conditional on benefit B being realized and not being realized, respectively. We refer to the principal not paying attention to the performance measure as turning a blind eye. 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. In the light of Proposition II.6, this immediately implies that the principal can eliminate any implementation problems by choosing p sufficiently high.

Besides alleviating possible implementation problems, turning a blind eye can also benefit the principal from a cost perspective. Differentiating the minimum cost of implementing action $\hat{a}$ under the transformed performance measure,

$$C(p;\hat{a}) = \overline{u} + \frac{k}{2}\hat{a}^2 + \frac{k\hat{a}(\lambda-1)(1-\gamma(\hat{a}))}{(\gamma^H - \gamma^L)} \frac{\gamma(\hat{a}) + p(1-\gamma(\hat{a}))}{1 - (\lambda - 1)\left[1 - 2\gamma(\hat{a}) - 2p(1-\gamma(\hat{a}))\right]}, \quad \text{(II.7)}$$

with respect to p reveals that $\text{sign}\{dC(p;\hat{a})/dp\} = \text{sign}\{2 - \lambda\}$. Hence, an increase in the probability of ignoring the performance measure 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.[24]

Proposition II.7: *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 $\lambda > 2$.*

To grasp this finding intuitively, remember the intuition underlying Proposition II.2: by implementation of a bonus contract, the principal reduces the ex ante probability of the agent incurring a loss by making it more likely that the agent receives what he expects to receive. By the same token, turning a blind eye allows the principal to reduce the agent's loss premium even beyond what is achieved by a deterministic bonus contract. While this reduction comes at the cost of making the performance measure less informative, according to Proposition II.7, the positive effect on the agent's loss premium outweighs the negative effect on incentives if the agent is sufficiently loss averse.

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

[24]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 if there is a cost of ignorance, however, the optimal probability of turning a blind eye is well defined.

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).[25] In this sense, while the optimal contract under standard risk aversion would specify only two distinct wages, loss aversion increases the complexity of the optimal contract.

We conclude this section by pointing out an interesting implication of the above analysis. Suppose the principal has no access to a randomization device. 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 II.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. 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 framework by Frøystein Gjesdal (1982).[26] Our findings, however, show that the latter is not the case anymore when the agent is loss averse.

6. Alternative Notions of Loss Aversion and Related Literature

With only little being known about how exactly expectations enter into the formation of a person's reference point, a discussion seems warranted to what extent our results depend on the notion of loss aversion according to Kőszegi and Rabin (2007). The agent in our model compares an obtained outcome with all other possible outcomes. This pairwise comparison, which may lead to one and the same outcome being perceived as both a gain and a loss at the same time, is in fact responsible for our main findings.[27] An increase in the margins of payments always increases the agent's expected loss. Even though they are closely related to the CPE concept, this latter effect does not arise under the forward-looking notions of loss aversion according to Bell (1985), Loomes

[25] The finding that stochastic contracts may be optimal is not novel to the principal-agent literature. Hans 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, Roland Strausz (2006) finds that deterministic contracts may be suboptimal in a screening context.

[26] The sufficiency part of Blackwell's theorem states that making use of more informative performance measure implies that the principal is not worse off. See David Blackwell (1951, 1953).

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

and Sugden (1986), or Gul (1991), which do not allow for mixed feelings. For the sake of argument, consider an agent with linear intrinsic utility for money who is loss averse in the sense of Bell (1985), i.e., his reference point is the arithmetic mean of the wage distribution. Suppose there are only three signals, $s = 1, 2, 3$, which are equiprobable for the action which is to be implemented. The associated wages are $w_1 < w_2 = w_3 =: w_{23}$. If the principal increases the wage for signal 3 by $\varepsilon > 0$ and reduces the wage for signal 2 by the same amount, then the average payment, and in consequence both the principal's cost and the agent's reference point remain unaffected. Moreover, given that ε is not too large, also the loss premium the principal has to pay turns out to be independent of ε,[28]

$$LP^{Bell}(\varepsilon) = (2/9)(\lambda - 1)(w_{23} - w_1). \tag{II.8}$$

Thus, the individual rationality constraint also holds under this new contract. Since an increase in the degree of wage differentiation often is accompanied by an improvement of incentives, it is easily imagined that the principal benefits from specifying more than two wages. With loss aversion à la Kőszegi and Rabin, in contrast, the loss premium is strictly increasing in ε:

$$LP^{KR}(\varepsilon) = (2/9)(\lambda - 1)(w_{23} - w_1) + (2/9)(\lambda - 1)\varepsilon. \tag{II.9}$$

In order to illustrate the differences between these two concepts more vividly, we discuss in more detail the sensations of losses and gains under both concepts. Under Bell's notion of loss aversion, if $s = 1$ occurs, the loss felt is the same under both contracts. For $s = 2$, under the new contract the agent feels a lower gain than under the original contract. This lower gain, however, is exactly offset in expectations by an increased gain for $s = 3$. With loss aversion à la Kőszegi and Rabin, under the new contract, if $s = 1$ is realized then the agent feels a lower loss compared to the outcome for $s = 2$ and a larger loss compared to the outcome for $s = 3$. In expectations, these changes exactly cancel out. For $s = 2$, in contrast, under the new contract the agent now feels a loss in comparison to the outcome for $s = 3$, while under the original contract this comparison did not lead to the sensation of a loss. Thus, under the more differentiated wage scheme, the ex ante probability of incurring a loss is higher, which in turn increases the agent's gain-loss disutility.

The above observations suggest that increasing the degree of wage differentiation always increases the principal's cost if the agent is loss averse à la Kőszegi and Rabin. If, on the other hand, the agent is loss-averse according to Bell, then paying slightly

[28] The independence of the loss premium of ε does neither rely on the wages being equiprobable nor on using Bell's concept instead of Loomes and Sugden's or Gul's.

different wages for different signals is costless, except when differentiating wages that originally were equal to the reference point. With wage differentiation being less costly in the absence of mixed feelings, one would expect the optimal contract to be more differentiated under Bell's notion of loss aversion. Nevertheless, with losses still being painful for the agent, a fully contingent contract, which maximizes the scope for the agent to incur a loss, hardly seems optimal for a rich performance measure even if the agent's reference point has no stochastic component.

This conjecture is highly in line with the extant literature on incentive design under loss aversion.[29] With no unifying approach provided how to determine a decision maker's reference point, it is little surprising that all contributions differ in this aspect. Nevertheless, none of the earlier contributions applies a notion of loss aversion that allows for mixed feelings. David de Meza and David C. Webb (2007) apply the concept of Gul (1991), which posits that the reference point is the certainty equivalent of the prospect and thus is closely related to Bell (1985). The optimal contract consists of three regions: first compensation increases with performance up to the reference point, thereafter for a range of signals the wage equals the reference point, and for high performance the wage is strictly increasing in performance. As an alternative to Gul's concept, de Meza and Webb also consider the median as reference wage, which captures the idea that a loss is incurred at all incomes for which it is odds-on that a higher income would be drawn. Now, the optimal contract is discontinuous after the flat-part, but otherwise qualitatively similar. Thus, the optimal contract derived by de Meza and Webb provides a theoretical underpinning for the usage of option-like incentive schemes in CEO compensation.

Focusing only on gain-loss utility, Emil P. Iantchev (2009) applies the concept of Luis Rayo and Gary S. Becker (2007) to a multi-principal/multi-agent environment in which an agent's reference point is determined by the equilibrium conditions in the market.[30] Next to a dismissal region for very low performance, the optimal contract is found to display a performance-independent flat part for intermediate performance, which is followed by a region where rewards are increasing in performance. Evidence for this theoretically predicted contractual form is shown to be found in panel data from Safelite Glass Corporation.

Also abstracting from intrinsic utility but assuming that the reference point equals previous year's income, Ingolf Dittmann, Ernst Maug, and Oliver G. Spalt (forthcom-

[29] Nonstandard risk preferences different from loss aversion are analyzed in a moral hazard framework by Ulrich Schmidt (1999), who applies Menahem E. Yaari's (1987) concept of dual expected utility theory, and by Ján Zábojník (2002), who incorporates Friedman-Savage utility.

[30] The assumption that only changes in wealth matter is based on Kahneman and Tversky's (1979) original formulation of prospect theory.

ing) find that a loss aversion model dominates an equivalent risk aversion model in explaining observed CEO compensation contracts. The resulting contract under loss aversion qualitatively resembles the optimal contract identified by Iantchev (2009).

The commonality of all loss aversion concepts, irrespective of mixed-feelings possibly arising or not, is that there typically is a range of signals where payment does not vary with performance.[31] Without mixed feelings, however, the optimal wage schedule displays high sensitivity of pay to performance at least for signals that are very indicative for high effort. Thus, none of the aforementioned papers provides a rationale for the prevalence of binary payment schemes.[32]

To the best of our knowledge, Kohei Daido and Hideshi Itoh (2007) is the only paper that also applies reference dependence à la Kőszegi and Rabin to a principal-agent setting. The focus of Daido and Itoh greatly differs from ours. Assuming that the performance measure comprises of only two signals, two types of self-fulfilling prophecy regarding the impact of expectations on performance are explained. While sufficient to capture these two effects, the assumption of a binary measure of performance does not allow to inquire into the form that contracts take under moral hazard.

Though not placed in the literature on incentive design, the findings in Paul Heidhues and Kőszegi (2005) in spirit are closely related to our results. Here it is shown that consumer loss aversion à la Kőszegi and Rabin can explain why monopoly prices react less sensitively to cost shocks than predicted by orthodox theory. The driving force underlying this price stickiness is the aforementioned comparison effect: the probability of the consumer buying the good at some price is negatively affected by the comparison of this price to lower prices in the distribution. Therefore, just like our principal lumps together wages despite possibly negative incentive effects in order to avoid the unfavorable comparison of some relatively low wage with higher wages, the monopolist has an incentive to lump together prices even though this means foregoing the benefit from differentiating production according to cost. In a similar vein, Heidhues and Kőszegi (2008) provide an answer to the question why nonidentical competitors charge identical prices for differentiated products.

7. Closing Discussion

In this paper, we explore the implications of loss aversion à la Kőszegi and Rabin (2007) on contract design in the presence of moral hazard. With a stochastic reference

[31] Put differently, due to first-order risk aversion Holmström's informativeness principle is violated.

[32] De Meza and Webb (2007) find conditions under which a 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.

point component, increasing the number of different wages increases the agent's gain-loss disutility significantly without necessarily supplying strong additional incentives. The most drastic implication is the use of binary payment schemes even in situations where the principal has access to an arbitrarily rich performance measure and the optimal contract thus would be fully contingent under the standard notion of risk aversion. Moreover, we find that under reasonable conditions the optimal contract needs to specify only a cut-off performance, e.g., a sales quota. Thus, loss aversion provides a theoretical rationale for bonus contracts, the wide application of which is hard to reconcile with obvious drawbacks—such as seasonality effects—that come along with this particular contractual form. In the aforementioned sense loss aversion leads to simpler contracts than predicted by orthodox theory. Reduced complexity of the contract, however, is not a general prediction of loss aversion. We derived circumstances under which the optimal contract consists of stochastic payments if the agent is loss averse but does not include randomization if the agent is risk averse in the usual sense.

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 is the timing of expectation formation and actual decision making.[33] Under UPE a decision maker first forms his expectations, which determine his reference point, and thereafter, given these expectations, chooses his preferred action. To guarantee internal consistency, UPE requires that individuals can only make plans that they will follow through. With expectations being met on the equilibrium path under UPE, the expected utility takes the same form under both concepts. Since the optimality of bonus schemes is rooted in the agent's dislike of being exposed ex ante to numerous outcomes, we would expect bonus contracts to be optimal also under UPE.

Throughout the analysis we ignored diminishing sensitivity of the gain-loss function. A more general gain-loss function complicates the analysis because neither the incentive constraint nor the participation constraint are linear functions in the intrinsic utility levels any longer. Nevertheless, we expect that a reduction of the pay-performance sensitivity will benefit the principal in this case as well. Diminishing sensitivity 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 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.

[33] A dynamic model of reference-dependent preferences which allows for changes in beliefs about outcomes is developed in Kőszegi and Rabin (2009).

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, may have favorable incentive effects.

Last, while several notions of loss aversion proposed in the literature are forward looking in the sense that the reference point is determined by the decision maker's rational expectations, our findings depend on the mixed-feelings approach embodied in the concept of Kőszegi and Rabin. With the exact way of how expectations enter the process of reference point formation being an understudied question, this issue clearly warrants further investigation.

III. On Horns and Halos: Confirmation Bias and Job Rotation

Confirmation bias, which refers to unintentional and unknowing selectivity in the use of evidence, belongs to the major problems faced by organizations. In this chapter, we discuss job rotation as a natural solution to this problem. In a nutshell, adopting job rotation provides an organization that is plagued by confirmation bias with a more reliable informational footing upon which to base its decisions. Job rotation, however, also comes with a cost, e.g. a loss of productivity or a disruption of work flows. We study this trade-off and identify conditions under which job rotation and specialization are each optimal.

1. Introduction

"If one were to attempt to identify a single problematic aspect of human reasoning that deserves attention above all others, the confirmation bias would have to be among the candidates for consideration."

– *Raymond S. Nickerson*

Confirmation bias refers to unintentional and unknowing selectivity in the acquisition and use of evidence. Ample empirical evidence supports the view that once one has come to believe in a position on an issue, one's primary purpose becomes that of justifying or defending that position.[1] In consequence, regardless of whether treatment of evidence was evenhanded before the position was taken, it can become highly biased afterward. Though confirmation bias is considered as one of the most widely accepted notions of inferential errors, as suggested by the above quote by Nickerson (1999), its implications for organizational design have not been subject of thorough formal investigation.[2] This is surprising because in organizations there seems to be ample room for confirmation bias to arise and in consequence to adversely affect intra-organizational decision processes and organizational performance. In this paper, we aim at making a first step toward drawing out potential responses of organizational design to confirmation bias and its effects.

One aspect of organizational life where confirmation bias has major impact immediately comes to mind: performance appraisal.[3] Many, if not most performance measures regarding a firm's employees are subjective rather than objective in nature.[4] This makes performance appraisal a process by which humans judge other humans, thereby opening the door for behavioral biases and inferential errors to enter and – more importantly – to distort this process. Raters' bias in performance appraisal is considered a severe problem in practice. According to Brian Davis, executive vice president of

[1]See Nickerson (1999) for an excellent survey.

[2]Other behavioral biases have been considered in the literature on organizational theory: leniency, favoritism, or centrality bias on the side of supervisors, reference-dependent preferences, inequity aversion, or violation of procedure-neutrality on the side of employees, just to name a few. Surveys regarding the former and the latter kind of biases in the context of organizations are found in Prendergast and Topel (1993) and Camerer and Malmendier (2009), respectively.

[3]In the community practicing performance appraisal, confirmation bias is also referred to as the horns-and-halo effect, which refers to supervisors' tendency to judge employees as either good or bad, and then to seek evidence that supports that opinion. This, in turn, is one possible explanation for the so-called Matthew effect, which suggests that no matter how hard an employee strives, their past appraisal records will prejudice their future attempts to improve. For more on this, see http://www.performance-appraisal.com/bias.htm.

[4]For papers emphasizing this point, see, for example, Prendergast (1999) and MacLeod (2003).

Personnel Decisions International, "[t]he problem with rater-bias is that it takes away the organization's ability to objectively use data from performance evaluations with any validity. [...] [Y]ou can't count on the objectivity or accuracy of a performance assessment, and you have no differentiating data that allows you to make confident decisions about promotions, training, or leadership development."[5] In consequence, with bad promotion decisions having dire consequences, the biggest of which are lower employee morale, decreased productivity, and lost customer share, organizations have a vested interest in identifying the right person for a job because the cost of getting it wrong is high.[6]

In this paper we argue that organizational design provides a tool which is capable of thwarting confirmation bias not only in performance appraisal but also in other situations: job rotation.[7] Under confirmation bias the outcome of a judgment process often is determined by early pieces of evidence which color all subsequently received pieces of information, i.e., first impressions matter. By its very nature, in many situations job rotation creates "multiple first impressions" – and thus unbiased evaluations – by regularly breaking up the matches of the judging person and the situation to be judged. The work practice of job rotation, however, commonly is acknowledged to be associated with some sort of cost, e.g. a serious loss of productivity caused by a disruption of work flows or the sacrifice of job-specific human capital. We show that, when organizational members are subject to confirmation bias, incuring this cost for implementing job rotation may well be worthwhile for an organization in order to obtain a more accurate probability assessment upon which to base its decisions.

In Section 2, we briefly review some of the many forms that confirmation bias can take, survey some (mostly psychological) evidence for these phenomena, and finally

[5] See http://www.management-issues.com/2007/6/7/research/bias-blights-performance-reviews.asp. Further information about Personnel Decisions International (PDI), a Minneapolis-based consultancy firm, can be found at http://www.personneldecisions.com/.

[6] According to a survey of 444 organizations throughout North America conducted by Right Management, a globally operating career transition and organizational consulting firm, the average cost of coping with an employee who does not work out is 2.5 times his salary. According to Rick Smith, Senior Vice President of Right Management, "[t]here is a smaller margin for error today in selection and promoting people into key positions, and a greater need to target development efforts to ensure that they really make a difference." For the corresponding press release from 04/11/2006, see http://phx.corporate-ir.net/phoenix.zhtml?c=65255&p=irol-newsArticle&ID=849080&highlight=.

[7] Job rotation refers to a job practice which assigns an employee not to a single specific task but to a set of several tasks (associated with a meaningful change in job content) among which he rotates with some frequency. For evidence on job rotation being used by a significant and increasing number of companies in the United States and other OECD countries, see Osterman (1994, 2000), Gittleman et al. (1998), and OECD (1999).

present the model of confirmation bias proposed by Rabin and Schrag (1999), which we are going to apply throughout the paper.

Inspired by the anecdotal evidence presented above, in Section 3 we turn to the most immediate situation one can think of when pondering where confirmation bias might take effect in organizations: promotion decisions based on the evaluation of workers by their supervisors. We present a simple model in which, with different types of jobs being available, the efficient allocation of a worker depends on his ability, which is assumed to be commonly unknown. If the firm wants to base these decisions on a more solid informational footing by gathering additional information, due to an exogenously given need to delegate some tasks, it has to rely on supervisors to do so. Since under the assumptions we impose no incentive-compatibility or truthful-revelation complications arise, supervisors are happy to truthfully report their observations to the firm. The only friction that we allow for is that supervisors are subject to confirmation bias. The firm can choose between two types of work design, specialization or job rotation. If the firm opts for specializing the worker, he remains in one and the same division which leads to an increase in his productivity in this field of activity. Under specialization, however, the worker is evaluated by this division's supervisor exclusively. When supervisors succumb to confirmation bias, this leads to later evaluations being biased due to earlier established beliefs. If the firm decides to implement job rotation, on the other hand, the worker is placed in various of the firm's divisions and becomes a generalist who is less productive than a specialist. Under job rotation, however, the firm regularly breaks up the matches of supervisors and their subordinates, thereby creating multiple unbiased evaluations of many supervisors regarding one particular employee. We show that preventing confirmation bias from affecting supervisors' judgment can indeed outweigh the loss of productivity due to implementing job rotation. Moreover, we show that job rotation is more likely to be the optimal form of work design the stronger the degree of supervisors' confirmation bias is.

After discussing our modeling assumptions in Section 4, in Section 5 we provide an alternative interpretation of our model in order to emphasize its applicability to situations different from supervisor-worker relationships. We consider an employee who has to evaluate where a productive asset might be put to use most profitably. In contrast to the supervisor-worker setting, here job rotation does not sever the link between the judging person and the situation to be judged. In consequence an unbiased evaluation in this case probably is not to be obtained. Nevertheless, empirical evidence documents that preferential treatment of information supporting existing beliefs as well as overconfidence in one's own judgment can be reduced by forcing people to evaluate their own views, especially when that includes providing reasons against their current opinion.[8] By placing them in various positions, by its very nature job rotation forces employees to look at their field of activity from different perspectives, thereby most likely broadening their view and making them less susceptible for one-sided treatment of evidence. By showing that there is scope for the firm to benefit from the resulting more reliable probability assessment even when confirmation bias is merely reduced but not fully eliminated, we provide an explanation for the often found statement that firms prefer "well-rounded employees", which neither relies on the folk wisdom that future managers should be equipped with a broad view of the entire firm, nor on the need of multi-skilled workers in order to cope efficiently with technological change.[9]

Section 6 concludes by briefly summarizing our results, relating our findings to alternative theories of job rotation, and drawing out potential implications for empirical analysis.

Related Literature When it comes to naming potential costs of implementing job rotation, there almost seems to be unanimity in the theoretical literature: Transferring individuals to new jobs sacrifices job-specific human capital, and frequent job rotation may in consequence entail a serious loss of productivity. With regard to benefits of this particular kind of work design, on the other hand, over the years many explanations have been put forth why it may be worthwhile to incur the afore-mentioned loss in productivity. One of these explanations, formalized in Cosgel and Miceli (1999), posits that workers dislike monotonous jobs. In consequence, regular job transfers increase employees' motivation and overall satisfaction by reducing their boredom and keeping them interested in their jobs, which in turn allows firms to economize on wages. A large part of the theoretical literature, however, focuses on the effects of job rotation on firm learning by placing firms and their employees on very unequal informational footing, with the firm being in a disadvantageous position. In a framework where the firm can neither observe workers' effort nor the productivity of the jobs the workers are placed in (which subsequently is observed by the respective workers), both Ickes and Samuelson (1987) and Arya and Mittendorf (2004) show that job transfers alleviate the ratchet effect.[10,11] Abstracting from any moral hazard problems, Ortega (2001) finds

[8] See Perkins et al. (1991), Fischhoff (1977), Hoch (1984, 1985), Koriat et al. (1980), Tetlock and Kim (1987).

[9] See, for example, Schaeffer (1983) and Koike (1993) on the former argument, and Carmichael and MacLeod (1993) on the latter.

[10] The ratchet effect refers to workers' shirking in order to disguise the productivity of their jobs and to prevent an increase in performance standards.

[11] Arguing that under the tie-breaking rule used by Ickes and Samuleson (1987) there exists a second equilibrium in which both agents shirk in the productive job and thus are overall better off, Ma (1988) proposes an alternative compensation mechanism which uniquely implements the second-

that a firm can benefit from implementing job rotation in order to optimally match employees to jobs when there is uncertainty about both the profitability of different jobs and the productivity of different persons at different jobs. Eguchi (2005) considers a multi-task situation where, next to regular work activities, the worker can engage in influence activities which become more profitable for the worker the longer he is in his current position. It is shown that when the firm is harmed by this rent-seeking behavior of its employees but cannot use incentive payment schemes effectively due to difficulties in measuring workers' performance, frequent job transfers are useful to limit these influence activities. Finally, when the firm faces workers of different but unobservable ability, Arya and Mittendorf (2006) argue that implementing optional job rotation programs can help firms to better match pay to an employee's true worth by achieving a self-selection of the workers: When undertaking different tasks is costly for workers but less costly for highly talented employees than for employees of low talent, the former opt for the job transfer program in order to prove their versatility, whereas the latter refrain from doing so because it is too costly.

We see this paper as complementing the existing theoretical literature on job rotation in the following sense: We abstract from any hidden action problems (e.g. hidden gaming by supervisors or shirking and influence abilities by workers) and we also remove any informational disadvantage in the afore-mentioned sense, which organizational members might profitably exploit (e.g. private information of workers with respect to their own ability or workplace productivity). Moreover, the preferences of the organizational members are completely standard with no inherent taste for diversity in their field of activity. The only friction that we allow for is that organizational members are subject to confirmation bias. We show that – even in the absence of any informational asymmetries – the firm may benefit from incuring the cost for implementing job rotation in order to obtain a more accurate probability assessment upon which to base its decisions. This observation adds a new item, which is based upon psychological foundations, to the list of benefits associated with job rotation as work design.

2. Confirmation Bias

Empirical Evidence Confirmation bias, which refers to unwitting selectivity in both the acquisition and evaluation of evidence, comes along in many guises. When seeking or interpreting information, people display the tendency to give greater weight to evidence that is supportive to beliefs they hold dear than to information that is counter indicative of those established opinions. Empirical evidence for this prefer-

best identified by Ickes and Samuleson.

ential treatment of evidence, also referred to as my-side bias, is provided by Baron (1991, 1995), Perkins et al. (1983), Perkins et al. (1991), and Kuhn (1989).[12] Another well-documented phenomenon is the primacy effect, which refers to the finding that when information is gathered and integrated over time, evidence acquired in the early stages is likely to carry more weight than evidence acquired later in the process. In consequence, opinions are formed early in the process and subsequently acquired information evaluated in a way that is partial to that opinion.[13] The primacy effect, which can be seen as possible manifestation of belief persistence,[14] can also lead to a biased evaluation and interpretation of evidence that is subsequently acquired: people tend to question conflicting information more willingly than information supportive of preexisting beliefs (Ross and Anderson, 1982), to see ambiguous evidence more likely as supporting rather than disconfirming an established opinion (Lord et al., 1979; Darley and Gross, 1983), to explain away events that are inconsistent with a held position (Henrion and Fischhoff, 1986), and even to interpret evidence that should count against a hypothesis as counting in favor of it (Pitz et al., 1967).

The explanations that have been put forth to account for confirmation bias are numerous, ranging from "the desire to believe" over pragmatism and error avoidance to educational effects. At this point, however, we take the occurrence of this phenomenon as given.

A Formal Model In order to formally draw out the implications of confirmation bias for organizational design, we adopt the model of confirmation bias and belief formation proposed by Rabin and Schrag (1999). There are two exhaustive and mutually exclusive states of the world, $\theta \in \{\theta_L, \theta_H\}$. A priori, an individual considers both states of the world equiprobable, i.e., $\text{prob}(\theta = \theta_L) = \text{prob}(\theta = \theta_H) = 0.5$. In every period $t \in \{1, 2, 3, \ldots\}$ the person receives a signal, $s_t \in \{L, H\}$, that is correlated with the true state of the world. Signals received over time are independently and identically distributed with $\text{prob}(s_t = L|\theta = \theta_L) = \text{prob}(s_t = H|\theta = \theta_H) = \mu$, where $\mu \in (0.5, 1)$.

When subject to confirmation bias, the person may misinterpret a received signal that contradicts her current belief about the true state of the world, which is based on the sequence of signals she perceived so far. More precisely, with probability $q \in (0, 1)$

[12]Even if there is no "my side", i.e., even when people have no vested interest in the truth of a particular hypothesis, they appear to seek confirmatory information regarding this hypothesis. See, for example, Maynatt et al. (1977), Schwartz (1982), Zuckerman et al. (1995).

[13]See, for example, Nisbett and Ross (1980), Lingle and Ostrom (1981), Sherman et al. (1983).

[14]Belief persistence refers to the resistance of once established opinions to change even when faced with compelling disconforming evidence. See, for example, Ross et al. (1975), Ross (1977), Ross and Lepper (1980).

the individual misreads a signal s_t that conflicts with her current belief as actually supportive of her current opinion. Signals that are supportive of the currently held belief, on the other hand, are always interpreted correctly. Let the individual's (possibly erroneous) perception of signal s_t be denoted by $\sigma_t \in \{h, l\}$ for $t \in \{1, 2, 3, \ldots\}$. Given her perception of the information she is receiving, the individual each period updates her beliefs according to Bayes' Rule.

As an illustration of this process of signal perception and belief formation, suppose that the true state of the world is θ_L, but that the person currently believes that θ_H is more likely, i.e., $\text{prob}(\theta = \theta_H | \sigma_1, \ldots, \sigma_{t-1}) > 0.5$. With probability $1 - \mu$ the person receives a signal $s_t = H$, which she interprets as $\sigma_t = h$ with certainty because it is supportive of the currently held belief. With probability μ, on the other hand, the person receives a disconfirming signal $s_t = L$. While this signal is interpreted correctly as a disconfirming signal $\sigma_t = l$ with probability $1 - q$, with probability q this signal is misread as a supportive signal $\sigma_t = h$. At the end of the period, given her perception of the information she has received, the individual updates her beliefs according to Bayes' Rule. Figure III.1 summarizes this example.

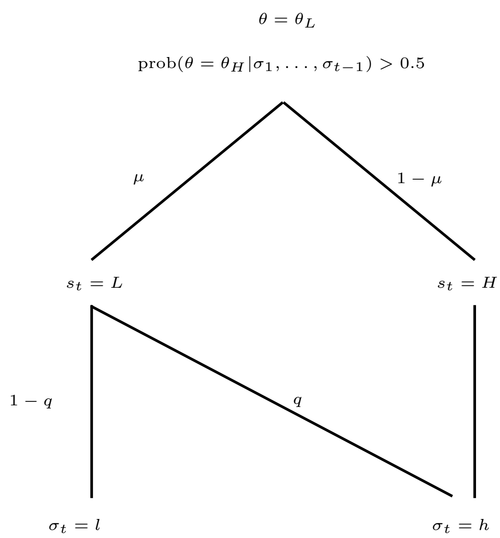

Figure III.1.: Perception of signals.

In order to summarize the distribution of a person's perceived signal σ_t more concisely, let $\mu^*(q)$ denote the probability that the person perceives a signal confirming her belief that one state is more likely when in fact the other state is the true state of the world. Analogously, let $\mu^{**}(q)$ denote the probability that the person perceives a signal confirming her belief that one state is more likely when in fact it is the true state of the world. Formally,

$$\begin{aligned}\mu^*(q) &= \text{prob}\,(\sigma_t = h\,|\,\text{prob}(\theta = \theta_H|\sigma^{t-1}) > 0.5,\ \theta = \theta_L) \\ &= \text{prob}\,(\sigma_t = l\ \,|\,\text{prob}(\theta = \theta_L\,|\sigma^{t-1}) > 0.5,\ \theta = \theta_H) \\ &= (1-\mu) + q\mu\end{aligned}$$

and

$$\begin{aligned}\mu^{**}(q) &= \text{prob}\,(\sigma_t = h\,|\,\text{prob}(\theta = \theta_H|\sigma^{t-1}) > 0.5,\ \theta = \theta_H) \\ &= \text{prob}\,(\sigma_t = l\ \,|\,\text{prob}(\theta = \theta_L\,|\sigma^{t-1}) > 0.5,\ \theta = \theta_L) \\ &= \mu + q(1-\mu).\end{aligned}$$

Note that $\mu^{**}(q) > \mu^*(q)$ for all $q \in (0,1)$ and $\mu \in (0.5, 1)$.

3. Supervision and Job Allocation

Suppose a firm hires a worker who has two periods of active work life. Both the firm and the worker are assumed to be risk neutral. The firm's objective is to maximize overall output over the two periods. The worker's ability is either high or low, $\theta \in \{\theta_L, \theta_H\}$ with $\theta_H > 0$. In the first period, neither the firm nor the worker know the worker's ability. It is common knowledge, however, that both types of workers are equally likely among the overall population, $\text{prob}(\theta = \theta_H) = 0.5$.

The firm comprises of two divisions. At the outset, the firm commits to one of two possible types of job design, specialization or job rotation. If the firm opts for specialization of the new worker, he is placed in one of these divisions and stays there for at least the first period. If the firm implements job rotation, the worker spends the first half of the first period in one division and the second half of the first period in the other division. Thus, under job rotation the worker becomes a generalist in the sense that he learns as much about one division as he learns about the other. Let $r \in \{1, 2\}$ denote the number of divisions that the worker is placed in during the first period, i.e., $r = 1$ corresponds to specialization and $r = 2$ to job rotation.[15]

We abstract from any moral hazard problems: presence of the worker is enough for the firm to benefit from his input, i.e., no costly effort from the worker is needed. In his first period with the firm, the worker has to be trained and has to learn work flows, organizational design, and communication channels. Since each new worker faces these basic tasks regardless of his talent or his work place, first-period output is assumed to be

[15] The assumption that the worker switches divisions only once under job rotation is shared with most contributions to the extant theoretical literature on job rotation, e.g. Ickes and Samuelson (1987), Cosgel and Miceli (1999), Ortega (2001), and Arya and Mittendorf (2004).

independent of both his ability and the division he is placed in. Moreover, first-period output is independent of the type of work design, job rotation or specialization. We normalize first-period output to zero. The worker's second-period output, on the other hand, depends on both his ability and the type of job he is allocated to in the following way: There are two types of jobs for the worker, $j \in \{A, B\}$, that the firm can install in the second period in any division the worker visited during the first period. Let y_j denote the worker's second-period output in job j. Output in job A is independent of the worker's ability, $y_A = \bar{y} > 0$. In job B, on the other hand, output depends on the worker's ability as follows:

$$y_B = \begin{cases} \bar{y} + k(r)\theta_H & \text{for} \quad \theta = \theta_H \\ 0 & \text{for} \quad \theta = \theta_L \end{cases},$$

where $0 < k(2) < k(1) = 1$. More vividly spoken, job A might be thought of as a back-office job where the worker has to do (possibly tedious but nevertheless straightforward) paperwork. Job B, on the other hand, could be that of a product designer or marketing manager, where skills like creativity or analytical thinking are important for success. With the impact of high talent on output being decreasing in the degree of rotation, $1 - k(2)$ represents the benefits of specialization. Let $\theta_H < \bar{y}$, which implies that the firm would place the worker in job A even under specialization if it had to rely on its prior beliefs when allocating the worker to a job in period 2.

For the sake of simplicity, we assume that once the worker starts working for the firm, he stays with that firm for both periods. Thus, all the firm has to do is to compensate the worker for his (discounted lifetime) reservation utility, which we assume to be zero.

Under these assumptions the only remaining decision the firm has to take is in what type of job to place the worker at the beginning of period 2. With $\theta_H \in (0, \bar{y})$, in order to allow for a meaningful analysis, the firm must be able to gather information about the worker's ability. Due to some exogenously given need for delegation, the firm itself cannot observe this information about the worker's ability, but has to rely on the divisions' supervisors for doing so. We assume that over his first period with the firm, there are two evaluation periods of equal length in each of which the worker is evaluated by the supervisor of the division in which he is currently placed.[16] Under specialization the worker is evaluated twice in one and the same division by this division's supervisor, whereas under job rotation he is evaluated exactly once in each division he is placed in and thus by two different supervisors. In each evaluation period, the current supervisor of the worker receives a signal $s_t \in \{L, H\}$, $t = 1, 2$, about the worker's ability. This signal represents, for example, the realization of some set of (at least to some extent)

[16] The (admittedly) ad hoc restriction to only two evaluation periods will be discussed at length in the following section.

subjective performance measures. Let $\mu = \text{prob}(s_t = H|\theta = \theta_H) = \text{prob}(s_t = L|\theta = \theta_L) \in (0.5, 1)$.

Supervisors are risk neutral and we abstract from any incentives for supervisors to lie about the signals they perceive, e.g. disutility from handing out bad evaluations. Moreover, we assume that supervisors costlessly observe signals and that the informativeness of the signals is independent of any costly effort of the supervisors. Under these assumptions, an arbitrarily small incentive to identify the true ability of the worker, e.g. an arbitrarily small stake in the firm's profits, will lead to the supervisors reporting truthfully. The only friction we allow for is that supervisors are subject to confirmation bias. As described in the previous section, with probability $q \in (0, 1)$ a supervisor misinterprets signals that contradict her current hypothesis about the worker's ability as supporting her hypothesis. Let the supervisor's perception of signal $s_t \in \{L, H\}$ be denoted by $\sigma_t \in \{l, h\}$. We assume that all supervisors share the same (common knowledge) prior about the worker being of high talent, $\text{prob}(\theta = \theta_H) = 0.5$, and that there is no communication among supervisors. In consequence, confirmation bias will only affect a supervisor's judgment under specialization when she receives subsequent signals about the same worker. Last, we assume that the firm is aware of the supervisors being subject to confirmation bias. If this was not the case, there would be no reason for the firm to implement anything else but specialization.

With regard to information transmission, at the end of each evaluation period, a supervisor reports her perceived signal immediately to the firm, where this information is stored. Thus, at the end of period 1, the firm is faced with a tuple of reports, $(\sigma_1, \sigma_2) \in \{(h, h), (h, l), (l, h), (l, l)\} \equiv \mathcal{M}$. We assume that both the content and the date of reception of these reports are verifiable, and that in consequence, when choosing the type of job design at the outset, the firm can commit to an allocation rule based on the content and the order of the supervisors' reports.[17] For a given job design which places the worker in $r \in \{1, 2\}$ divisions during the first period, this allocation rule $\mathcal{B}_r$ prescribes for which pairs of reports the worker is allocated to job B. Formally, either $\mathcal{B}_r \subseteq \mathcal{M}$ or $\mathcal{B}_r = \emptyset$, where the latter refers to the worker being allocated to job A no matter what.[18] Clearly, the optimal allocation rule depends on the updated posterior belief of the worker being of high talent, which in turn depends on the type of job design implemented. The timing of events is summarized in Figure III.2.

[17]This assumption allows us to sidestep the issue whether the firm itself is subject to confirmation bias. We will comment on this assumption in the next section.

[18]More precisely, an allocation rule for a job design with $r \in \{1, 2\}$ is a mapping $\mathcal{B}_r : \mathcal{M} \to \{A, B\}$, which prescribes for each pair of possible reports $(\sigma_1, \sigma_2) \in \mathcal{M}$ in which job the worker is placed in period 2. The above "operationalization" of such an allocation rule, however, will turn out to be quite convenient.

In a first-best situation, i.e., when the worker's ability is known to the firm, the firm would place the worker in job A when $\theta = \theta_L$ and in job B when $\theta = \theta_H$, where in the latter case the worker stays in one and the same division over the first period in order to capitalize on the benefits of specialization. When the worker's talent is unknown to the firm, it has to rely on the reports of the supervisors when allocating the worker to a job in period 2.

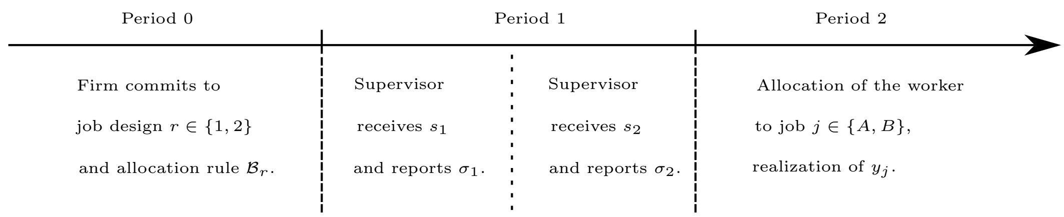

Figure III.2.: Timing of events.

Allocation under Specialization First, suppose the firm decides to reap the benefits of specialization and does not implement job rotation. Under specialization, after two evaluation periods the worker will be allocated to job B if and only if, given the updated posterior belief that the worker is highly talented, the expected output in job B exceeds the ability-independent output in job A, or equivalently, if and only if the firm's posterior belief about the worker being of high talent exceeds

$$\bar{p} := \frac{\bar{y}}{\bar{y} + \theta_H}.$$

Note that $\bar{p} \in (0.5, 1)$ due to our assumptions that $\theta_H \in (0, \bar{y})$. With supervisors being subject to confirmation bias, if a supervisor receives in the second evaluation period a signal which contradicts her current opinion about the worker's ability, with probability $q \in (0, 1)$ she misinterprets that signal as supporting her current opinion. When forming its updated posterior belief based on the supervisor's report at the end of the first period, the firm has to take into account the supervisor's possible misperception of the signals she received. In consequence, the order in which signals are received, or more precisely perceived, is important. Suppose, for example, the supervisor reports that she has observed two h signals. The firm now has to take into account that the supervisor, after having received an H signal in the first evaluation period, at the beginning of the second evaluation period considered the agent more likely to be of high ability than of low ability. Therefore, since with probability q she misinterprets an L signal as supporting her opinion, the probability that the supervisor perceived a second h signal is higher than the probability that she actually received a second H signal.[19] Let $p(\sigma_1, \sigma_2; q) := \text{Prob}(\theta = \theta_H | \sigma_1, \sigma_2; q)$ denote the firm's posterior believe about the worker being of high ability after the supervisor reports $(\sigma_1, \sigma_2) \in \mathcal{M}$ under specialization. Then, according to Bayes' rule,

$$p(h, h; q) = \frac{\mu\mu^{**}(q)}{\mu\mu^{**}(q) + (1-\mu)\mu^{*}(q)}.$$

Analogously we obtain $p(h, l; q) = p(l, h; q) = 0.5$, and $p(l, l; q) = (1-\mu)\mu^{*}(q)/[(1-\mu)\mu^{*}(q) + \mu\mu^{**}(q)]$. It is readily verified that $\mu > 0.5$ implies $p(l, l; q) < 0.5 < p(h, h; q)$ for all $q \in (0, 1)$. From above we know that the firm will allocate the worker to job B only if the ex post belief about the worker being of high ability exceeds $\bar{p} > 0.5$. Thus, under specialization, the worker will be placed in job B only if the supervisor reports two h signals and $p(h, h; q) \geq \bar{p}$.

Lemma III.1: *If $p(h, h; q) \geq \bar{p}$, then $\mathcal{B}_1 = \{(h, h)\}$. Otherwise, $\mathcal{B}_1 = \emptyset$.*

Allocation under Job Rotation Under our assumptions on intra-organizational information transmission, job rotation helps the firm to get rid of the supervisors' confirmation bias. Each evaluation period the worker is evaluated by a different supervisor, and each of these supervisors shares the common prior about the worker's ability since she encounters the worker for the first time. Thus, job rotation creates multiple unbiased "first impressions", which in turn allows the firm to derive a more accurate probability assessment about the worker's talent. Clearly, in this situation the order in which signals are observed is of no importance for the updated posterior belief. Formally, let $p(n_h, n_l) = \text{Prob}(\theta = \theta_H | n_h, n_l)$ denote the firm's updated posterior belief about the agent being highly talented, where n_h and n_l are the overall number of h signals and l signals, respectively, reported by the supervisors over the two evaluation periods. According to Bayes' rule we have

$$p(2, 0) = \frac{\mu^2}{\mu^2 + (1-\mu)^2}.$$

Analogously, we obtain $p(0, 2) = (1-\mu)^2/[\mu^2 + (1-\mu)^2]$ and $p(1, 1) = 0.5$. Since $\mu > 1/2$, we have $p(0, 2) < 0.5 < p(2, 0)$. Removing the distortion due to confirmation bias, however, comes at the cost of sacrificing the benefits of specialization since $k(2) < 1$. Under job rotation, the worker will be placed in job B only if the posterior belief about

[19] While the probability of receiving a second H signal when $\theta = \theta_H$ is μ, the probability that the supervisor perceives a second h signal is $\mu^{**}(q) > \mu$. Analogously, while the probability of receiving a second H signal when $\theta = \theta_L$ is $1 - \mu$, the probability that the supervisor perceives a second h signal is $\mu^{*}(q) > 1 - \mu$.

the worker being of high talent exceeds

$$\bar{\bar{p}} := \frac{\bar{y}}{\bar{y} + k(2)\theta_H}.$$

Since $k(2) \in (0,1)$, we have $0.5 < \bar{p} < \bar{\bar{p}} < 1$. Thus, under job rotation, the worker will be allocated to job B if and only if two h signals have been reported and $p(2,0) \geq \bar{\bar{p}}$.

Lemma III.2: *If $p(2,0) \geq \bar{\bar{p}}$, then $\mathcal{B}_2 = \{(h,h)\}$. Otherwise, $\mathcal{B}_2 = \emptyset$.*

Comparison of Job Designs The question of interest is whether job allocation under job rotation can outperform job allocation under specialization in terms of ex-ante expected output. So far we know that the allocation rule under specialization depends on whether or not $p(h,h;q)$ exceeds $\bar{p}$, whereas under job rotation it depends on whether or not $p(2,0)$ exceeds $\bar{\bar{p}}$. Since $\mu > 0.5$ and $q > 0$, we have $p(h,h;q) < p(2,0)$, which reflects that the firm is more confident that the worker is highly talented after two h signals being reported under job rotation than under specialization due to a more accurate probability assessment. With $k(2) < k(1) = 1$, on the other hand, we have $\bar{p} < \bar{\bar{p}}$, which accounts for the loss of productivity under job rotation. Thus, we have to distinguish the following cases:

(a) $\bar{p} \leq p(h,h;q)$ and $\bar{\bar{p}} \leq p(2,0)$;

(b) $\bar{p} \leq p(h,h;q)$ and $p(2,0) < \bar{\bar{p}}$;

(c) $p(h,h;q) < \bar{p}$ and $\bar{\bar{p}} \leq p(2,0)$;

(d) $p(h,h;q) < \bar{p}$ and $p(2,0) < \bar{\bar{p}}$.

Obviously, case (d) is of little interest since under both forms of job design the worker will always be allocated to job A, $\mathcal{B}_1 = \mathcal{B}_2 = \emptyset$, which yields output $\bar{y}$ with certainty. In case (a), the allocation rule is identical under both types of job design, since the worker is allocated to job B whenever two h signals are reported, and to job A otherwise, $\mathcal{B}_1 = \mathcal{B}_2 = \{(h,h)\}$; ex-ante expected output, however, may differ under both types of job design due to a different probability assessment on the one hand, and the benefit of specialization on the other hand. Cases (b) and (c) obviously give rise to different allocation rules: In case (b), while the worker is always placed in job A under job rotation, $\mathcal{B}_2 = \emptyset$, he is allocated to job B if two h signals are reported under specialization, $\mathcal{B}_1 = \{(h,h)\}$. In case (c), allocation rules are vice versa.

In order to compare job designs in cases (a)-(c), we first characterize these cases in terms of the underlying model parameters μ and $k(2)$ for given values $\bar{y}$, θ_H and

$q \in (0,1)$.[20] It is readily verified that $p(2,0) \geq \bar{\bar{p}}$ if and only if $k(2) \geq \bar{k}$, where

$$\bar{k} := \frac{(1-\mu)^2}{\mu^2}\frac{\bar{y}}{\theta_H}.$$

Since $k(2) < 1$, for $k(2) \geq \bar{k}$ to be possible we must have $\bar{k} < 1$. Regarding $\bar{k}$ as a function of μ, we find that $\bar{k} < 1$ if and only if $\mu > \bar{\mu}$, where

$$\bar{\mu} := \frac{\sqrt{\bar{y}}}{\sqrt{\bar{y}} + \sqrt{\theta_H}}.$$

By the assumption that $\theta_H \in (0, \bar{y})$ we have $\bar{\mu} \in (0.5, 1)$. Next, note that $p(h,h;q) < \bar{p}$ if and only if $\mu < \bar{\bar{\mu}}(q)$, where

$$\bar{\bar{\mu}}(q) := \frac{2\bar{y} - q(\bar{y} - \theta_H) - \sqrt{q^2(\bar{y}-\theta_H)^2 + 4\bar{y}\theta_H}}{2(1-q)(\bar{y}-\theta_H)}.$$

In the appendix we show that $\lim_{q\to 0}\bar{\bar{\mu}}(q) = \bar{\mu}$ and that, for all $q \in (0,1)$, $d\bar{\bar{\mu}}(q)/dq > 0$ and $\bar{\bar{\mu}}(q) < 1$, which implies that $\bar{\bar{\mu}}(q) \in (\bar{\mu}, 1)$ for $q \in (0,1)$. The fact that $\bar{\bar{\mu}}(q)$ is increasing in q reflects that if the distortion through confirmation bias becomes stronger, for the firm to be willing to allocate the worker to the ability-dependent job B under specialization the signal itself must become more reliable. Taken together, these observations allow us to establish the following lemma.

Lemma III.3: *Given $\bar{y}$, θ_H, and $q \in (0,1)$, we have*

(a) $\bar{p} \leq p(h,h;q)$, $\bar{\bar{p}} \leq p(2,0)$ *iff* $\mu \in [\bar{\bar{\mu}}(q), 1)$ *and* $k(2) \geq \bar{k}$;

(b) $\bar{p} \leq p(h,h;q) < p(2,0) < \bar{\bar{p}}$ *iff* $\mu \in [\bar{\bar{\mu}}(q), 1)$ *and* $k(2) < \bar{k}$;

(c) $p(h,h;q) < \bar{p} < \bar{\bar{p}} \leq p(2,0)$ *iff* $\mu \in (\bar{\mu}, \bar{\bar{\mu}}(q))$ *and* $k(2) \geq \bar{k}$.

Proof: See Appendix.

Note that μ has to be sufficiently large ($\mu > \bar{\mu}$) to allow for the possibility of job rotation being the optimal choice of work design. Intuitively, if the correlation of the (unbiased) signal with the true state of the world is too low per se, it does not pay off for the firm to incur the cost of job rotation in order to prevent this bad signal from becoming somewhat more distorted.

To compare job rotation and specialization in terms of ex-ante expected output, we introduce one further piece of notation. Let $P(r)$ denote the probability of two h signals being reported when the number of divisions the worker is placed in equals r. Then $P(1) = (1/2)(\mu\mu^{**}(q) + (1-\mu)\mu^*(q))$ and $P(2) = (1/2)(\mu^2 + (1-\mu)^2)$. Moreover, let $\mathbb{E}[y|r]$ denote the ex-ante expected output under a job design with $r \in \{1, 2\}$.

[20]For details, see the proof of Lemma III.3 in the Appendix.

Case (a): Under both specialization and job rotation the same allocation rule is implemented, $\mathcal{B}_1 = \mathcal{B}_2 = \{(h,h)\}$. Thus, $\mathbb{E}[y|2] > \mathbb{E}[y|1]$ if and only if

$$P(2)p(2,0)(\bar{y} + k(2)\theta_H) + (1 - P(2))\bar{y} > P(1)p(h,h;q)(\bar{y} + \theta_H) + (1 - P(1))\bar{y},$$

or equivalently, if and only if $k(2) > \bar{\bar{k}}(q)$, where

$$\bar{\bar{k}}(q) := 1 - \frac{1-\mu}{\mu} q \left[\frac{\bar{y}}{\theta_H} - 1\right].$$

First, note that $\bar{\bar{k}}(q) < 1$ for all $q \in (0,1)$. Moreover, it is readily verified that $\bar{\bar{k}}(q) \geq \bar{k}$ if and only if $\mu \geq \bar{\bar{\mu}}(q)$. Thus, in case (a), we have $0 < \bar{k} \leq \bar{\bar{k}}(q) < 1$.

Case (b): While the allocation rule under job rotation is $\mathcal{B}_2 = \emptyset$, under specialization we have $\mathcal{B}_1 = \{(h,h)\}$. Thus, $\mathbb{E}[y|2] \leq \mathbb{E}[y|1]$ if and only if

$$\bar{y} \leq P(1)p(h,h;q)(\bar{y} + \theta_H) + (1 - P(1))\bar{y},$$

or equivalently, if and only if $\mu \geq \bar{\bar{\mu}}(q)$. Since this last inequality is satisfied in case (b), specialization unconditionally outperforms job rotation. This result follows more immediately from the fact that under specialization the firm prefers to implement allocation rule $\mathcal{B}_1 = \{(h,h)\}$ instead of $\mathcal{B}_1 = \emptyset$.

Case (c): Under specialization the allocation rule is $\mathcal{B}_1 = \emptyset$, whereas under job rotation we have $\mathcal{B}_2 = \{(h,h)\}$. Thus, $\mathbb{E}[y|2] > \mathbb{E}[y|1]$ if and only if

$$P(2)p(2,0)(\bar{y} + k(2)\theta_H) + (1 - P(2))\bar{y} > \bar{y},$$

or equivalently, if and only if $k(2) > \bar{k}$. Since this last inequality is satisfied in case (c), job rotation unconditionally outperforms specialization. This result follows more immediately from the fact that under job rotation the firm prefers to implement allocation rule $\mathcal{B}_2 = \{(h,h)\}$ instead of $\mathcal{B}_2 = \emptyset$.

We summarize the above observations in the following proposition.

Proposition III.1: *Given $\bar{y}$, θ_H, $q \in (0,1)$, job rotation strictly outperforms specialization, $\mathbb{E}[y|2] > \mathbb{E}[y|1]$, if and only if (i) $\mu \in [\bar{\bar{\mu}}(q), 1)$ and $k(2) > \bar{\bar{k}}(q)$, or (ii) $\mu \in (\bar{\mu}, \bar{\bar{\mu}}(q))$ and $k(2) > \bar{k}$.*

Thus, given that the benefits of specialization are sufficiently small, there are two reasons for job rotation being superior compared to specialization. First, in case (c), there are different allocation rules implemented under the different types of job design. Under specialization, confirmation bias is so strong that the worker will always be placed in the ability-independent job A because the firm is (justifiedly) pessimistic – even if two h signals are reported – about the worker's talent.[21] Under job rotation, in contrast, with an unbiased probability assessment, the firm dares to place the worker in job B when two h signals are reported, which ex ante generates higher expected profits. Secondly, in case (a), both types of job design nominally implement the same allocation rule, i.e., the worker is allocated to job B if two h signals are reported and to job A otherwise. Under job rotation, however, due to unbiased reports, the probability of actually facing a highly-talented worker is higher than under specialization, which, again, leads to ex ante higher expected profits.

Having characterized the circumstances where job rotation outperforms specialization and vice versa, allows us to establish the following comparative static result.

Proposition III.2: *Given $\bar{y}$, θ_H, μ, and q such that $\mu \in [\bar{\bar{\mu}}(q), 1)$ and $k(2) \in [\bar{k}, \bar{\bar{k}}(q)]$. An increase in the degree of confirmation bias from q to $q' > q$ makes it more likely that job rotation strictly outperforms specialization.*

The intuition for this result is straightforward. In the original situation, a subcase of case (a) in Lemma III.3, under both types of job design the worker is allocated to job B if two h signals are reported and to job A otherwise. According to Proposition III.1, however, specialization outperforms job rotation in terms of ex ante expected output because the benefits of specialization are large. Under specialization, an increase in the degree of confirmation bias, q, reduces the posterior belief about the worker being highly talented after two h signals have been reported. The posterior belief under job rotation, in contrast, is unaffected by an increase in q. There are two reasons why this might lead to job rotation becoming the optimal form of job design. First, if the posterior belief under specialization is lowered sufficiently, the firm will adopt a different allocation rule under specialization and place the worker in job A no matter what, in which case job rotation unconditionally becomes superior. Formally, the increase in q raises $\bar{\bar{\mu}}(q)$. Letting $q < q'$, if $\bar{\bar{\mu}}(q) \leq \mu < \bar{\bar{\mu}}(q')$, then the shift from q to q' leads to a transition from case (a) to case (c) in Lemma III.3. Secondly, even if the firm sticks to the original allocation rule, since the reliability of the supervisor's report decreases under specialization, the threshold which the cost of implementing job rotation must not exceed in order for job rotation to be optimal, becomes less stringent, $\bar{\bar{k}}(q') < \bar{\bar{k}}(q)$. If $\bar{\bar{k}}(q') < k(2) \leq \bar{\bar{k}}(q)$, job rotation becomes the optimal form of job design. Thus, when confirmation bias becomes a more severe problem, the stronger distortion of the supervisor's reports under specialization is more likely to outweigh the loss in productivity that comes along with job rotation.

[21]Formally, given $\bar{y}$, θ_H, and μ, q is sufficiently large such that $\mu < \bar{\bar{\mu}}(q)$, and in turn, $p(h, h; q) < \bar{p}$.

Before we move on to a discussion of our modeling assumptions, we want to relate the above analysis to a statement found in Ickes and Samuelson (1987). There we read that " [w]hile uncertainty about employee productivity may be important, job transfers can optimally arise only if there is also uncertainty about the productivity of the job. Allowing uncertainty about employee characteristics [...] cannot serve as an alternative explanation for job transfers." As we have seen, however, when we allow for another type of friction in form of confirmation bias of supervisors, job rotation may be the optimal form of work design even if there is no uncertainty regarding job characteristics but only regarding employee characteristics.

4. Discussion

Multiple periods While we stripped our model bare of hidden action and hidden information problems on purpose, the restriction to two evaluation periods is not that voluntarily but imposed by Rabin and Schrag (1999)'s model of confirmation bias. To illustrate, suppose the firm comprises of three divisions, implements three evaluation periods, and rotates the worker three times when opting for job rotation. The allocation rule under specialization depends on how the firm's posterior belief compares to $\bar{p}$, whereas the allocation rule under job rotation depends on how the firm's posterior belief compares to $\bar{\bar{p}} = \bar{y}/(\bar{y} - k(3)\theta_H)$, where $k(3) < 1$ represents the cost of rotating the worker three times compared to specialization. Under both types of work design, with $0.5 < \bar{p} < \bar{\bar{p}}$, a necessary condition for the worker to be placed in the ability-dependent job B is that the firm's posterior belief about the worker being of high talent exceeds 0.5. Application of Bayes' rule and straightforward calculations reveal that the firm's posterior belief exceeds 0.5 if and only if at least two H signals have been reported. More precisely, $p(3,0) > p(h,h,h;q) > p(2,1) = p(h,l,h;q) = p(l,h,h;q) = \mu > p(h,h,l;q) > 0.5$. Thus, in order to compare specialization and job rotation in terms of expected output, we have to distinguish ten different cases, as illustrated in Table 1.

While dealing with that many cases clearly would be tedious enough, we run into further problems when characterizing these cases in terms of the underlying model parameter μ. For example, in order to determine the values of μ for which $p(h,h,h;q) < \bar{p}$, we have to figure out when

$$\frac{\mu(\mu + q(1-\mu))^2}{\mu(\mu + q(1-\mu))^2 + (1-\mu)((1-\mu) + q\mu)^2} < \frac{\bar{y}}{\bar{y} - \theta_H},$$

which basically boils down to finding the zeros of a polynomial of third order. With both the number of cases to consider and the degree of the polynomials characterizing the

	$\mathcal{B}_1$	$\mathcal{B}_2$
$p(h,h,h;q) < \bar{p},\ p(3,0) < \bar{\bar{p}}$	$\emptyset$	$\emptyset$
$\mu < \bar{p} \leq p(h,h,h;q),\ p(3,0) < \bar{\bar{p}}$	$\{(h,h,h)\}$	$\emptyset$
$p(h,h,l;q) < \bar{p} \leq \mu,\ p(3,0) < \bar{\bar{p}}$	$\{(h,l,h),(l,h,h),(h,h,h)\}$	$\emptyset$
$0.5 < \bar{p} \leq p(h,h,l;q),\ p(3,0) < \bar{\bar{p}}$	$\{(h,h,l),(h,l,h),(l,h,h),(h,h,h)\}$	$\emptyset$
$p(h,h,h;q) < \bar{p},\ \mu < \bar{\bar{p}} \leq p(3,0)$	$\emptyset$	$\{(h,h,h)\}$
$\mu < \bar{p} \leq p(h,h,h;q),\ \mu < \bar{\bar{p}} \leq p(3,0)$	$\{(h,h,h)\}$	$\{(h,h,h)\}$
$p(h,h,l;q) < \bar{p} \leq \mu,\ \mu < \bar{\bar{p}} \leq p(3,0)$	$\{(h,l,h),(l,h,h),(h,h,h)\}$	$\{(h,h,h)\}$
$0.5 < \bar{p} \leq p(h,h,l;q),\ \mu < \bar{\bar{p}} \leq p(3,0)$	$\{(h,h,l),(h,l,h),(l,h,h),(h,h,h)\}$	$\{(h,h,h)\}$
$p(h,h,l;q) < \bar{p} \leq \mu,\ 0.5 < \bar{\bar{p}} \leq \mu$	$\{(h,l,h),(l,h,h),(h,h,h)\}$	$\{(h,h,l),(h,l,h),(l,h,h),(h,h,h)\}$
$0.5 < \bar{p} \leq p(h,h,l;q),\ 0.5 < \bar{\bar{p}} \leq \mu$	$\{(h,h,l),(h,l,h),(l,h,h),(h,h,h)\}$	$\{(h,h,l),(h,l,h),(l,h,h),(h,h,h)\}$

Table III.1.: Allocation rules for three evaluation periods.

threshold levels for μ increasing in the number of evaluations, the necessary calculations soon become infeasible. Thus, in order to inquire into interesting questions of optimal rotation intervals or the optimal number of evaluation periods a more tractable model of confirmation bias is needed.

Commitment to an allocation rule The assumption that the firm can commit to an allocation rule at the outset can be dropped when the firm itself is not subject to confirmation bias when making the allocation decision based on the reports it received. With the firm's perception being unbiased, its updated posterior from report (σ_1, σ_2) is the same irrespective of whether it is calculated ex ante, before actually receiving this report, or ex post, i.e., after having received (and thus after having evaluated) this report. Therefore, if commitment to an allocation rule is not possible, after receiving the reports there is no incentive for the firm to deviate from the allocation rule which is optimal if commitment is possible before receiving the reports.[22]

The assumption that the head of the firm is not subject to confirmation bias does not seem that far-fetched. As we already mentioned in Section 2, for confirmation bias to arise, pieces of evidence have to be received and evaluated subsequently, and there also often needs to be some degree of vagueness that leaves room for misinterpretation. Both can be imagined not to be the case for the head of the organization: First, when deciding in which job to place a worker, the head of the firm refers to the reports and recommendations of the supervisors that evaluated this worker. With forced dis-

[22]The firm's updated posterior beliefs ex ante and ex post, however, would not coincide if the firm also succumbs to confirmation bias when evaluating the reports it received. In this case, the ability to commit to an allocation rule clearly makes a difference.

tribution rankings or specified evaluation schemes often being used in practice, there probably is less room for arbitrariness in interpretation of these reports than in the original evaluation of the worker by his supervisor(s). Secondly, with bosses and CEOs almost always having got their hands full, the head of the firm will not spend days over days in advance pondering where to place the worker, but more likely he will focus on the decision shortly before it is due with most (all) relevant information available. Though only one evaluation report or memo can be read at a time, for the head of the firm this removes the sequential character of information acquisition at least to some degree. Third, while probably only one supervisor at a time is responsible for evaluation of the worker, when making the allocation decision, the head of the organization most likely comprises of several actors, like the personnel manager, the managers in whose division the worker might be placed in, and so on. One might imagine that preferential treatment of evidence might be less likely to occur when discussing pros and cons of a decision with other equally skilled people. Last, the assumption of a rational head of the organization also is in line with the largest part of the literature on behavioral industrial organization, where rational firms/principals interact with behaviorally biased consumers/agents: while the former know about the biases of the latter, the latter often are assumed to be naive and do not know about their own bias.[23]

5. An Alternative Interpretation

In this section we want to emphasize applicability of the above analysis to situations different from supervisor-worker relationships. In order to do so, we provide an alternative interpretation of our model. Apart from very few exceptions, we basically just relabel variables. Therefore, as long as there is no danger of confusion, we will make use of notation and definitions introduced before without explicitly saying so.

Suppose a risk neutral firm, which comprises of several divisions, e.g. production and marketing, faces two opportunities where to deploy an asset in the second period, project A or project B. For example, the asset might be a machine used in production, and the two projects represent the production of different products. While the return of project A is assumed to be riskless, $R_A = \bar{R} > 0$, the return of project B depends on the true state of the world, $\theta \in \{\theta_L, \theta_H\}$, as follows:

$$R_B = \begin{cases} \bar{R} + \Delta & \text{for} \quad \theta = \theta_H \\ 0 & \text{for} \quad \theta = \theta_L \end{cases},$$

[23] See, for example, Eliaz and Spiegler (2006), DellaVigna and Malmendier (2004), Gilpatric (2008), and Gabaix and Laibson (2006). For a survey on bounded rationality in industrial organization, see Ellison (2006).

where $\Delta \in (0, \bar{R})$. Thus, in the above example, we can think of project A as the production of a product which is already established in the market and for which the firm is familiar with the production process. Project B, on the other hand, can be thought of as the production of a newly developed product, the success of which depends on factors like market acceptance or whether there will be complications in the production process. In this sense, θ_H can be interpreted as a situation, i.e., a constellation of market characteristics and technological circumstances, in which the launch of the new product will be successful, whereas it will be a failure if the state is θ_L. A priori, the two states are equally likely. Thus, without any further information, the firm allocates the asset to project A.

In order to obtain further information about the true state of the world, the firm can hire a risk neutral worker in period 1, who lives and stays with the firm for two periods. The worker's (discounted lifetime) reservation utility equals zero. Once again, we abstract from hidden information or hidden action problems: there is no uncertainty about the worker's talent, and the worker's output is (for expositional purposes only) equal to zero in both periods. The only meaningful task the firm can assign to the worker is to gather information about the true state of the world in order to improve the decision where to place the productive asset.

Over the first period, the worker costlessly receives two subsequent signals about the true state of the world, $s_t \in \{L, H\}$ with $t = 1, 2$. These signals are identically and independently distributed according to $\mu \in (0.5, 1)$. Due to confirmation bias, however, the signal the worker receives may differ from the signal he actually perceives: once he comes to believe that one state of the world is more likely than the other, with probability $q \in (0, 1)$ the agent misinterprets a signal which contradicts this hypothesis as actually supporting this hypothesis. Let the worker's perception of signal $s_t \in \{L, H\}$ be denoted by $\sigma_t \in \{l, h\}$. We assume that the firm itself does not receive the signals, and therefore has to rely on the reports of the worker. The firm, however, is aware of the worker's confirmation bias. With the gathering of information being costless and the worker being risk neutral, an arbitrarily small incentive to identify the true state of the world, e.g. an arbitrarily small stake in the return of the second-period project, will lead to the worker reporting truthfully what signals he has perceived. Both the content and the order of these reports, which we can identify with the worker's perceived signals $(\sigma_1, \sigma_2) \in \mathcal{M}$, are verifiable.

At the outset, the firm can commit to a particular type of job design, specialization or job rotation. Letting $r \in \{1, 2\}$ denote the number of divisions the worker is placed in during the first period, $r = 1$ corresponds to specialization and $r = 2$ to job rotation. If the firm opts for implementing job rotation and makes the worker switch divisions during his first period with the firm, it incurs a cost $c > 0$.[24] Moreover, the firm can commit to an allocation rule, which prescribes for which reports of the worker the asset is placed in project B. Thus, for a job design with $r \in \{1, 2\}$, this allocation rule is either $\mathcal{B}_r = \emptyset$ or $\mathcal{B}_r \subseteq \mathcal{M}$.

We do not believe it to be too far-fetched to assume that job rotation will reduce confirmation bias in this scenario as well. Suppose the firm opts for specialization and worker is placed in the production division throughout his first period with the firm. Though he might be able to get all the relevant information from the marketing division, it is easy to imagine that he will see all bits of information he gathers through the eyes of a production engineer, thus attaching too much weight to technological aspects and too little weight to market related data.[25] If the firm implements job rotation, on the other hand, during the first period the worker switches from production into marketing, and thus basically is forced to open his eyes more widely with respect to the market-related data as well. Thus, while the agent already holds some belief about the true state of the world when being placed in the marketing division, the new perspective from which he now has to assess the problem might make him more willing to let go of this hypothesis. However, since one and the same worker evaluates one and the same problem, it is likely that confirmation bias will merely be reduced by job rotation but not fully eliminated. Thus, letting q_r denote the probability that the agent misinterprets a contradicting signal under a job design which places the worker in $r \in \{1, 2\}$ divisions during the first period, we assume $0 < q_2 < q_1 < 1$.

Defining

$$\bar{c} := \frac{\mu(1-\mu)(q_1 - q_2)(\bar{R} - \Delta)}{2}$$

and

$$\bar{\bar{c}} =: \frac{\mu(\mu + q_2(1-\mu))\Delta - (1-\mu)((1-\mu) + q_2\mu)\bar{R}}{2},$$

and following the lines of the analysis in Section 3, we obtain the following result.

[24] Basically $c > 0$ may reflect any cost possibly associated with job rotation. Campion et al. (1994), for example, identify productivity losses and disruption of work flows for both the department gaining a rotating employee and the department losing the employee as potential costs of job rotation, resulting from training requirements in the first case and from having a vacancy in the second case. Also Burke and Moore (2000) draw attention to reverberating negative effects of job rotation on nonrotaters' perception of organizational justice.

[25] Evidence for experts being more confident than justified when making judgments in their own areas of expertise is provided by Kidd (1970), Loftus and Wagenaar (1988), Oskamp (1965) regarding engineers, attorneys, and psychologists, respectively.

Proposition III.3: *Given $\bar{R}$, Δ, q_1, q_2, job rotation strictly outperforms specialization, $\mathbb{E}[R|2] > \mathbb{E}[R|1]$ if and only if (i) $\mu \in [\bar{\bar{\mu}}(q_1), 1)$ and $c < \bar{c}$, or (ii) $\mu \in (\bar{\bar{\mu}}(q_2), \bar{\bar{\mu}}(q_1))$ and $c < \bar{\bar{c}}$.*

Proof: See Appendix.

The above result is familiar by now: if the benefit of specialization is sufficiently small, job rotation may be superior to specialization for two reasons. First, in case (ii), with confirmation bias being strong under specialization, the firm implements a very conservative allocation rule under specialization which places the asset always in project A, whereas under job rotation the asset is placed in project B if two h signals are reported and in project A otherwise. This more "daring" allocation rule, which is based on a more reliable probability assessment, yields higher expected profits. In case (i), on the other hand, even though allocation rules are identical under both types of work design, under job rotation, due to unbiased reports, the probability of the worker being of high talent after two h signals have been reported is higher than under specialization, which yields higher expected profits. Moreover, if the degree of confirmation bias under specialization increases or job rotation becomes more efficient in reducing the employees degree of confirmation bias, i.e., if q_1 increases or q_2 decreases, it becomes more likely that job rotation is the optimal form of work design. This follows from the fact that the threshold which the cost of implementing job rotation must not exceed in order for job rotation to be optimal becomes less stringent, $d\bar{c}/q_2 = -d\bar{c}/q_1 < 0$ and $d\bar{\bar{c}}/dq_2 < 0$.

As we briefly mentioned in the introduction, the literature on management development and employee learning recommends job rotation in order to endow the managers-to-be with a deeper understanding of more aspects of business, which they will need as they move up to broader jobs, or to help employees to cope better with uncertainty and technological change. The above analysis suggests, however, that a firm may have an incentive to provide its employees with a broader view of the organization even in the absence of such considerations: if knowledge of different organizational aspects makes employees less susceptible to confirmation bias, and if the firm has to base some of its decisions on its employees' judgments, job rotation may help to provide a more profound informational footing for the firm's decisions.

6. Conclusion

In this paper we examined a setting in which an organization is faced with its members being subject to confirmation bias, i.e., the tendency to treat subsequent information partially after an initial position has been taken. Given that job rotation is able (i) to sever the link between the judge and the situation to be judged, or (ii) to force the judge to be more open-minded for contradicting evidence, we have shown that implementing this particular form of job design may be profitable for the organization, even if it comes along with certain costs. The reason is that job rotation leads to a more reliable informational footing for the organization's decision making. We do not, however, obtain a call for a universal mandate for job rotation, but we find that optimality of job rotation is circumstance specific. In particular, the higher the degree of confirmation bias and the lower the cost associated with the implementation of job rotation, the more likely it is that job rotation is superior to specialization.

As briefly mentioned in the introduction, there are three major approaches to explain why work place organization may take the particular form of job rotation: employee motivation, employee learning, and employer learning. The employee motivation theory posits that job rotation helps to make work more interesting, thereby in particular providing motivation for so-called "plateaued" employees, i.e., employees with limited promotion prospects. The employee learning theory, on the other hand, contents that job rotation is an effective way to develop employees' abilities and to improve organizational knowledge in order to help prepare junior employees to become top managers or to better cope with uncertainty. Last, according to the employer learning theory, job rotation improves job assignments by providing the employer with information about the employee's abilities, both general and job-specific, and also job-specific factors unrelated to the employee. Though we do not see an immediate connection to the first of these approaches, the two tales told in this paper suggest that the presence of confirmation bias in organizations might interact with the two latter explanations, employee and employer learning. As for employer learning, our model about employees' evaluation by supervisors indicates that job rotation may become an even more valuable learning device for the firm when confirmation bias is an issue because it may prevent distortion of the signal that the employer receives. The alternative interpretation of our model, on the other hand, in a sense links employer and employee learning theory: though the ultimate goal of the employer is to learn where best to deploy the asset, when confirmation bias is present this may be achieved most profitably by making the employee learn to know the different building blocks of the organization in order to broaden his view and make him less susceptible for partial treatment of information.

In particular this last observation might be relevant for empirical analysis. In a rigorous test of the afore-mentioned explanations for the practice of job rotation, Eriksson and Ortega (2006) find "only very limited support for the employee motivation hypothesis, [but that] statistical evidence is more amenable to the employee learning hypothesis and employer learning hypothesis."[26] This is correct in the sense that a number of the hypothesized relationships between job rotation and the set of relevant variables were found to be in the predicted direction at a statistically significant level, e.g. a positive correlation between the use of job rotation and firm size or the number of hierarchical levels, which is consistent with both employee and employer learning theory. Regarding hypotheses for which the two learning theories predict different directions, however, there is no clear-cut result which theory better explains the data. For example, the finding that firms that spend more to train their employees are more likely to use job rotation schemes is favorable to the employee learning hypothesis but contradicts the employer learning hypothesis. Tenure in the industry not having a statistically significant effect on rotation, on the other hand, is consistent with the employer learning theory but contradicts the employee learning theory. In the light of our second story, we believe that these two theories sometimes cannot be treated separately but have to be seen as interwoven with each other. Therefore, in order to obtain even sharper predictions, it might be insightful to differentiate cases where the ultimate goal of employee learning is firm learning from cases of pure employee learning.

Last, we want to point out a more directly testable implication of this paper. We have seen that the stronger the degree of confirmation bias, the more likely is job rotation the optimal form of workplace design. In consequence, we should expect to find rotation arrangements more often in firms where there is more scope for confirmation bias to arise. While the degree of confirmation bias might be quite difficult to measure per se, there might be several ways to operationalize its measurement. For example, based on the observation that for confirmation bias to arise there needs to be some room for misinterpretation of evidence, the extent to which evaluation of employees is based upon subjective performance measures, which (by their very nature) are more vague and thus more susceptible to misinterpretation than objective performance measures, might serve as an indicator for the presence and strength of confirmation bias.

[26] Arguing that a satisfactory test of the three major theories of job rotation should combine a representative sample of establishments with data on employee characteristics, Eriksson and Ortega (2006) merge a representative survey of Danish firms with the employer-employee linked panel constructed by Statistics Denmark, which provides data on each employee at the sampled firms. The resulting database is richer than most surveys of establishments and provides more representative evidence than do single-firm case studies.

IV. Price Discrimination in Input Markets: Downstream Entry and Welfare

The extant theory on price discrimination in input markets takes the structure of the intermediate industry as exogenously given. This paper endogenizes the structure of the intermediate industry by allowing for costly entry and examines the effects of banning third-degree price discrimination on market structure and welfare. We identify situations where banning price discrimination leads to either higher or lower prices for all downstream firms. These findings are driven by the fact that upstream profits are discontinuous due to entry being costly. Moreover, permitting price discrimination fosters entry which in many cases improves welfare. Nevertheless, entry can also reduce welfare because it may lead to a severe inefficiency in production.

1. Introduction

"There are several ways in which the [upstream] manufacturer may influence the number of [downstream] retailers. [...] [T]he manufacturer may indirectly control the number of dealers through his pricing policy [...] ."

— *Michael L. Katz (1989)*

An ubiquitous assumption in the extant theory on third-degree price discrimination in input markets is that the structure of the intermediate industry is rigid. Abstracting from entry into the intermediate industry ignores the fact that pricing decisions of the upstream supplier are a major determinant of the resulting industry structure and market outcome. These pricing decisions in turn are determined by the pricing instruments available to the upstream supplier, in particular whether price discrimination is feasible or not. In this paper, we endogenize the structure of the intermediate industry and examine the effects of banning price discrimination in input markets on industry structure and welfare.

Our modeling assumptions are shared by a large part of the extant literature: a monopolistic upstream firm supplies an input that is used by firms in an intermediate industry to produce a final product. The upstream supplier makes a take-it-or-leave-it offer to each of the downstream firms, specifying a per-unit wholesale price at which that firm can procure any desired quantity of the input. The new feature in our model is that one of the downstream firms has yet to decide whether to incur a strictly positive entry cost in order to become active in the intermediate industry.

If downstream firms operate in separate markets and if the entry cost imposes a binding restriction on the choice of wholesale prices under either regime, then, depending on the relative efficiency of the potential entrant, price discrimination can lead to lower or higher wholesale prices for all downstream firms compared to uniform pricing. This immediately translates into price discrimination being strictly welfare enhancing or welfare reducing, respectively. Irrespective of whether downstream firms operate in separate markets or compete in the same market, price discrimination fosters entry. With separate downstream markets, opening of a new market under price discrimination but not under uniform pricing is a sufficient condition for a ban on price discrimination to be welfare harming. If downstream firms compete à la Cournot, then entry alleviates the distortion arising from double marginalization. Under discriminatory wholesale pricing, however, this beneficial effect of entry can be offset by entry being costly and an allocative inefficiency in production induced by the upstream supplier's discrimination against the more efficient firm.

The theoretical debate about the welfare effects of banning third-degree price discrimination in intermediate-goods markets was initiated by Katz (1987). His seminal paper considers a vertically related industry where the upstream market is monopolized and the downstream industry consists of a large chain that competes in several downstream markets with a small local store. Katz shows that permitting price discrimination reduces welfare unless it prevents inefficient backward integration by the chain of stores. The finding of Katz is generalized by DeGraba (1990) to a long-run analysis where downstream firms can invest in cost reduction. Here, a ban on price discrimination does not only increase welfare in the short run, but also is beneficial in the long run. The reason is that the more efficient downstream firm is charged a higher wholesale price under price discrimination than under uniform pricing. Thus, under price discrimination the benefit of lower production costs is partially offset by a higher wholesale price, which reduces a firm's investment incentives. Yoshida (2000) extends the previous models to the case where downstream firms operate with Leontief-type technologies.[1]

More recent contributions relax the assumption that the upstream firm has all the bargaining power. Inderst and Valetti (2009) posit that downstream firms have access to an alternative source of input supply. In their model the more efficient firm receives a discount. In consequence, price discrimination provides higher incentives to invest in cost reduction and thus—at least for linear demand—can result in higher welfare than uniform pricing. While Inderst and Valetti still assume that the upstream firm makes a take-it-or-leave-it offer, O'Brien (forthcoming) assumes that the wholesale prices are determined by Nash bargaining. This also gives rise to circumstances where banning price discrimination is socially harmful.

Last, O'Brien and Shaffer (1994) and Inderst and Shaffer (2009) relax the assumption that the upstream supplier is restricted to linear wholesale prices and allow for two-part tariffs. In O'Brien and Shaffer, while a ban on price discrimination may benefit downstream firms, it always does so at the expense of consumers and total welfare.[2] In the setting of Inderst and Shaffer (2009), optimal wholesale prices are shown to amplify differences in downstream firms' competitiveness. A ban on price discrimination in consequence reduces allocative efficiency and may lead to higher wholesale prices for all downstream firms, resulting in lower welfare.

All the aforementioned papers take the structure of the intermediate industry as exogenously given. This paper, in contrast, endogenizes the structure of the intermediate

[1] Valetti (2003) generalizes the results obtained in Yoshida (2000) beyond the case of linear demand.

[2] Analyzing a similar model but assuming that the upstream firm competes against a fringe, Caprice (2006) shows that banning price discrimination may cause welfare to increase.

industry by allowing for costly entry, and derives implications of banning price discrimination for industry structure, consumers' surplus, and welfare.[3] As was first reasoned by Bork (1978), allowing a final-good monopolist to price discriminate can lead to more markets being served, which in turn improves welfare.[4] This entry-promoting and in turn welfare-improving effect of price discrimination is also operative in our model. But even when all markets are served under either pricing regime, we derive circumstances where price discrimination leads to either overall higher or overall lower prices than uniform pricing. These cases arise from entry being costly and do not crucially rely on any assumptions on the demand function. Thus, in a nutshell, the established welfare implications of banning third-degree price discrimination with an endogenous market structure for final-good markets do not extend to the case of intermediate-good markets.

The rest of the paper is organized as follows: In Section 2, we introduce our model with downstream firms operating in separate markets. This model is analyzed for the cases of a less efficient entrant and a more efficient entrant in Section 3 and Section 4, respectively. Section 5 introduces Cournot competition between downstream firms. We conclude in Section 6.

2. A Model of Separate Markets

Consider a vertically related industry where the upstream market is monopolized by firm U. The upstream monopolist produces an essential input that is supplied to a downstream sector. For simplicity we assume that U produces without costs. There are potentially two downstream firms, $i \in \{I, E\}$, that transform one unit of input into one unit of a final good. While firm I, the incumbent, is already active in the downstream industry, firm E, the entrant, has to expend an entry cost $F > 0$ to become active in the downstream industry. Downstream firm i produces at constant marginal cost $k_i \in \{0, k\}$, $k > 0$, and without fixed cost.

The sequence of events is as follows: First, U can make a take-it-or-leave-it offer to each downstream firm.[5] Under price discrimination, U offers each downstream firm

[3] In a linear demand model, Haucap and Wey (2007) also consider endogeneity of market structure in intermediate good markets. Abstracting from any real entry decision in the sense of incuring an entry cost, their findings, in contrast to our results, closely parallel the established findings for final-good markets.

[4] This finding is formally established by Hausman and MacKie-Mason (1988).

[5] The assumption of the upstream supplier having all the bargaining power, "which arguably can be justified on the grounds that for antitrust purposes the considerations of price discrimination in intermediate-goods markets is primarily relevant if the supplier enjoys a dominant position" (Inderst and Shaffer, 2009, p.4) is common in the extant literature. Exceptions are O'Brien and

a possibly different wholesale price w_i, whereas under uniform pricing the same price $w_i = w$ applies to both firms.[6] Thus, upon accepting U's offer, downstream firm i's effective marginal cost is $c_i = w_i + k_i$. In stage two, after observing the contracts offered by U, firm E decides whether or not to enter the downstream industry at cost $F > 0$. In stage three, all active firms in the downstream industry purchase a nonnegative quantity of the input from U, transform this input into the final good, and sell the produced output to consumers. We abstract from any commitment problems and assume that U can credibly commit to the prices quoted in this first stage.[7]

First, we focus on the case where the downstream firms serve independent markets. We assume that both markets are symmetric and thus characterized by the same inverse demand function $P(q)$. The inverse demand function is assumed to be strictly decreasing and thrice differentiable where $P > 0$. Moreover, we impose the standard assumption $P'(q) < \min\{0, -qP''(q)\}$ where $P > 0$.[8] The equilibrium concept employed is subgame perfect Nash equilibrium in pure strategies.

We impose an additional assumption that ensures that U's maximization problem is well-behaved under either pricing regime.

Assumption (A1): *Downstream marginal revenue is concave,* $3P''(q) + qP'''(q) \leq 0$*, whenever* $P > 0$.

Next to technical issues, there is another reason for this assumption: as was shown by Katz (1987), if downstream firms engage in Cournot competition, then under price discrimination the downstream firm with the lower marginal cost will be charged a higher wholesale price than the downstream firm with the higher marginal cost.[9] The firm with lower own marginal cost has the more inelastic demand for the input, which causes the supplier to charge this firm a higher price. While the peculiarity of Cournot

---

Shaffer (1994) and O'Brien (forthcoming).

[6] In restricting the upstream supplier to linear wholesale contracts we follow Katz (1987), DeGraba (1990), Yoshida (2000), O'Brien (forthcoming), and Inderst and Valetti (2009, forthcoming). Though obviously restrictive, this assumption "can be defended on grounds of possible realism", as argued in Inderst and Valetti (2009, forthcoming). From a theoretical perspective, Iyer and Villas-Boas (2003) and Milliou et al. (2004) provide some support for the use of linear wholesale contracts. Both these papers show that linear wholesale contracts can emerge as equilibrium outcome when upstream and downstream firms can bargain over the form of their contractual arrangement.

[7] At a later point we make clear, which of our results are driven by this assumption.

[8] See, for example, Vives (1999).

[9] This result, which is also obtained by DeGraba (1990) and Yoshida (2000), holds as long as there are no additional restrictions on the input supplier's price setting, such as backward integration into the production process by downstream firms considered by Katz (1987) or demand-side substitution considered by Inderst and Valetti (2009, forthcoming).

competition that total output only depends on the sum of effective marginal cost allows Katz (1987) to obtain this result in considerable generality, this is not possible in the case of separate markets. Here, Assumption (A1) provides a sufficient condition for the demand of the more efficient downstream firm being less elastic, which in turn implies that price discrimination results in a higher wholesale price for the more efficient firm. Thus, next to reasons of analytical convenience, we impose this assumption in order to maintain comparability to the earlier models of price discrimination in input markets.

In order to state our results as concise as possible, we restrict attention to situations where U considers it optimal to serve both firms under uniform pricing at least for sufficiently small entry cost. A sufficient condition for this is that the less efficient firm is not too inefficient in the sense that it demands a strictly positive quantity when charged the optimal discriminatory wholesale price $w^d(0)$ for the more efficient firm. Formally, letting the optimal quantity produced by an active downstream firm i be denoted by $q(c_i) := \arg\max_q \{q[P(q) - c_i]\}$, we impose the following assumption:

Assumption (A2): *Marginal cost k is such that $q(w^d(0) + k) > 0$.*

As a tie-breaking rule, we assume that U serves only the incumbent market when indifferent between the two possible structures the intermediate industry can take.

3. The Analysis

In this section, it is assumed that the potential entrant is less efficient than the incumbent, i.e., $0 = k_I < k_E = k$. As a preliminary consideration, note that an active downstream firm in stage 3 realizes gross profits $\pi(c_i) := q(c_i)[P(q(c_i)) - c_i]$, and that both $q(c_i)$ and $\pi(c_i)$ are strictly decreasing in effective marginal cost c_i where $q > 0$. Firm E will enter the intermediate industry if and only if its profits in stage 3 exceed the entry cost, i.e., iff $\pi(w_E + k) \geq F$. In all that follows, we focus on the case where F is not too high,

$$F < \pi(k) =: \tilde{F}(k), \tag{IV.1}$$

such that there are positive gains from trade to be realized between U and firm E. Nevertheless, the entry constraint may impose a restriction on U in its setting of wholesale prices.

3.1. Optimal Wholesale Pricing

First, suppose that F is sufficiently low such that the entry constraint is not binding. If price discrimination is permitted, the optimal wholesale price for U to charge from firm i is

$$w^d(k_i) := \arg\max_w \{wq(w + k_i)\}. \tag{IV.2}$$

Under uniform pricing, U chooses the common wholesale price

$$w^u(k) := \arg\max_w \{wq(w + k) + wq(w)\}. \tag{IV.3}$$

We now can establish the following result: if U is unrestricted in its choice of wholesale prices under both regimes, then the optimal uniform wholesale price is bracketed by the two discriminatory prices. More precicely, under discrimination the less efficient firm receives a discount compared to uniform pricing. This discount, however, does not outweigh its cost disadvantage.

Lemma IV.1: *Given (A1) and (A2), then* $w^d(k) < w^u(k) < w^d(0) < w^d(k) + k$.

Under either pricing regime, if the entry fee is high, then charging the optimal unrestricted wholesale price(s) leads to a violation of firm E's entry constraint. Letting $\bar{F}^j(k)$ denote the highest value the entry fee can take such that U is not restricted in its price setting under pricing regime $j \in \{d, u\}$, we have

$$\bar{F}^u(k) := \pi(w^u(k) + k) \text{ and } \bar{F}^d(k) := \pi(w^d(k) + k). \tag{IV.4}$$

From Lemma IV.1 and (IV.1) it follows immediately that $\bar{F}^u(k) < \bar{F}^d(k) < \tilde{F}(k)$.

In order to induce entry, U optimally charges firm E wholesale price w^R at which firm E is indifferent between entering and staying out of the intermediate industry. Wholesale price w^R is implicitly defined by

$$\pi(w^R + k) \equiv F. \tag{IV.5}$$

Obviously, $w^R = w^R(F; k)$ is strictly decreasing in F.

Under price discrimination it is optimal to offer wholesale price w^R to firm E as long as positive gains from trade are to be realized between U and firm E, i.e., for $F < \tilde{F}(k)$. When restricted to a uniform wholesale price, U has to pass-through this discount price w^R also to firm I. If the entry fee only slightly exceeds $\bar{F}^u(k)$, it remains optimal for U to serve both downstream firms just as in the case where the entry fee does not restrict wholesale pricing. Since w^R is strictly decreasing in F, if the entry fee exceeds some critical threshold $\hat{F}$, U prefers serving only firm I at wholesale price $w^d(0)$. Formally, $\hat{F}$ is implicitly defined by

$$w^R(\hat{F}; k)[q(w^R(\hat{F}; k) + k) + q(w^R(\hat{F}; k))] = w^d(0)q(w^d(0)). \tag{IV.6}$$

Obviously, $\hat{F} = \hat{F}(k)$. From $w^R(\bar{F}^u(k); k) \equiv w^u(k)$ together with $w^R(F; k)$ tending to zero as F tends to $\tilde{F}(k)$, it follows that $\bar{F}^u(k) < \hat{F}(k) < \tilde{F}(k)$.

Letting wholesale prices in equilibrium be denoted by $\{w_E^d, w_I^d\}$ and w^u under price discrimination and uniform pricing, respectively, we summarize the above discussion as follows:

Observation 1: *In equilibrium, the optimal wholesale price(s)*

(i) under price discrimination are $w_E^d = w^d(k)$ *for* $0 < F \leq \bar{F}^d(k)$, $w_E^d = w^R(F;k)$ *for* $\bar{F}^d(k) < F < \tilde{F}(k)$, *and* $w_I^d = w^d(0)$.

(ii) under uniform pricing is $w^u = w^u(k)$ *for* $0 < F \leq \bar{F}^u(k)$, $w^u = w^R(F;k)$ *for* $\bar{F}^u(k) < F < \hat{F}(k)$, *and* $w^u = w^d(0)$ *for* $\hat{F}(k) \leq F < \tilde{F}(k)$.

3.2. Welfare Implications of Banning Price Discrimination

The measure of total welfare applied in this paper is the unweighted sum of consumer and producer surplus. We express changes in economic variables due to a regime shift from uniform pricing to price discrimination using symbol Δ. If both firms are active in the downstream industry, then the change in total welfare due to a regime shift amounts to

$$\Delta W \equiv \int_{q(w^u+k)}^{q(w_E^d+k)} P(x)dx - \int_{q(w_I^d)}^{q(w^u)} P(x)dx - k[q(w_E^d+k) - q(w^u+k)]. \qquad \text{(IV.7)}$$

To compare the two pricing regimes one needs a complete ordering of the entry cost's critical threshold levels. With the relationship between $\bar{F}^d(k)$ and $\hat{F}(k)$ being undetermined in general, there are two possible orderings of the thresholds as depicted in Figure IV.1.

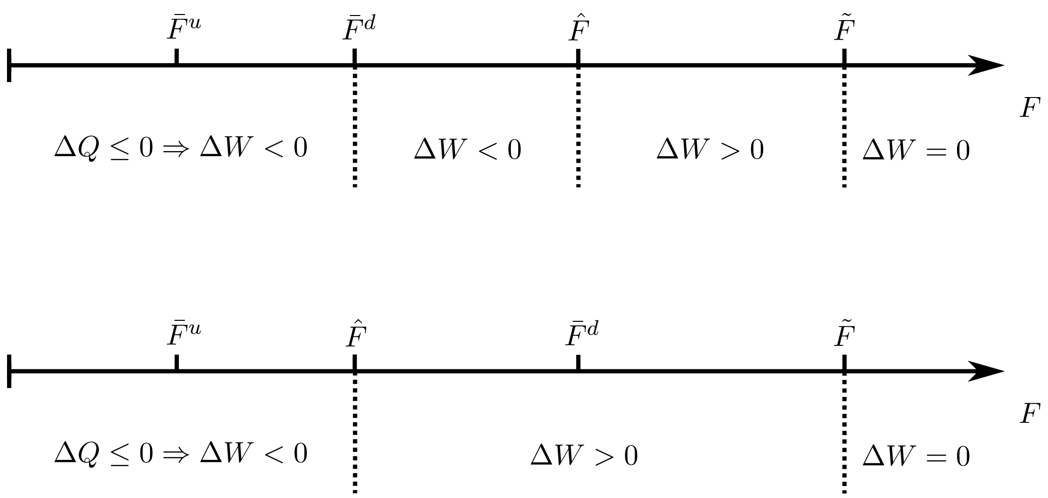

Figure IV.1.: Welfare comparison with separate markets and a less efficient entrant.

It is important to note that ceteris paribus entry into the downstream industry is always beneficial from a welfare point of view, since E enters only if it generates a surplus that exceeds the entry cost. Letting Q^r denote the total quantity sold under pricing regime $r \in \{d, u\}$, the following welfare implications are readily obtained.

Proposition IV.1: *Given* $(A1)$ *and* $(A2)$*, if*

(i) $F < \min\{\hat{F}(k), \bar{F}^d(k)\}$*, then* $\Delta Q \leq 0$ *implies* $\Delta W < 0$*;*

(ii) $\bar{F}^d(k) \leq F < \hat{F}(k)$*, then* $\Delta W < 0$*;*

(iii) $\hat{F}(k) \leq F < \tilde{F}(k)$*, then* $\Delta W > 0$.

Proposition IV.1 is illustrated in Figure IV.1. Since a situation with downstream firms operating in independent markets has a flavor of a final-good monopoly, we next relate our findings to the existant literature on monopolistic third-degree price discrimination in final-good markets.

For low values of the entry fee, in case (i), entry occurs under both pricing regimes. With the uniform wholesale price lying strictly between the two discriminatory wholesale prices, a clear welfare result is not to obtain. Nevertheless, we can derive a sufficient condition for permitting price discrimination to reduce welfare: if total output under price discrimination is not higher than total outout under uniform pricing, then welfare under price discrimination is strictly lower than welfare under uniform pricing. This finding clearly paralells Schmalensee's (1981) result on third-degree price discrimination in final-good markets: while in general price discrimination may be both welfare increasing as well as welfare decreasing, a necessary condition for price discrimination to improve welfare is that it leads to an expansion of aggregate output.[10]

For high values of the entry fee, i.e., in case (iii), price discrimination fosters entry which in turn improves welfare. With entry occuring only if discriminatory pricing is permitted, the market outcome and thus welfare in the incumbent market is independent of the pricing regime, whereas welfare in the entrant's market is strictly positive only under price discrimination. This obviously is the intermediate-market analogue to Hausman and MacKie-Mason's (1988) finding on third-degree price discrimination in final-good markets: here, with opening of a market assumed to be costless, price discrimination yields a better welfare outcome if it results in opening of a new market while uniform pricing does not.

Case (ii), on the other hand, embodies a novelty with regard to the extant literature on monopolistic third-degree price-discrimination in final-good markets, which

[10] A series of papers elaborates on Schmalensee's basic insight, see Varian (1981), Schwartz (1990), and Malueg (1993).

abstracts from opening of new markets being costly. For intermediate values of the entry fee both downstream firms are served under either pricing regime. With the upstream supplier being restricted in its price setting under both pricing regimes, the entrant receives wholesale price w^R irrespectively of the regime. This low wholesale price is passed on to the incumbent firm only under uniform pricing but not under price discrimination. In consequence, welfare in the entrant's market is unchanged when permitting price discrimination but welfare in the incumbent's market is strictly reduced. Thus, even though market-opening occurs under both pricing regimes, price discrimination is unambiguously found to be detrimental for welfare.[11]

We are fairly optimistic that results analogous to Proposition IV.1—as well as Proposition IV.2 in the next section—can also be derived in a model of monopolistic third-degree price discrimination in final-good markets. Note, however, that for the effect in case (ii) of Proposition IV.1 to arise it must not be the final-good monopolist who bears the cost of becoming active in a new market but the consumers of this potentially opened market. To the best of our knowledge, such a situation has not been considered in the existant literature on third-degree price discrimination in final-good markets.[12] Nevertheless, we refrain from framing our analysis in terms of a final-good monopoly for at least two reasons. First, while it is natural to imagine a downstream firm to bear the cost of opening a new market in the form of purchasing a franchise fee or setting up a production site, most stories one could tell for a final consumers seem somewhat contrieved.[13] Second, with most of the principal legislations regarding price discrimination—like the Robinson-Patman Act in the US and Article 82 of the European Treaty—being primarily concerned about input price discrimination, inquiring into the welfare effects of price discrimination in intermediate-good markets seems of greater practical relevance. As was already pointed out by Katz (1987), welfare results regarding price discrimination obtained in final-good markets do not necessarily carry over to intermediate-good markets due to different market characteristics. Though most likely of little loss in the above situation with independent downstream markets,

[11] Note that case (ii) exists only if $\bar{F}^d(k) < \hat{F}(k)$. The effect arising in case (ii) depends on the upstream supplier's ability to commit to its offers. If commitment is not possible and the entry decision is made before wholesale prices are set, then for $F > \bar{F}^d(k)$ entry occurs under neither pricing regime.

[12] For extensive overviews on price discrimination in final-good markets, see Armstrong (2007) and Stole (2007).

[13] For example, one might think of a monopolistic seller of dvds who faces two groups of distinguishable consumers, students and elderly people. While students most likely possess a dvd player in one form or the other—regular dvd player, home computer, laptop, Playstation 3—eldery people can easily be imagined not to possess these devices. In this sense, in order for the market consisting of elderly people to be opened, each senior has to incur the fixed cost of purchasing a dvd player.

at the latest when downstream competition is considered, as we do in Section 5, caution should be exercised when infering validity of welfare statements from final-good markets in intermediate-good markets. Therefore, we place our analysis in the realms of intermediate-good market price discrimination right away from the beginning.

A Linear Demand Application: Suppose $P(q) = \max\{1 - q, 0\}$, which satisfies (A1), and $0 < k < 1$, which relaxes (A2). The profit of an active downstream firm is $\pi(c_i) = (1 - c_i)^2/4$. If U is unrestricted by the entry constraint, the optimal wholesale prices are $w^d(k_i) = (1 - k_i)/2$ and $w^u(k) = (2 - k)/4$ under price discrimination and under uniform pricing, respectively. The wholesale price that makes firm E indifferent between entering and staying out is $w^R = 1 - k - 2\sqrt{F}$. It can be shown that $\hat{F}(k) > \bar{F}^d(k)$ if and only if $k > 1/2$, where $\hat{F}(k) = (1/64)[2 - 3k + 4\sqrt{k^2 - 4k + 2}]^2$. Since we have not imposed (A2) there exists a critical marginal cost $\bar{k}$ such that under uniform pricing U optimally serves only firm I if $k > \bar{k}$ even for $F = 0$.

Figure IV.2 depicts the critical thresholds for firm E's marginal cost k and the entry cost F, where for illustrative purposes we rephrased the thresholds in terms of $\sqrt{F}$. As is well-known, with linear demand total output is the same under price discrimination and uniform pricing, given that the entry constraint does not impose a binding restriction. Hence, according to Proposition IV.1(i), for low values of F banning price discrimination improves welfare. This case corresponds to the white area on the left bottom of Figure IV.2. The dark gray shaded area of Figure IV.2 corresponds to case (ii) of Proposition IV.1. Here, permitting price discrimination is harmful for total welfare. On the other hand, in the light gray shaded area price discrimination encourages entry which in turn supports welfare, case (iii) of Proposition IV.1.

4. More Efficient Entrant

Suppose the entrant is more efficient than the incumbent, $0 = k_E < k_I = k$. Otherwise the model is the same as before: in particular $(A1)$ and $(A2)$ hold and $F < \pi(0) =: \tilde{F}(0)$. Lemma IV.1 immediately implies that the unrestricted uniform wholesale price is bracketed by the two unrestricted discriminatory wholesale prices, with the less efficient firm I receiving a discount under price discrimination.

If discriminatory offers are allowed, U charges wholesale price $w^d(k)$ from firm I. The wholesale price offered to firm E depends on whether the entry constraint imposes a binding restriction. If the entry constraint is slack, U sets $w^d_E = w^d(0)$. If the entry fee exceeds $\bar{F}^d(0) := \pi(w^d(0))$, the entry constraint becomes binding and U charges $w^d_E = w^R(F; 0)$ implicitely defined by $\pi(w^R(F; 0)) = F$.

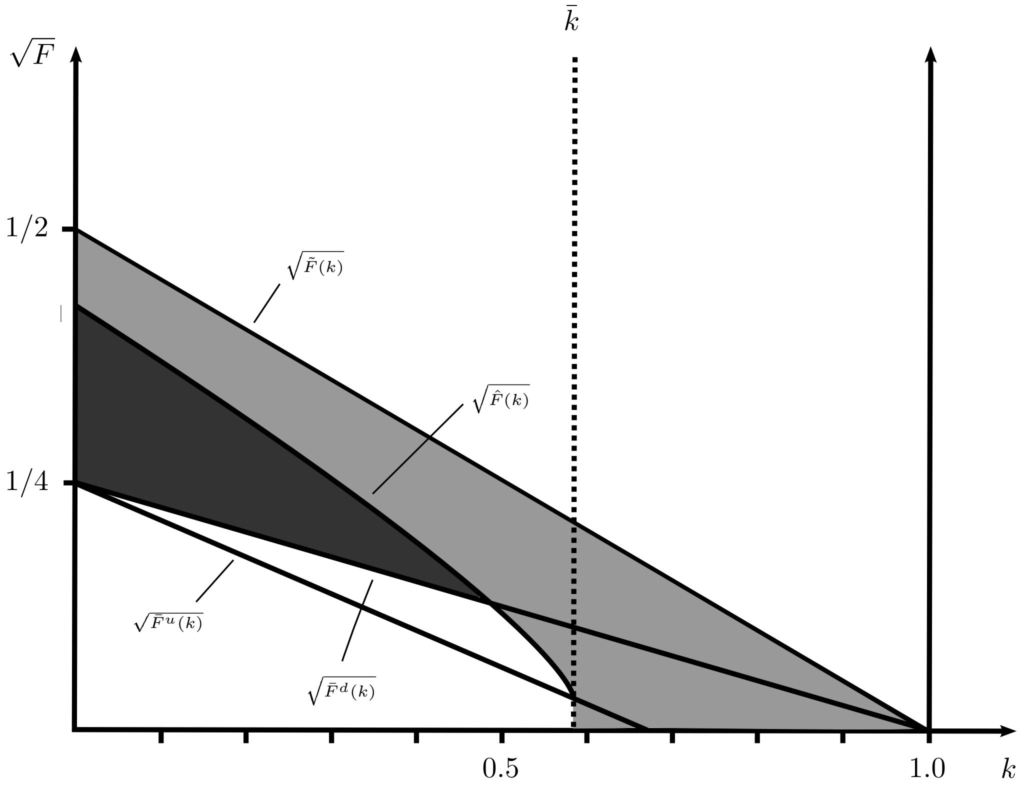

Figure IV.2.: Welfare comparison with separate markets and linear demand.

The optimal uniform wholesale price is $w^u(k)$ if the entry cost is low. If the entry cost exceeds $\bar{\bar{F}}^u(k) := \pi(w^u(k))$, then U is restricted when choosing a uniform wholesale price. Note that $\bar{F}^d(0) < \bar{\bar{F}}^u(k)$. For intermediate values of the entry fee the optimal uniform wholesale price is $w^R(F;0)$ which makes firm E just willing to enter the industry. For sufficiently high entry cost, $F \geq \hat{\hat{F}}$, U prefers to serve only firm I at price $w^d(k)$. The critical entry fee $\hat{\hat{F}}$ is implicitly defined by

$$w^R(\hat{\hat{F}};0)\Big[q(w^R(\hat{\hat{F}};0)+k)+q(w^R(\hat{\hat{F}};0))\Big] \equiv w^d(k)q(w^d(k)+k), \qquad \text{(IV.8)}$$

where $\hat{\hat{F}} = \hat{\hat{F}}(k) < \tilde{F}(0)$.

There exists an additional threshold for the entry cost that turns out to be important to characterize the welfare implications of banning price discrimination. Since firm I receives a discount under price discrimination, there exists an entry cost $\check{F}(k) \in (\bar{\bar{F}}^u(k), \hat{\hat{F}}(k))$, at which the restricted wholesale price equals the discriminatory wholesale price of firm I, i.e., $w^R(\check{F}(k);0) \equiv w^d(k)$.[14] For entry costs slightly below $\check{F}(k)$,

[14]Note that $\bar{\bar{F}}^u(k) < \check{F}(k)$ follows immediately from $w^R(\check{F}(k);0) = w^d(k) < w^u(k) = w^R(\bar{\bar{F}}^u(k);0)$.

price discrimination leads to (weakly) lower wholesale prices for both firms compared to uniform pricing. For F slightly above $\check{F}(k)$, on the other hand, the uniform wholesale price is (weakly) below both discriminatory prices.

From the above observations, the following welfare implications follow immediately.

Proposition IV.2: *Suppose* $0 = k_E < k_I = k$. *Given* $(A1)$ *and* $(A2)$, *if*

(i) $F < \bar{\bar{F}}^u(k)$, *then* $\Delta Q \leq 0$ *implies* $\Delta W < 0$;

(ii) $\bar{\bar{F}}^u(k) \leq F < \check{F}(k)$, *then* $\Delta W > 0$;

(iii) $\check{F}(k) < F < \hat{\hat{F}}(k)$, *then* $\Delta W < 0$;

(iv) $\hat{\hat{F}}(k) \leq F < \tilde{F}(k)$, *then* $\Delta W > 0$.

Cases (i), (iii), and (iv) are basically known from the previous analysis of a less efficient entrant. For very low values of the entry cost, in case (i), with the uniform wholesale price being bracketed by the two discriminatory wholesale prices, a clear welfare result is intricate to obtain. Nevertheless, if price discrimination does not lead to an expansion of total output, then permitting price discrimination is harmful for social welfare. For high-intermediate values of the entry cost, in case (iii), banning price discrimination unambiguously improves welfare. Under both pricing regimes, the upstream supplier serves both firms. The low restricted wholesale price which is necessary to induce entry is passed on to the incumbent only under uniform pricing, however. In case (iv), if entry is very costly, only price discrimination leads to the opening of the new market. Thus, for high values of the entry cost the known entry-promoting and in turn welfare-improving effect of permitting price discrimination prevails.

An interesting novelty is found in case (ii). Here, for low-intermediate values of the entry cost, under both pricing regimes the upstream supplier is restricted in its price setting but nevertheless induces entry. Surprisingly, a discriminatory pricing regime leads to (weakly) lower wholesale prices for both downstream firms. In consequence, permitting price discrimination strictly increases welfare, even though it does not lead to more markets being served than under a ban of price discrimination.

5. Downstream Competition

In this section, we inquire into the implications of downstream competition for the welfare effects associated with a ban of discriminatory wholesale pricing.[15] We now

$\check{F}(k) < \hat{\hat{F}}(k)$, on the other hand, follows from the fact that it is profitable to serve both downstream firms at a price only slightly below $w^d(k)$ instead of serving only firm I at price $w^d(k)$.

[15] A detailed account of the following discussion is found in Appendix B.

assume that active downstream firms produce a homogeneous final good and compete in quantities. Thus, if firm E, which is assumed to be the less efficient downstream firm, becomes active in the downstream industry, this is not associated with opening of a new market but with entry into firm I's market. Except for firms competing à la Cournot in stage 3, we stick to the sequence of events introduced in Section 2. Without further assumptions on the demand function welfare results are hard to obtain with downstream competition. Therefore, we focus on linear demand, i.e., $P(q) = \max\{1 - q, 0\}$. Moreover, we assume $0 < k < 1/2$ and focus on the case where $0 < \sqrt{F} < 1/3 - (2/3)k =: \tilde{f}(k)$. While the first assumption guarantees that both downstream firms produce positive quantities at the optimal unrestricted uniform wholesale price, the latter rules out the case where U prefers to serve only firm I under both pricing regimes.

Before proceeding with the analysis, a remark regarding the upstream supplier's incentives to serve the inefficient entrant next to the incumbent is in order: being restricted to linear wholesale contracts, the manufacturer's interest in inducing entry and thereby promoting downstream competition arises from the desire to reduce double marginalization. If the input supplier nevertheless prefers to serve only one downstream firm in equilibrium, then this is always the incumbent firm at wholesale price $w_M = 1/2$.

Suppose both downstream firms are active in equilibrium. Given the rival's effective marginal cost c_j, downstream firm i with effective marginal cost c_i demands quantity $q(c_i, c_j) = (1/3)(1 - 2c_i - c_j)$ and realizes gross profits $\pi(c_i, c_j) = (1/9)(1 - 2c_i - c_j)^2$. Similar as before, for low values of the entry cost, firm E's entry constraint does not impose a binding restriction on U's choice of wholesale prices. In this case, the optimal wholesale prices under uniform pricing and under price discrimination are $w^u(k) = (1/4)(2 - k)$ and $w^d(k_i) = (1/2)(1 - k_i)$, respectively. In consequence, the entry constraint does not impose a binding restriction under uniform pricing and under price discrimination if

$$\pi(w^u(k) + k, w^u(k)) \geq F \iff \sqrt{F} \leq \frac{1}{6} - \frac{7}{12}k =: \bar{f}^u(k) \tag{IV.9}$$

and

$$\pi(w^d(k) + k, w^d(0)) \geq F \iff \sqrt{F} \leq \frac{1}{6} - \frac{1}{3}k =: \bar{f}^d(k), \tag{IV.10}$$

respectively. Note that $\bar{f}^u(k) < \bar{f}^d(k)$, i.e., the entry constraint imposes a stronger restriction on U under uniform pricing than under price discrimination since the less efficient firm receives a discount if price discrimination is permitted.

For higher values of the entry cost, $\sqrt{F} > \bar{f}^r(k)$ with $r \in \{d, u\}$, to make firm E enter the downstream market U needs to offer a discount wholesale price such that firm

E can just recover its fixed cost. With competition firm E's profit does not only depend on its own wholesale price but also on firm I's wholesale price. Thus, in contrast to the case with separate downstream markets, the restricted wholesale price is not necessarily identical under the two pricing regimes. The restricted uniform wholesale price, w^{Ru}, is defined by $\pi(w^{Ru}+k, w^{Ru}) \equiv F$, or equivalently, $w^{Ru}(\sqrt{F};k) = 1-2k-3\sqrt{F}$. Under price discrimination, on the other hand, U chooses two wholesale prices, and thus the restricted wholesale price is not pinned down by firm E's binding entry constraint alone. Here, U optimally offers wholesale price $w_I^R = 1/2$ and $w_E^R(\sqrt{F};k) = 3/4-k-(3/2)\sqrt{F}$ to firm I and firm E, respectively.

Is it always in U's interest to serve both downstream firms? Under discriminatory pricing it can be shown that U prefers to implement a downstream duopoly if $\sqrt{F} < \tilde{f}(k)$. If U is forced to offer a uniform wholesale price, it prefers that firm I monopolizes the downstream market when firm E is very inefficient or when entry costs are too high. Formally, U optimally serves both downstream firms if $\sqrt{F} < \hat{f}(k)$, where

$$\hat{f}(k) := \begin{cases} (1/12)[2-7k+\sqrt{1-4k+k^2}] & ,\text{for } k < 2-\sqrt{3} \\ 0 & ,\text{for } k \geq 2-\sqrt{3} \end{cases} . \tag{IV.11}$$

Note that $\hat{f}(k) < \tilde{f}(k)$. Hence, there exists a range of entry costs where entry occurs under price discrimination but not under uniform pricing, i.e., price discrimination promotes entry also when downstream firms compete. The thresholds characterized above are depicted in Figure IV.3.

Welfare Comparisons In order to compare welfare under the two pricing regimes, we distinguish five cases with respect to the resulting downstream market structure, as illustrated in Figure IV.3. We label these cases with Roman numerals, I - V.

In cases I - III, both pricing regimes lead to implementation of a downstream duopoly. Moreover, in these three cases allowing for discriminatory wholesale prices lowers welfare. This is most obvious in case I, where both the entry fee and the entrant's marginal cost of production are sufficiently low such that the input supplier is not constrained in its choice of wholesale prices under either pricing regime.[16] While total output is unaffected by the pricing regime, which is a direct result of linear demand, under price discrimination the upstream supplier "subsidizes" the less efficient firm by charging a higher wholesale price to the more efficient firm, thereby (at least partly) removing the incumbent firm's cost advantage. In consequence, under price discrimination output is shifted from the low-cost firm to the high-cost firm, which raises the total cost of production and thus reduces welfare. This negative effect of price discrimination on

[16]This situation exactly corresponds to the short-run analysis in DeGraba (1990).

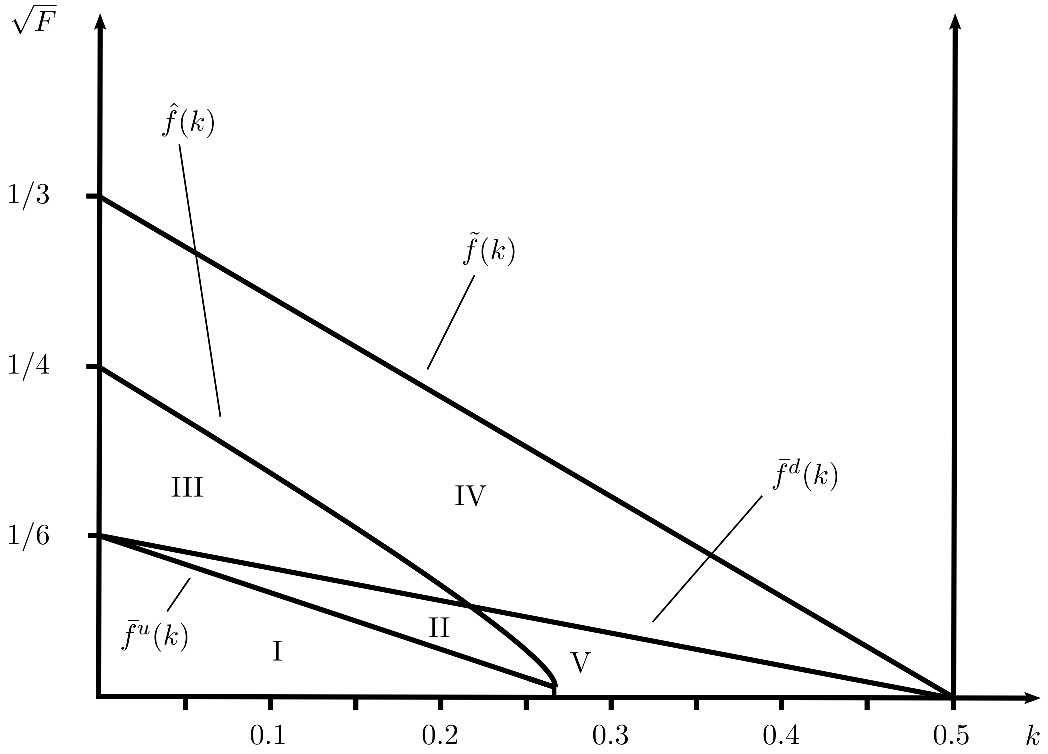

Figure IV.3.: Downstream market structure with competition.

the allocation of producing the final output is even more severe if the upstream firm is restricted by firm E's entry decision, since this increases the discount the less efficient entrant receives. If the entry constraint imposes a restriction under uniform pricing, on the other hand, the lowered wholesale price applies for all downstream firms, such that no such misallocation in production shares occurs.[17]

The more interesting cases are IV and V. Here, the downstream market is monopolized under uniform pricing while under price discrimination both downstream firms compete for final customers. The reason is that the relatively high entry fee and/or the relatively high marginal cost of the entrant render the concession in the uniform wholesale price necessary to induce entry unprofitable for the upstream monopolist. Price discrimination, on the other hand, provides the input supplier with a tool to profitably implement a downstream duopoly even in these cases.

With separate markets, entry into the intermediate industry taking place only under a discriminatory pricing regime but not under uniform pricing is a sufficient condition

[17]Moreover, due to a lower input price, total output is increased, which in turn improves welfare compared to the situation where the upstream firm is unrestricted under uniform pricing.

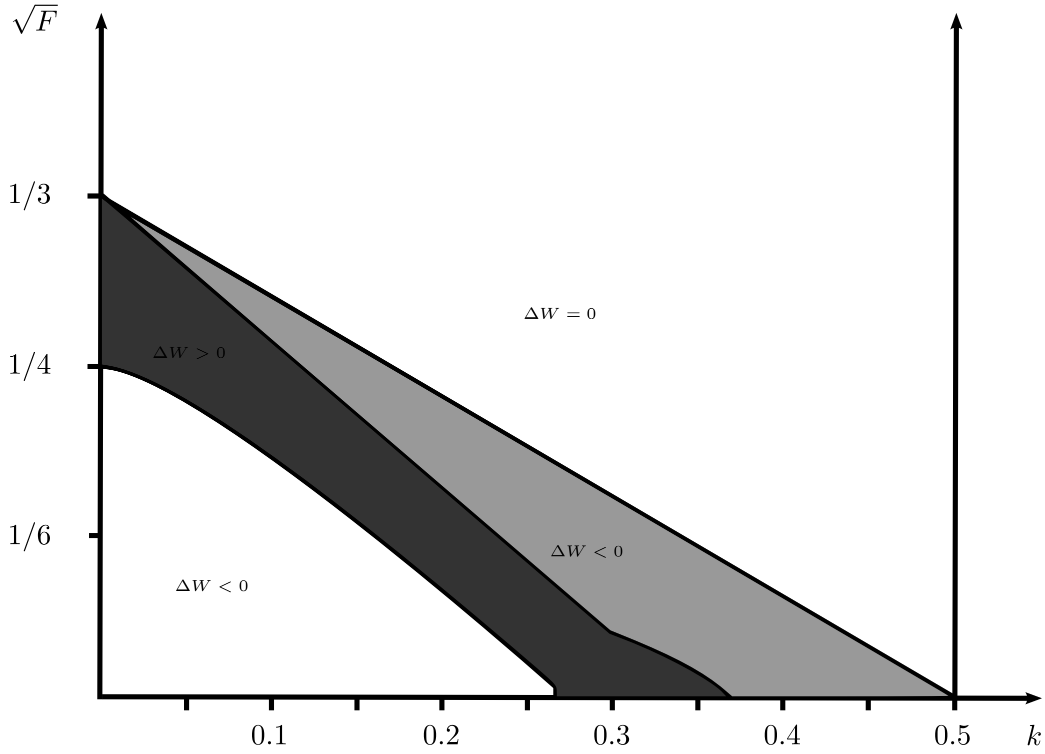

Figure IV.4.: Welfare comparison with downstream competition.

for welfare to be higher under price discrimination, see Proposition IV.1. For moderate values of the entry cost and a not too inefficient entrant—represented by the dark-gray shaded area in Figure IV.4—this result carries over to the case of downstream competition: here, society benefits from output being produced inefficiently rather than not being produced at all. If, however, the entry fee is high and/or the entrant is very inefficient—represented by the light-gray shaded area in Figure IV.4—entry becomes undesirable from a social perspective: while entry into the downstream market alleviates the quantity distortion arising under downstream monopoly, thereby increasing upstream profits and benefiting consumers through higher quantity and lower final-good prices, the increase in aggregate output comes at the cost of a reduction in the efficient downstream firm's output brought about by competition in the downstream market. Thus, the increase in consumer surplus and the upstream supplier's profits is gained at the price of burdening society with the cost of entry and higher production costs. In consequence, since the major effect of the discriminatory pricing policy is not the creation of value but shifting rents away from the incumbent firm to the upstream supplier, here price discrimination leads to a strictly inferior welfare result even though market entry is promoted.

With regard to a formal welfare result, we define

$$f^W(k) := \begin{cases} 1/3 - (8/9)k & ,\text{for } k \leq 3/10 \\ \sqrt{(23/72)k^2 - (5/18)k + 17/288} & ,\text{for } 3/10 < k \leq 17/46 \\ 0 & ,\text{for } k > 17/46 \end{cases} . \quad \text{(IV.12)}$$

With this notation we are prepared summarize the above discussion as follows:

Proposition IV.3: *For the case with linear demand and downstream Cournot competition; (i)* $\Delta W > 0$ *if* $\hat{f}(k) \leq \sqrt{F} < f^W(k)$ *and (ii)* $\Delta W < 0$ *if either* $\sqrt{F} < \hat{f}(k)$ *or* $f^W(k) < \sqrt{F} < \tilde{f}(k)$.

We refrain from an analysis of downstream competition with a more efficient entrant. In this case, if entry occurs under price discrimination but not under uniform pricing, production shares are shifted from the less efficient to the more efficient downstream firm, i.e., in tendency overall production becomes less costly. Therefore, in these cases, we would expect a discriminatory pricing regime to be welfare improving more often, because one major inefficiency identified in the analysis of a less efficient entrant does not arise.

6. Conclusion

This paper attempts to provide answers to the following two questions: First, how does potential entry into the downstream industry affect wholesale prices set by an upstream monopolist? Second, under what circumstances is banning third-degree price discrimination beneficial for welfare and consumer surplus if there is potential entry into the downstream sector?

Compared to a situation with a rigid structure of the intermediate industry, the optimal uniform wholesale price as well as the optimal discriminatory wholesale price charged from the potential entrant may be lower if costly entry is possible. The optimal wholesale price charged from incumbent firms, in contrast, does not depend on whether entry into the intermediate industry is possible or not. As a consequence, when downstream firms operate in distinct markets, there are situations—in terms of the entrant's efficiency in production and the cost of entry—where price discrimination may lead to either higher or lower prices for all downstream firms than uniform pricing. In these cases, with wholesale prices being clearly favorable under one of the two pricing regimes, we obtain unambiguous implications of banning price discrimination regarding welfare and consumer surplus. If downstream firms are Cournot competitors, permitting price discrimination has the beneficial effect that it supports entry which in turn reduces double marginalization. This beneficial effect, however, can be outweighed by entry being costly and an allocative inefficiency in production induced by discrimination against the more efficient firm. With costly entry being possible in our model, these results are novel to the extant literature on third-degree price discrimination in intermediate-good markets.

A. Appendices

1. Appendix to Chapter I

1.1. Proofs of Propositions and Lemmas

Proof of Proposition I.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,

$$g'(\hat{e}_1 + e_2^{TC}(\hat{e}_1)) = c'(e_2^{TC}(\hat{e}_1)) \ . \tag{A.1}$$

Differentiating (A.1) 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:

$$\beta g'\left(e_1^N + e_2^{TC}(e_1^N)\right) = c'(e_1^N) \ . \tag{A.2}$$

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

$$\beta g'\left(e_1^N + e_2^N\right) = c'(e_2^N) \ . \tag{A.3}$$

Comparison of (A.1)-(A.3) allows to establish the proposition. We prove each part of the proposition in turn.

(iii) Comparison of (A.1) and (A.3) 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. ■

Proof of Proposition I.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 (I.5) and (I.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. ■

Proof of Lemma I.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 I.1 and I.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 (I.5). Together with $de_2^S(e_1)/de_1 \in (-1, 0)$, the desired result follows. ■

Proof of Lemma I.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:

$$\begin{aligned} &- c(e_1^N) - \beta c(e_2^{TC}(e_1^N)) + \beta g(e_1^N + e_2^{TC}(e_1^N)) \\ &\qquad\qquad \geq -c(e_1^S) - \beta c(e_2^{TC}(e_1^S)) + \beta g(e_1^S + e_2^{TC}(e_1^S)) \end{aligned}$$

and

$$\begin{aligned} &- c(e_1^S) - \beta c(e_2^S(e_1^S)) + \beta g(e_1^S + e_2^S(e_1^S)) \\ &\qquad\qquad \geq -c(e_1^N) - \beta c(e_2^S(e_1^N)) + \beta g(e_1^N + e_2^S(e_1^N)) \end{aligned}$$

Taken together these two inequalities imply

$$\begin{aligned} &\left[g(e_1^N + e_2^{TC}(e_1^N)) - c(e_2^{TC}(e_1^N))\right] - \left[g(e_1^N + e_2^S(e_1^N)) - c(e_2^S(e_1^N))\right] \\ &\qquad \geq \left[g(e_1^S + e_2^{TC}(e_1^S)) - c(e_2^{TC}(e_1^S))\right] - \left[g(e_1^S + e_2^S(e_1^S)) - c(e_2^S(e_1^S))\right] . \quad \text{(A.4)} \end{aligned}$$

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.4) 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.1) and (I.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,

$$\frac{dF(e_1)}{de_1} = \left[g'(e_1 + e_2^{TC}(e_1)) - g'(e_1 + e_2^S(e_1))\right] - (1-\beta) g'(e_1 + e_2^S(e_1)) \frac{de_2^S(e_1)}{de_1} . \quad \text{(A.5)}$$

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 sohisticate 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

$$\left[g'(e_1^N + e_2^S(e_1^N)) - g'(e_1^N + e_2^{TC}(e_1^N))\right] - (1-\beta)g'(e_1^N + e_2^S(e_1^N))\frac{de_2^S(e_1)}{de_1} = 0. \quad \text{(A.6)}$$

Note that the left-hand side of (A.6) 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. ■

Proof of Proposition I.3: Follows immediately from Lemmas I.1 and I.2. ■

Proof of Proposition I.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 explicitely 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

$$\beta g'(\alpha_1 e_1 + e_2^S(e_1, \alpha_1)) = c'(e_2^S(e_1, \alpha_1)) \ . \tag{A.7}$$

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.7) 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

$$-c'(e_1) + \alpha_1 \beta g'(\alpha_1 e_1 + e_2^S(e_1, \alpha_1)) + (1 - \alpha_1)\beta g'((1 - \alpha_1)e_1) > 0 \, . \tag{A.8}$$

Combining (A.7) and (A.8) yields

$$c'(e_2^S(e_1, \alpha_1)) - c'(e_1) - (1 - \alpha_1)\beta \left[g'(\alpha_1 e_1 + e_2^S(e_1, \alpha_1)) - g'((1 - \alpha_1)e_1)\right] > 0 \, . \tag{A.9}$$

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.9) 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 bacause $e_A < e_B$ 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 $\partial U_1(e_1, \alpha_1)/\partial e_1|_{e1=0} > 0$), we are left with $e_1 \in (0, \infty)$ and $\alpha_1 = 0$. For $\alpha_1 = 0$, second-period behavior is characterized by

$$\beta g'(e_2) = c'(e_2). \tag{A.10}$$

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

$$\beta g'(e_1) = c'(e_1) \, . \tag{A.11}$$

Taken together, (A.10) and (A.11) contradict the assumption that effort smoothing over tasks is not optimal in period 2, 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 (I.8), i.e.,

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

Differentiation of (I.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 first-order condition given by (I.9),

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

From (I.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 (I.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:

$$\frac{\partial U_1^{SD}}{\partial e_A} = 0 \iff \beta g'(e_A) - c'(e_A + e_{B1}) = 0 \ , \tag{A.12}$$

$$\frac{\partial U_1^{S^D}}{\partial e_{B1}} = 0 \iff \beta g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - c'(e_A + e_{B1}) + \frac{de_{B2}^{S^D}(e_{B1})}{de_{B1}} \beta \left[g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - c'(e_{B2}^{S^D}(e_{B1})) \right] = 0 \,. \quad \text{(A.13)}$$

Combining (A.12) and (A.13) 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 (I.10). This last inequality implies that $e_{B1} + e_{B2}^{S^D}(e_{B1}) < e_A$. From (A.12) it follows that e_A decreases as e_{B1} increases. Comparing (I.10) and (A.12) 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 (I.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

$$\beta g'(0) - c'(e_{B1}) \leq 0, \quad \text{(A.14)}$$

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:

$$\beta g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - c'(e_{B1}) + \frac{de_{B2}^{S^D}(e_{B1})}{e_{B1}} \beta \left[g'(e_{B1} + e_{B2}^{S^D}(e_{B1})) - c'(e_{B2}^{S^D}(e_{B1})) \right] = 0.$$

The last term of the left-hand side of the above equation is negative, which implies that $\beta g'(e_{B1} + e^{SD}_{B2}(e_{B1})) - c'(e_{B1}) > 0$. Taken together with (A.14) this yields $\beta g'(e_{B1} + e^{SD}_{B2}(e_{B1})) > g'(0)$. This in turn implies $e_{B1} + e^{SD}_{B2}(e_{B1}) < 0$, 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 $e^{SND}_1 \geq e^{SD}$. From (I.8) and (I.12) we know, respectively, that $c'(e^{SND}_2(e^{SND}_1)) = \beta g'(\frac{1}{2}(e^{SND}_1 + e^{SND}_2(e^{SND}_1)))$ and $c'(e^{SD}) = \beta g'(\frac{1}{2}(e^{SD} + e^{SD}))$. Since $de^{SND}_2(e^{SND}_1)/de_1 \in (-1, 0)$, $e^{SND}_1 \geq e^{SD}$ implies that $e^{SND}_2(e^{SND}_1) \leq e^{SD}$, which in turn in implies $e^{SND}_1 + e^{SND}_2(e^{SND}_1) \geq 2e^{SD}$. From (I.9), however, we know that $\beta g'(\frac{1}{2}(e^{SND}_1 + e^{SND}_2(e^{SND}_1))) - c'(e^{SND}_1) > 0$. Together with (I.12) this implies that $\beta g'(\frac{1}{2}(e^{SND}_1 + e^{SND}_2(e^{SND}_1))) - \beta g'(\frac{1}{2}(e^{SD} + e^{SD})) > c'(e^{SND}_1) - c'(e^{SD}) \geq 0$, where the last inequality holds by our initial assumption that $e^{SND}_1 \geq e^{SD}$. With $g'(\cdot)$ being strictly decreasing, this last expression implies $e^{SND}_1 + e^{SND}_2 < 2e^{SD}$, a contradiction. Therefore we must have $e^{SND}_1 < e^{SD}$. Together with $e^{SND}_2(e^{SD}) = e^{SD}$, which follows from (I.8) in combination with (I.11) or (I.12), and $de^{SND}_2(e_1)/de_1 \in (-1, 0)$, $e^{SND}_1 < e^{SD}$ immediately implies $e^{SND}_1 + e^{SND}_2(e^{SND}_1) < 2e^{SD}$. It remains to show that a sophisticate indeed is better off under a deadline from a long-run perspective, i.e., $U^{SD}_0 > U^{SND}_0$. 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 $e^{SND}_1 + e^{SND}_2(e^{SND}_1) < 2e^{SD} < 2e^{TC^{ND}}$ we have $U^{SD}_0 = U_0(2e^{SD}, \frac{1}{2}) > U_0(e^{SND}_1 + e^{SND}_2(e^{SND}_1), \frac{1}{2})$. 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, $U^{SND}_0 < U_0(e^{SND}_1 + e^{SND}_2(e^{SND}_1), \frac{1}{2})$, which establishes the desired result. ■

Proof of Proposition I.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:

$$\frac{\partial U_1^{NND}(e_1, e_2, \alpha)}{\partial e_1} = 0 \iff -c'(e_1^{NND}) + \beta g'(\alpha^{NND}(e_1^{NND} + e_2^{TCND}))\alpha^{NND} + \beta g'((1-\alpha^{NND})(e_1^{NND} + e_2^{TCND}))(1-\alpha^{NND}) = 0\,, \quad \text{(A.15)}$$

$$\frac{\partial U_1^{NND}(e_1, e_2, \alpha)}{\partial e_2} = 0 \iff -\beta c'(e_2^{TCND}) + \beta g'(\alpha^{NND}(e_1^{NND} + e_2^{TCND}))\alpha^{NND} + \beta g'((1-\alpha^{NND})(e_1^{NND} + e_2^{TCND}))(1-\alpha^{NND}) = 0\,, \quad \text{(A.16)}$$

$$\frac{\partial U_1^{NND}(e_1, e_2, \alpha)}{\partial \alpha} = 0 \iff \beta g'(\alpha^{NND}(e_1^{NND} + e_2^{TCND})) = \beta g'((1-\alpha^{NND})(e_1^{NND} + e_2^{TCND})). \quad \text{(A.17)}$$

From (A.17) it follows that $\alpha^{NND} = 1/2$. With $\alpha^{NND} = 1/2$, e_1^{NND} and e_2^{TCND} are characterized by

$$\beta g'((1/2)(e_1^{NND} + e_2^{TCND})) = c'(e_1^{NND})\,, \quad \text{(A.18)}$$

$$g'((1/2)(e_1^{NND} + e_2^{TCND})) = c'(e_2^{TCND})\,, \quad \text{(A.19)}$$

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 allocaation as even as possible, all second-period effort is invested in task A and therefore is characterized by

$$\beta g'(e_{A1}^{NND} + e_2^{NND}) = c'(e_2^{NND})\,. \quad \text{(A.20)}$$

From (A.18), we know that

$$\beta g'((1/2)(e_{A1}^{NND} + e_{B1}^{NND} + e_2^{TCND})) = c'(e_{A1}^{NND} + e_{B1}^{NND})\,. \quad \text{(A.21)}$$

Equations (A.20) and (A.21) 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:

$$\beta g'((1/2)(e_1^{NND} + e_2^{NND}) = c'(e_2^{NND}), \tag{A.22}$$

Comparing (A.19) and (A.22) yields $e_2^{NND} < e_2^{TCND}$, which in combination with (A.18) and (A.22) 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

$$g'(e_{B1} + e_2^{TC}(e_{B1})) = c'(e_2^{TC}(e_{B1})). \tag{A.23}$$

The first-order conditions of utility maximization take the following form:

$$\frac{\partial U_1^{ND}}{\partial e_A} = 0 \iff \beta g'(e_A) - c'(e_A + e_{B1}) = 0 \tag{A.24}$$

$$\frac{\partial U_1^{ND}}{\partial e_{B1}} = 0 \iff \beta g'(e_{B1} + e_2^{TC}(e_{B1})) - c'(e_A + e_{B1}) = 0 \tag{A.25}$$

If the above maximization problem has interior solutions, $e_A > 0$ and $e_{B1} > 0$, then these solutions are characterized by (A.24) and (A.25). 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.24), however, it is immediate that e_A is decreasing in e_{B1}. Moreover, comparing (A.23) and (A.24) 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) $e_A^{ND} > 0$ and $e_{B1}^{ND} = 0$ or (ii) $e_A^{ND} = 0$ and $e_{B1}^{ND} > 0$. For (ii) to be the solution to the naive agent's first-period problem, the following conditions have to hold:

$$\beta g'(0) - c'(e_{B1}^{ND}) \leq 0 \tag{A.26}$$

$$\beta g'(e_{B1}^{ND} + e_{B2}^{TC}(e_{B1}^{ND})) - c'(e_{B1}^{ND}) = 0 \tag{A.27}$$

Obviously, for (A.26) and (A.27) to hold simultaneously it is required that $e_{B1}^{ND} + e_2^{TC}(e_{B1}^{ND}) \leq 0$, which can never be the case. Therefore we are left with $e_A^{ND} > 0$ and $e_{B1}^{ND} = 0$, that is, in the first period the agents invests only in task A. This first-period effort is characterized by

$$\beta g'(e_A^{ND}) = c'(e_A^{ND}). \tag{A.28}$$

The second-period utility of a naive agent under a regime with a deadline takes the following form:

$$U_2^{ND} = -c(e_{B2}) + \beta g(e^{ND}) + \beta g(e_{B2}) .$$

The optimal second-period effort then satisfies

$$\beta g'(e_{B2}^{ND}) = c'(e_{B2}^{ND}). \tag{A.29}$$

Comparing (A.28) and (A.29) yields $e_A^{ND} = e_{B2}^{ND}$, 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^{ND}.

To show that a naive agent chooses a higher effort level in the first period when facing a deadline, suppose, in contradiction, that $e_1^{NND} \geq e^{ND}$. Then $\beta g'(e^{ND}) = c'(e^{ND}) \leq c'(e_1^{NND}) = \beta g'(\frac{1}{2}(e_1^{NND} + e_2^{TC}(e_1^{NND}))$, where the first equality holds by (A.28) and the second equality holds by (A.18). 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.28) 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 I.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. ■

1.2. 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

$$\hat{\beta} g'(\hat{e}_1 + e_2^B(\hat{e}_1)) = c'(e_2^B(\hat{e}_1)) \ . \tag{A.30}$$

Differentiating (A.30) 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

$$-c(e_1) - \beta c(e_2^B(e_1)) + \beta g(e_1 + e_2^B(e_1)) \ . \tag{A.31}$$

The corresponding first-order condition for optimality is given by

$$-c'(e_1^P) + \beta g'\left(e_1^P + e_2^B(e_1^P)\right) + \frac{de_2^B(e_1^P)}{de_1} \beta\left[g'\left(e_1^P + e_2^B(e_1^P)\right) - c'(e_2^B(e_1^P))\right] = 0. \tag{A.32}$$

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

$$\beta g'(e_1^P + e_2^P) = c'(e_2^P) \; . \tag{A.33}$$

A comparison of (A.30) and (A.33) 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.30) and (A.32) 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 I.1 and I.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 I.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

$$c'(e_2^{BND}(\hat{e}_1)) = \hat{\beta} g'((1/2)(\hat{e}_1 + e_2^{BND}(\hat{e}_1))). \tag{A.34}$$

The effort level chosen by a partially naive agent in the first period is determined by the following first-order condition,[1]

$$\begin{aligned} &\beta g'\left((1/2)(e_1^{BND} + e_2^{BND}(e_1^{BND}))\right) - c'(e_1^{BND}) \\ &+ \frac{de_2^{BND}(e_1)}{de_1} \beta \left[g'\left((1/2)(e_1^{BND} + e_2^{BND}(e_1^{BND}))\right) - c'(e_2^{BND}(e_1^{BND}))\right] = 0\,. \end{aligned} \tag{A.35}$$

The actual second-period effort, e_2^{PND}, is characterized by

$$c'(e_2^{PND}) = \hat{\beta} g'((1/2)(e_1^{PND} + e_2^{PND})). \tag{A.36}$$

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., $e_{B1}^{PD} = 0$. 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:

$$c'(e_A^{PD}) = \beta g'(e_A^{PD}) \tag{A.37}$$

$$c'(e_{B2}^{PD}) = \beta g'(e_{B2}^{PD}) \tag{A.38}$$

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 I.4 (respectively I.5) for the case of a partially naive agent.

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

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 I.4.

2. Appendix to Chapter II

2.1. Proofs of Propositions and Lemmas

Proof of Lemma II.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

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

Since signals are ordered according to their likelihood ratio, we have $\sum_{s=\hat{s}}^{S}(\gamma_s^H - \gamma_s^L) > 0$ and $\sum_{s=1}^{\hat{s}-1}(\gamma_s^H - \gamma_s^L) < 0$ 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. ■

Proof of Proposition II.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

$$\frac{\partial \mathcal{L}}{\partial u_s} = 0 \iff h'(u_s) = \mu_0 + \mu_1 \frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}, \quad \forall s \in \mathcal{S}. \tag{A.39}$$

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.39) 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.39) 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. ■

Proof of Proposition II.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$^+$):

$$b_s \beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a}) = c'(\hat{a}) \iff b_s = \frac{c'(\hat{a})}{\beta_s(\hat{\boldsymbol{\gamma}}, \lambda, \hat{a})}. \tag{A.40}$$

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

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

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, $\left\{(\gamma_s^H, \gamma_s^L)\right\}_{s=1}^S$, and the agent's degree of loss aversion, λ. Now, fix $\hat{a}$ and $\left\{(\gamma_s^H, \gamma_s^L)\right\}_{s=1}^S$. 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. ■

Proof of Proposition II.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,

$$\frac{dY(P_{\mathcal{B}})}{dP_{\mathcal{B}}} = \frac{1}{P_{\mathcal{B}}^2(1-P_{\mathcal{B}})^2}\left[2P_{\mathcal{B}}^2 + \frac{2-\lambda}{\lambda-1}(2P_{\mathcal{B}}-1)\right] . \tag{A.41}$$

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, $\sum_{s\in\mathcal{B}^*}(\gamma_s^H - \gamma_s^L) < \sum_{s\in\mathcal{B}^*\setminus\{\hat{s}\}}(\gamma_s^H - \gamma_s^L)$ because $\hat{s}$ is a bad signal. Moreover, $Y(\mathcal{B}^*) < Y(\mathcal{B}^*\setminus\{\hat{s}\})$ because $Y(\cdot)$ increases when signals are excluded of $\mathcal{B}^*$. Thus $X(\mathcal{B}^*) < X(\mathcal{B}^*\setminus\{\hat{s}\})$, 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. ■

Proof of Proposition II.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 $(\gamma_s^H - \gamma_s^L)/\gamma_s(\hat{a}) < (\gamma_t^H - \gamma_t^L)/\gamma_t(\hat{a})$. 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 $\bar{\mathcal{B}} \equiv (\mathcal{B}^* \setminus \{s\}) \cup \{t\}$.

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

Rearranging yields

$$\left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right]>$$
$$\left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H-\gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\mathcal{B}^*})}{(\lambda-1)P_{\mathcal{B}^*}(1-P_{\mathcal{B}^*})}-\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right]. \quad \text{(A.42)}$$

With $Y(P_{\mathcal{B}})$ being defined as in the proof of Proposition II.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.42) 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.42) yields

$$\left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right]>$$
$$\left(\sum_{j\in\mathcal{B}^*}(\gamma_j^H-\gamma_j^L)\right)\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}}-2\kappa)}{(\lambda-1)[P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})+\kappa(1-2P_{\bar{\mathcal{B}}})-\kappa^2]}-\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right]. \quad \text{(A.43)}$$

The right-hand side of (A.43) becomes arbitrarily close to zero for $\kappa\to 0$, thus it remains to show that

$$\left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\left[\frac{1-(\lambda-1)(1-2P_{\bar{\mathcal{B}}})}{(\lambda-1)P_{\bar{\mathcal{B}}}(1-P_{\bar{\mathcal{B}}})}\right]>0\,. \quad \text{(A.44)}$$

For (A.44) to hold, we must have $(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)>0$. From the proof of Proposition II.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

$$\left[(\gamma_t^H-\gamma_t^L)-(\gamma_s^H-\gamma_s^L)\right]\geq\gamma_t(\hat{a})\gamma_s(\hat{a})\left[\frac{1}{\gamma_s(\hat{a})}\frac{\gamma_t^H-\gamma_t^L}{\gamma_t(\hat{a})}-\frac{1}{\gamma_s(\hat{a})+\kappa}\frac{\gamma_s^H-\gamma_s^L}{\gamma_s(\hat{a})}\right], \quad \text{(A.45)}$$

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.45) 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.42) 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.42), 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, $\max_{s\in\mathcal{S}\setminus\mathcal{B}^*}\{\gamma_s^H/\gamma_s^L\} < \min_{s\in\mathcal{B}^*}\{\gamma_s^H/\gamma_s^L\} =: K$, which establishes the result. ■

Proof of Proposition II.5:

We first prove part (ii). Consider a feasible partition $\mathcal{B}$. The corresponding bonus to implement $\hat{a}$ is given by

$$b = \frac{c'(\hat{a})}{\sum_{s\in\mathcal{B}}(\gamma_s^H - \gamma_s^L) - (\lambda - 1)\left[\sum_{s\in\mathcal{B}}(\gamma_s^H - \gamma_s^L)\right][1 - 2P_{\mathcal{B}}]}. \tag{A.46}$$

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 for a feasible partition $\sum_{s\in\mathcal{B}}(\gamma_s^H - \gamma_s^L) > 0$, the desired result follows.

To prove part (i), let $\mathcal{B}^+ \equiv \left\{\mathcal{B} \subset \mathcal{S} \mid \sum_{s\in\mathcal{B}}(\gamma_s^H - \gamma_s^L) > 0\right\}$. 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

$$C_{\tilde{\mathcal{B}}} \;=\; \underline{u}_{\tilde{\mathcal{B}}} + b_{\tilde{\mathcal{B}}}P_{\tilde{\mathcal{B}}} \;=\; \bar{u} + c(\hat{a}) + \frac{c'(\hat{a})(\lambda - 1)P_{\tilde{\mathcal{B}}}(1 - P_{\tilde{\mathcal{B}}})}{\left[\sum_{s\in\tilde{\mathcal{B}}}(\gamma_s^H - \gamma_s^L)\right][1 - (\lambda - 1)(1 - 2P_{\tilde{\mathcal{B}}})]}. \tag{A.47}$$

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, $dC_{\mathcal{B}}/d\lambda > 0$ 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. ■

Proof of Proposition II.6:

First consider $b \geq 0$. We divide the analysis for $b \geq 0$ into three subcases.

Case 1 ($a_0 < 0$): 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 $dEU(a)/da$, the bonus has to be chosen according to

$$b^*(\hat{a}) = \frac{k\hat{a}}{(\gamma^H - \gamma^L)\left[2 - \lambda + 2\gamma(\hat{a})(\lambda - 1)\right]}.$$

Since $a_0 < 0$, $b^*(a)$ 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 $dEU(a)/da$ with a nonnegative bonus. By choosing the bonus according to $b^*(\hat{a})$, $\hat{a}$ satisfies, by construction, the first-order condition. Inserting $b^*(\hat{a})$ into $d^2EU(a)/da^2$ 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 nonnegative bonus. To see this, consider bonus

$$b_0 = \frac{k}{2(\gamma^H - \gamma^L)^2(\lambda - 1)}.$$

For $b < b_0$, $dEU(a)/da < 0$ 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 $dEU(a)/da > 0$. 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 $b \leq b_0$, $dEU(a)/da < 0$ 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 $d^2EU(a)/da^2 > 0$, 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 $EU(0)$ exceeds $EU(1)$ 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

$$EU(a) = u + \gamma(a)b^- + \gamma(a)(1-\gamma(a))\lambda b^- + (1-\gamma(a))\gamma(a)(-b^-) - \frac{k}{2}a^2. \tag{A.48}$$

The first derivative with respect to effort,

$$\frac{dEU(a)}{da} = \underbrace{(\gamma^H - \gamma^L)b^- \left[\lambda - 2\gamma(a)(\lambda-1)\right]}_{MB^-(a)} - \underbrace{ka}_{MC(a)},$$

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 nonnegative 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. ■

Proof of Proposition II.7:

Throughout the analysis we restricted attention to nonnegative 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 II.6. Let $a_0(p)$, $a_0^-(p)$, $b^*(p;\hat{a})$, and $u^*(p;\hat{a})$ 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 $\gamma(p,\hat{a})$, $\gamma^L(p)$, and $\gamma^H(p)$. From the proof of Proposition II.6 we know that (i) action $\hat{a}$ is implementable with a nonnegative bonus (negative bonus) if and only if $a_0(p) \leq 0$ ($a_0^-(p) \leq 0$), and (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 nonnegative bonus.

Consider first the case where $a_0^-(p) < 0$. The negative bonus payment satisfying incentive compatibility is given by

$$b^-(p;\hat{a}) = \frac{k\hat{a}}{(\gamma^H(p) - \gamma^L(p))\left[\lambda - 2\gamma(p,\hat{a})(\lambda-1)\right]}.$$

It is easy to verify that the required punishment to implement $\hat{a}$ is larger in absolute value than than the respective nonnegative bonus which is needed to implement $\hat{a}$, that is, $b^*(p;\hat{a}) < |b^-(p;\hat{a})|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$. When punishing the agent with a negative bonus $b^-(p;\hat{a})$, $u^-(p;\hat{a})$ will be chosen to satisfy the corresponding participation constraint with equality, that is,

$$u^-(p;\hat{a}) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^-(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda-1)\right].$$

Remember that, if $\hat{a}$ is implemented with a nonnegative bonus, we have

$$u^*(p;\hat{a}) = \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[2 - \lambda + \gamma(p,\hat{a})(\lambda-1)\right].$$

It follows immediately that the minimum cost of implementing $\hat{a}$ with a nonnegative bonus is lower than the minimum implementation cost with a strictly negative bonus:

$$\begin{aligned}
C^-(p;\hat{a}) &= u^-(p;\hat{a}) + \gamma(p,\hat{a})b^-(p;\hat{a}) \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^-(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda-1) - 1\right] \\
&> \bar{u} + \frac{k}{2}\hat{a}^2 + \gamma(p,\hat{a})b^*(p;\hat{a})\left[\lambda - \gamma(p,\hat{a})(\lambda-1) - 1\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[1 - \lambda + \gamma(p,\hat{a})(\lambda-1)\right] \\
&= \bar{u} + \frac{k}{2}\hat{a}^2 - \gamma(p,\hat{a})b^*(p;\hat{a})\left[2 - \lambda + \gamma(p,\hat{a})(\lambda-1)\right] + \gamma(p,\hat{a})b^*(p;\hat{a}) \\
&= u^*(p;\hat{a}) + \gamma(p,\hat{a})b^*(p;\hat{a}) \\
&= C(p;\hat{a}).
\end{aligned}$$

The same line of argument holds when $a_0^- = 0$: the bonus which satisfies the (IC) is

$$b_0^-(p;\hat{a}) = -\frac{k}{2(\gamma^H(p) - \gamma^L(p))^2(\lambda-1)},$$

and so $b^*(p;\hat{a}) < |b_0^-(p;\hat{a})|$ for all $\hat{a} \in (0,1)$ and all $p \in [0,1)$. ■

Proof of Corollary II.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 II.7. 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 $\hat{\boldsymbol{\zeta}}$. ■

2.2. 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* $EU'(\hat{a}) = 0$.

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, for a given contract with the property $b_s > 0$ only if $\beta_s > 0$, we show that 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 twice continuously differentiable 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

$$EU''(a) = -2(\lambda - 1) \sum_{s=2}^{S} b_s \sigma_s - c''(a) \ , \tag{A.49}$$

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

$$\sum_{s=2}^{S} b_s \beta_s = c'(\hat{a}) \iff \sum_{s=2}^{S} b_s \frac{\beta_s}{\hat{a}} = \frac{c'(\hat{a})}{\hat{a}} \ . \tag{A.50}$$

Action $\hat{a}$ locally maximizes the agent's utility if

$$-2(\lambda - 1) \sum_{s=2}^{S} b_s \sigma_s < c''(\hat{a}) \ . \tag{A.51}$$

Under Assumption (A3), we have $c''(\hat{a}) > c'(\hat{a})/\hat{a}$. Therefore, if

$$\sum_{s=2}^{S} b_s \big[-2(\lambda - 1)\sigma_s - \beta_s/\hat{a}\big] < 0 \ , \tag{A.52}$$

then (A.50) implies (A.51), 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.52) obviously is satisfied if each element of the sum is negative. Summand s is negative if and only if

$$- 2(\lambda - 1) \left(\sum_{i=1}^{s-1} (\gamma_i^H - \gamma_i^L) \right) \left(\sum_{i=s}^{S} (\gamma_i^H - \gamma_i^L) \right) \hat{a}$$
$$- \left(\sum_{\tau=s}^{S} (\gamma_\tau^H - \gamma_\tau^L) \right) \left[1 - (\lambda - 1) \left(\sum_{t=1}^{s-1} \gamma_t(\hat{a}) \right) \right] + (\lambda - 1) \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) \right] \left(\sum_{t=1}^{s-1} (\gamma_t^H - \gamma_t^L) \right) < 0 \ .$$

Rearranging the above inequality yields

$$\begin{aligned} \left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\left\{\lambda+2(\lambda-1)\left[\hat{a}\sum_{i=1}^{s-1}(\gamma_i^H-\gamma_i^L)-\sum_{i=1}^{s-1}\gamma_i(\hat{a})\right]\right\} &> 0 \\ \iff \left(\sum_{i=s}^{S}(\gamma_i^H-\gamma_i^L)\right)\left\{\lambda\left(1-\sum_{i=1}^{s-1}\gamma_i^L\right)+(2-\lambda)\sum_{i=1}^{s-1}\gamma_i^L\right\} &> 0. \quad \text{(A.53)} \end{aligned}$$

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.

$$C(\boldsymbol{b}) = \sum_{s=1}^{S}\gamma_s(\hat{a})h\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right), \quad \text{(A.54)}$$

where $$u_1(\boldsymbol{b}) = \bar{u}+c(\hat{a})-\sum_{s=2}^{S}b_s\left[\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})-(\lambda-1)\left(\sum_{\tau=s}^{S}\gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{s-1}\gamma_t(\hat{a})\right)\right].$$

The partial derivative of the cost function with respect to an arbitrary b_k is

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} = \sum_{s=1}^{k-1}\gamma_s(\hat{a})h'\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\left[\frac{\partial u_1}{\partial b_k}\right]+\sum_{s=k}^{S}\gamma_s(\hat{a})h'\left(u_1(\boldsymbol{b})+\sum_{t=2}^{s}b_t\right)\left[\frac{\partial u_1}{\partial b_k}+1\right].$$

Rearranging yields

$$\begin{aligned} \frac{\partial C(\boldsymbol{b})}{\partial b_k} &= \sum_{s=1}^{k-1}\gamma_s(\hat{a})h'(u_s)\underbrace{\left[(\lambda-1)\left(\sum_{\tau=k}^{S}\gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1}\gamma_t(\hat{a})\right)-\sum_{\tau=k}^{S}\gamma_\tau(\hat{a})\right]}_{<0} \\ &+\sum_{s=k}^{S}\gamma_s(\hat{a})h'(u_s)\underbrace{\left[(\lambda-1)\left(\sum_{\tau=k}^{S}\gamma_\tau(\hat{a})\right)\left(\sum_{t=1}^{k-1}\gamma_t(\hat{a})\right)-\sum_{\tau=k}^{S}\gamma_\tau(\hat{a})+1\right]}_{>0}. \quad \text{(A.55)} \end{aligned}$$

Note $u_s \leq u_{s+1}$ which implies that $h'(u_s) \leq h'(u_{s+1})$. Thus, the following inequality holds

$$\frac{\partial C(\boldsymbol{b})}{\partial b_k} \geq \sum_{s=1}^{k-1} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right]$$
$$+ \sum_{s=k}^{S} \gamma_s(\hat{a}) h'(u_k) \left[(\lambda - 1) \left(\sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) \right) \left(\sum_{t=1}^{k-1} \gamma_t(\hat{a}) \right) - \sum_{\tau=k}^{S} \gamma_\tau(\hat{a}) + 1 \right] . \quad \text{(A.56)}$$

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. ■

2.3. The General Case: Loss Aversion and Risk Aversion

In this part of the Appendix we provide a thorough discussion of 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 II.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

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

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

$$u_S \geq u_{S-1} \geq \ldots \geq u_1 \ . \qquad (\text{OC}_G)$$

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 A.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 $\boldsymbol{u}^* = (u_1^*, \dots, u_S^*)$.*

Proof:

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 II.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 Dimitri Bertsekas (1974). Since the constraint set is closed, the existence of a minimum follows from Weierstrass' theorem. ■

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 A.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 $\mu_{IR} > 0$.*

Proof:

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

$$u_s^N = \alpha u_s^* + (1-\alpha)\sum_{t=1}^{S}\gamma_t(\hat{a})u_t^* \ , \tag{A.57}$$

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

$$\begin{aligned}
&\sum_{s=1}^{S}\gamma_s(\hat{a})u_s^N - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^N - u_s^N\right] \\
&= \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - \hat{\alpha}(\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&> \sum_{s=1}^{S}\gamma_s(\hat{a})u_s^* - (\lambda-1)\sum_{s=1}^{S-1}\sum_{t=s+1}^{S}\gamma_s(\hat{a})\gamma_t(\hat{a})\left[u_t^* - u_s^*\right] \\
&= \bar{u} + c(\hat{a}) \ ,
\end{aligned} \tag{A.58}$$

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. ■

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 $\mu_{OC,s} = 0$ for all $s \in \{2, \dots, S\}$. In this case, 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:

$$h'(u_s) = \underbrace{\left(\mu_{IR} + \mu_{IC}\frac{\gamma_s^H - \gamma_s^L}{\gamma_s(\hat{a})}\right)}_{=:H_s} \underbrace{\left[1 - (\lambda - 1)\left(2\sum_{t=1}^{s-1}\gamma_t(\hat{a}) + \gamma_s(\hat{a}) - 1\right)\right]}_{=:\Gamma_s} - \underbrace{\mu_{IC}(\lambda - 1)\left[2\sum_{t=1}^{s-1}(\gamma_t^H - \gamma_t^L) + (\gamma_s^H - \gamma_s^L)\right]}_{=:\Lambda_s}. \quad \text{(A.59)}$$

For $\lambda = 1$ we have $h'(u_s) = H_s$, the standard "Holmström-formula".[2] 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:

$$H_s \times \Gamma_s - \Lambda_s \geq H_{s-1} \times \Gamma_{s-1} - \Lambda_{s-1}. \quad \text{(A.60)}$$

Note that for the given ordering of signals, if there exists any pair of signals $s, s-1$ such that (A.60) 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 (A.60) 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 (A.60) is violated. In summary, it may well be that

[2]See Holmström (1979).

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.

Application with Constant Relative Risk Aversion.—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 A.3: *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 $u_L^* < u_H^*$.*

Proof:

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

$$C(\boldsymbol{u}|r) \approx \sum_{s \in \mathcal{S}} \gamma_s(\hat{a}) u_s + \Omega(\boldsymbol{u}|r) \, , \tag{A.61}$$

where

$$\Omega(\boldsymbol{u}|r) \equiv \sum_{s \in \mathcal{S}} \gamma_s(\hat{a}) \left[(u_s \ln u_s)(r-1) + (1/2) u_s (\ln u_s)^2 (r-1)^2 \right] . \tag{A.62}$$

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

$$u_1 = \bar{u} + c(\hat{a}) - \sum_{s=2}^{S} b_s \left\{ \sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) - (\lambda - 1) \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) \right] \left[\sum_{t=1}^{s-1} \gamma_t(\hat{a}) \right] \right\} \equiv u_1(\boldsymbol{b}) \tag{A.63}$$

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

$$C(\boldsymbol{b}|r) \approx \bar{u} + c(\hat{a}) + (\lambda - 1) \sum_{s=2}^{S} b_s \left[\sum_{\tau=s}^{S} \gamma_\tau(\hat{a}) \right] \left[\sum_{t=1}^{s-1} \gamma_t(\hat{a}) \right] + \tilde{\Omega}(\boldsymbol{b}|r) \, . \tag{A.64}$$

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 II.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. ■

3. Appendix to Chapter III

Proof of Lemma III.3

In order to give the proof some structure, we proceed in several steps.

Claim 1: $p(2,0) > \bar{\bar{p}}$ iff $k(2) > \bar{k}$.

Proof: Follows immediately from rearranging. ||

Claim 2: $\bar{k} < 1$ iff $\mu > \bar{\mu}$.

Proof: Rearranging yields

$$\bar{k} < 1 \quad \Longleftrightarrow \quad \mu^2 - \frac{2\bar{y}}{(\bar{y} - \theta_H)}\mu + \frac{\bar{y}}{(\bar{y} - \theta_H)} < 0.$$

Define

$$f(\mu) := \mu^2 - \frac{2\bar{y}}{(\bar{y} - \theta_H)}\mu + \frac{\bar{y}}{(\bar{y} - \theta_H)}.$$

Straight-forward differentiation reveals that $f(\mu)$ is a strictly convex function, $f''(\mu) > 0$, which reaches its minimum at $\mu = \bar{y}/(\bar{y} - \theta_H)$. The zeros of $f(\mu)$ are obtained by solving

$$f(\mu) = 0 \iff \bar{\mu}_{1,2} \;=\; \frac{\sqrt{\bar{y}}(\sqrt{\bar{y}} \pm \sqrt{\theta_H})}{(\sqrt{\bar{y}} + \sqrt{\theta_H})(\sqrt{\bar{y}} - \sqrt{\theta_H})}.$$

Let

$$\bar{\mu}_1 = \frac{\sqrt{\bar{y}}}{(\sqrt{\bar{y}} + \sqrt{\theta_H})} \quad \text{and} \quad \bar{\mu}_2 = \frac{\sqrt{\bar{y}}}{(\sqrt{\bar{y}} - \sqrt{\theta_H})}.$$

Obviously, $\bar{\mu}_2 > 1$, which allows us to focus on $\bar{\mu}_1$ because we are interested only in values of μ from the interval $(0.5, 1)$. Since, by assumption, $\theta_H \in (0, \bar{y})$, we have $\bar{\mu}_1 \in (0.5, 1)$. Thus, with $f(\mu)$ being a strictly convex function which is strictly decreasing for $\mu < 1$, letting $\bar{\mu} = \bar{\mu}_1$ concludes the proof. ||

Claim 3: $p(h, h; q) < \bar{p}$ iff $\mu < \bar{\bar{\mu}}$.

PROOF: Rearranging yields

$$p(h,h;q) < \bar{p} \iff \mu^2 - \frac{2\bar{y} - q(\bar{y} - \theta_H)}{(1-q)(\bar{y} - \theta_H)}\mu + \frac{\bar{y}}{(1-q)(\bar{y} - \theta_H)} > 0$$

Define

$$g(\mu) := \mu^2 - \frac{2\bar{y} - q(\bar{y} - \theta_H)}{(1-q)(\bar{y} - \theta_H)}\mu + \frac{\bar{y}}{(1-q)(\bar{y} - \theta_H)}.$$

Differentiation with respect to μ reveals that $g(\mu)$ is a strictly convex function, $g''(\mu) > 0$, which reaches its minimum at $\mu = (2\bar{y} - q(\bar{y} - \theta_H))/2(1-q)(\bar{y} - \theta_H) > 1$. The zeros of $g(\mu)$ are obtained by solving

$$g(\mu) = 0 \iff \bar{\bar{\mu}}_{1,2} = \frac{2\bar{y} - q(\bar{y} - \theta_H) \pm \sqrt{q^2(\bar{y} - \theta_H)^2 + 4\theta_H\bar{y}}}{2(1-q)(\bar{y} - \theta_H)}$$

Once again, we are interested in values of μ from the interval $(0.5, 1)$. Since for all $q \in (0.5, 1)$ we have

$$\bar{\bar{\mu}}_2 = \frac{2\bar{y} - q(\bar{y} - \theta_H) + \sqrt{q^2(\bar{y} - \theta_H)^2 + 4\theta_H\bar{y}}}{2(1-q)(\bar{y} - \theta_H)} > 1$$

we can focus on

$$\bar{\bar{\mu}}_1 = \frac{2\bar{y} - q(\bar{y} - \theta_H) - \sqrt{q^2(\bar{y} - \theta_H)^2 + 4\theta_H\bar{y}}}{2(1-q)(\bar{y} - \theta_H)}.$$

Straightforward calculations reveal that for $q \in (0,1)$ we have $\bar{\bar{\mu}}_1 \in (0.5, 1)$. Thus, with $g(\mu)$ being a strictly convex function which is strictly decreasing for $\mu < 1$, letting $\bar{\bar{\mu}}(q) = \bar{\bar{\mu}}_1$ concludes the proof. ||

CLAIM 4: $d\bar{\bar{\mu}}(q)/dq > 0$.
PROOF: First, note that $\bar{\bar{\mu}}(q)$ is continuously differentiable with respect to q for all $q \in (0,1)$. By definition of $\bar{\bar{\mu}}(q)$, the following identity holds:

$$g(\bar{\bar{\mu}}(q)) = \bar{\bar{\mu}}(q)^2 - \frac{2\bar{y} - q(\bar{y} - \theta_H)}{(1-q)(\bar{y} - \theta_H)}\bar{\bar{\mu}}(q) + \frac{\bar{y}}{(1-q)(\bar{y} - \theta_H)} \equiv 0$$

Rearranging and differentiation with respect to q yield

$$\frac{d\bar{\bar{\mu}}(q)}{dq} = \frac{(\bar{y} - \theta_H)\bar{\bar{\mu}}(q)(\bar{\bar{\mu}}(q) - 1)}{2\bar{y}(\bar{\bar{\mu}}(q) - 1) - 2\theta_H\bar{\bar{\mu}}(q) - q(\bar{y} - \theta_H)(2\bar{\bar{\mu}}(q) - 1)}.$$

In the proof of Claim 3 we established that $\bar{\bar{\mu}}(q) \in (0.5, 1)$ for $q \in (0,1)$, which immediately implies $d\bar{\bar{\mu}}(q)/dq > 0$. ||

CLAIM 5: $\forall q \in (0,1)$, $0.5 < \bar{\mu} < \bar{\bar{\mu}}(q) < 1$.

PROOF: In Claim 2 and 3 we have already established that $\bar{\mu} \in (0.5, 1)$ and $\bar{\bar{\mu}}(q) \in (0.5, 1)$. It remains to show that $\bar{\mu} < \bar{\bar{\mu}}(q)$ for all $q \in (0, 1)$. Note that $\bar{\bar{\mu}}(q)$ is a continuous and continuously differentiable function on the interval $(-\infty, 1)$, thus $\lim_{q\to 0} \bar{\bar{\mu}}(q)$ exists and is given by

$$\lim_{q\to 0} \bar{\bar{\mu}}(q) = \frac{2\bar{y} - \sqrt{4\bar{y}\theta_H}}{2(\bar{y} - \theta_H)} = \frac{\sqrt{\bar{y}}}{\sqrt{\bar{y}} + \sqrt{\theta_H}} = \bar{\mu}.$$

From Claim 4, we know that $d\bar{\bar{\mu}}(q)/dq > 0$ for $q \in (0, 1)$, which establishes the result. ||

Combining Claims 1-5 establishes the desired result. ■

Proof of Proposition III.3

Let $p(\sigma_1, \sigma_2; q_r)$ denote the firm's updated posterior belief about the true state of the world being $\theta = \theta_H$ after receiving report $(\sigma_1, \sigma_2) \in \mathcal{M}$ from the worker under a job design which places the worker in $r \in \{1, 2\}$ divisions during the first period. The expected return from allocating the asset to project B exceeds the expected return from allocating the asset to project A if and only if $p(\sigma_1, \sigma_2; q_r)$ exceeds

$$\bar{p} = \frac{\bar{R}}{\bar{R} + \Delta} \in (0.5, 1).$$

Since $\mu \in (0.5, 1)$ implies that $p(l, l; q_r) < p(l, h; q_r) = p(h, l; q_r) = 0.5 < p(h, h; q_r)$ for $r \in \{1, 2\}$, the following observation follows immediately.

Lemma A.4: *For $r \in \{1, 2\}$, if $p(h, h; q_r) \geq \bar{p}$, then $\mathcal{B}_r = \{(h, h)\}$. Otherwise $\mathcal{B}_r = \emptyset$.*

It is readily verified, that $\mu^*(q)/\mu^{**}(q)$ is increasing in q, which implies that $p(h, h; q)$ is decreasing in q. With the question of interest being whether the firm can benefit from implementing job rotation compared to specialization, this observation renders the case where $p(h, h; q_2) < \bar{p}$ uninteresting. In this case, $\mathcal{B}_1 = \mathcal{B}_2 = \emptyset$, i.e., under both types of job design the asset is allocated to the riskless project A irrespective of the worker's report. Thus, job rotation can never be optimal because it comes along with additional costs without providing any benefit. This leaves us with two cases in which there is scope for job rotation to outperform specialization due to a more accurate probability assessment. In the first of these cases, $\bar{p} \leq p(h, h; q_1)$, the allocation rule is the same under both types of job design, $\mathcal{B}_1 = \mathcal{B}_2 = \{(h, h)\}$. In the second case, $p(h, h; q_1) < \bar{p} \leq p(h, h; q_2)$, allocation rules differ, $\mathcal{B}_1 = \emptyset$ and $\mathcal{B}_2 = \{(h, h)\}$.

It can be shown that $p(h, h; q) > \bar{p}$ if and only if $\mu > \bar{\bar{\mu}}(q)$, where

$$\bar{\bar{\mu}}(q) = \frac{2\bar{R} - q(\bar{R} - \Delta) - \sqrt{q^2(\bar{R} - \Delta)^2 + 4\bar{R}\Delta}}{2(1 - q)(\bar{R} - \Delta)}$$

with $\bar{\bar{\mu}}'(q) > 0$ and $\bar{\bar{\mu}}(q) \in (0.5, 1)$ for all $q \in (0, 1)$. These properties of $\bar{\bar{\mu}}(q)$ follow immediately from the proof of Lemma III.3. The following observation then is immediate.

Lemma A.5: *Given $\bar{R}$, Δ, and $0 < q_2 < q_1 < 1$, we have*

(a′) $\bar{p} \leq p(h, h; q_1) < p(h, h; q_2)$ *iff* $\mu \in [\bar{\bar{\mu}}(q_1), 1)$ *;*

(b′) $p(h, h; q_1) < \bar{p} \leq p(h, h; ; q_2)$ *iff* $\mu \in [\bar{\bar{\mu}}(q_2), \bar{\bar{\mu}}(q_1))$ *.*

In both cases (a′) and (b′), job rotation can outperform specialization if the cost of job rotation is sufficiently small. To formally establish this result, let $P(q)$ denote the probability of two h signals being reported for a given q. Moreover, let $\mathbb{E}[R|r]$ denote the firm's ex-ante expected return from asset allocation under a job design with $r \in \{1, 2\}$.

Case (a′): Both types of job design lead to the same allocation rule, $\mathcal{B}_1 = \mathcal{B}_2 = \{h, h\}$. Thus, $\mathbb{E}[R|2] > \mathbb{E}[R|1]$ if and only if

$$P(q_2)p(h, h; q_2)(\bar{R} + \Delta) + (1 - P(q_2))\bar{R} - c > P(q_1)p(h, h; q_1)(\bar{R} + \Delta) + (1 - P(q_1))\bar{R},$$

or equivalently, if and only if $c < \bar{c}$, where

$$\bar{c} = \frac{\mu(1 - \mu)(q_1 - q_2)(\bar{R} - \Delta)}{2}.$$

Case (b′): Under specialization we have $\mathcal{B}_1 = \emptyset$, whereas under job rotation the allocation rule is $\mathcal{B}_2 = \{h, h\}$. Thus, $\mathbb{E}[R + y|2] > \mathbb{E}[R + y|1]$ if and only if

$$P(q_2)p(h, h; q_2)(\bar{R} + \Delta) + (1 - P(q_2))\bar{R} - c > \bar{R},$$

or equivalently, if and only if $c < \bar{\bar{c}}$, where

$$\bar{\bar{c}} = \frac{\mu(\mu + q_2(1 - \mu))\Delta - (1 - \mu)((1 - \mu) + q_2\mu)\bar{R}}{2}.$$

It is readily verified that $\bar{\bar{c}} > 0$ whenever $\mu \in (\bar{\bar{\mu}}(q_2), \bar{\bar{\mu}}(q_1))$.

This establishes the desired result. ■

4. Appendix to Chapter IV

4.1. Proofs of Propositions and Lemmas

Proof of Lemma IV.1:

For any wholesale price $w_i \geq P(0) - k_i$ firm i's input demand equals zero, whereas for

$w_i < P(0) - k_i$ the optimal input demand, $q(c_i)$, is strictly positive and characterized by

$$MR(q(c_i)) := q(c_i)P'(q(c_i)) + P(q(c_i)) = c_i. \tag{A.65}$$

Under the assumptions imposed on the inverse demand function, whenever $q(c_i) > 0$ we have $q'(c_i) < 0$, $q''(c_i) \leq 0$, $\pi'(c_i) < 0$, $MR'(q) < 0$ and $MR''(q) \leq 0$.

The upstream supplier's profit from charging an active downstream firm with own marginal cost $k_i < P(0)$ a wholesale price $w < P(0) - k_i$ is $\Pi(w; k_i) := wq(w + k_i)$. With $\Pi(w; k_i)$ being strictly concave on the interval $[0, P(0) - k_i]$ the optimal unconstrained discriminatory wholesale price $w^d(k_i)$ satisfies

$$q(w^d(k_i) + k_i) + w^d(k_i)q'(w^d(k_i) + k_i) = 0. \tag{A.66}$$

We first show that $w^d(k) < w^d(0)$. Suppose, in contradiction, that $w^d(k) \geq w^d(0)$. Differentiation of (A.65) with respect to c_i yields

$$q'(c_i) = \frac{1}{2P'(q(c_i)) + q(c_i)P''(q(c_i))} = \frac{1}{MR'(q(c_i))}, \tag{A.67}$$

where the second equality follows from the definition of $MR(q)$. From (A.66) it follows that the optimal discriminatory wholesale price charged to a downstream firm with own marginal cost k_i satisfies

$$w^d(k_i) = -\frac{q(w^d(k_i) + k_i)}{q'(w^d(k_i) + k_i)} = -q(w^d(k_i) + k_i)MR'(q(w^d(k_i) + k_i)). \tag{A.68}$$

In consequence, $w^d(0) \leq w^d(k)$ if and only if $-q(w^d(0))MR'(q(w^d(0))) \leq -q(w^d(k) + k)MR'(q(w^d(k) + k))$. Since

$$\frac{d}{dc}\left[-q(c)MR'(q(c))\right] = -q'(c)\left[MR'(q(c)) + q(c)MR''(q(c))\right] < 0, \tag{A.69}$$

$w^d(k) \geq w^d(0)$ implies $w^d(0) \geq w^d(k) + k$, a contradiction. Therefore, $w^d(k) < w^d(0)$.

Knowing that $w^d(k) < w^d(0)$, we next show that $w^d(0) < w^d(k) + k$. Suppose, in contradiction, that $w^d(0) \geq w^d(k) + k$. Then $q(w^d(0)) \leq q(w^d(k) + k)$, and in consequence

$$0 > MR'(q(w^d(0))) \geq MR'(q(w^d(k) + k)) \tag{A.70}$$

by marginal revenue being decreasing and concave. From above, we know that $w^d(0) > w^d(k)$, and thus, according to (A.68), we have

$$-q(w^d(0))MR'(q(w^d(0))) > -q(w^d(k) + k)MR'(q(w^d(k) + k)). \tag{A.71}$$

Taken together (A.70) and (A.71) imply $q(w^d(k) + k) < q(w^d(0))$ and in consequence $w^d(k) + k > w^d(0)$, a contradiction. Thus, $w^d(k) + k > w^d(0)$.

When unrestricted in its price setting, U's profit from charging a common wholesale price w is

$$\Pi^u(w;k) := \begin{cases} \Pi(w;0) + \Pi(w;k) & \text{for} \quad w < P(0) - k \\ \Pi(w;0) & \text{for} \quad P(0) - k \leq w < P(0) \\ 0 & \text{for} \quad w \geq P(0) \end{cases},$$

Obviously, serving no firm clearly is not optimal. Moreover, under Assumption (A2), it is never optimal to serve only firm I, i.e., we must have $w^u(k) < P(0) - k$. Note that $\Pi^u(w;k)$ is strictly concave on $[0, P(0) - k]$. By definition of $w^d(0)$ and $w^d(k)$, $w^d(0) > w^d(k)$ from above, and concavity of $\Pi(w;k_i)$ on $[0, P(0) - k_i]$ for $i \in \{I, E\}$, we have

$$\frac{d\Pi^u(w;k)}{dw} = \frac{d\Pi(w;0)}{dw} + \frac{d\Pi(w;k)}{dw} > 0$$

for all $w \in [0, w^d(k)]$, which immediately implies that $w^d(k) < w^u(k)$.

It remains to show that $w^u(k) < w^d(0)$. With $w^d(0) < P(0) - k$, under Assumption (A2) we have

$$\left.\frac{d\Pi^u(w;k)}{dw}\right|_{w=w^d(0)} = \left.\frac{d\Pi(w;0)}{dw}\right|_{w=w^d(0)} + \left.\frac{d\Pi(w;k)}{dw}\right|_{w=w^d(0)} = \left.\frac{d\Pi(w;k)}{dw}\right|_{w=w^d(0)} < 0,$$

where the last equality follows from definition of $w^d(0)$, and the inequality follows from the fact that $w^d(0) > w^d(k)$ and $\Pi(w;k)$ being strictly concave on $[0, P(0) - k]$. Strict concavity of $\Pi^u(w;k)$ on $[0, P(0) - k]$ then immediately implies $w^u(k) < w^d(0)$, which establishes the desired result. ■

Proof of Proposition IV.1:

We prove part (i) first. As a preliminary consideration, consider two active downstream firms i and j with own marginal cost $k_j < k_i$. For $w < P(0) - k_i$, we have $0 < q(w+k_i) < q(w+k_j)$, and $q'(c) < 0$ for all $c \in [w + k_j, w + k_i]$. The optimal quantity demanded by a downstream firm with own marginal cost $\tilde{k} \in [k_j, k_i]$ at wholesale price w satisfies

$$P(q(w + \tilde{k})) - \tilde{k} \equiv w - q(w + \tilde{k})P'(q(w + \tilde{k})). \tag{A.72}$$

Differentiation of this expression with respect to $\tilde{k}$ yields

$$\frac{d}{d\tilde{k}}[P(q(w + \tilde{k})) - \tilde{k}]$$
$$= -q'(w + \tilde{k}) \left[P'(q(w + \tilde{k})) + q(w + \tilde{k})P''(q(w + \tilde{k}))\right] < 0. \tag{A.73}$$

Thus, a more efficient downstream firm charges a higher mark-up.

Now, in case (i), with $F < \min\{\bar{F}^d(k), \hat{F}(k)\}$, we always have the optimal uniform price bracketed by the optimal discriminatory wholesale prices: for $F \leq \bar{F}^u(k)$ we have $w^d(k) < w^u(k) < w^d(0)$ by Lemma IV.1; for $F \in (\bar{F}^u(k), \min\{\bar{F}^d(k), \hat{F}(k)\})$ the optimal uniform wholesale price equals $w^R(F;k)$ where $w^R(\bar{F}^u(k);k) = w^u(k)$, $w^R(\bar{F}^d(k);k) = w^d(k)$, and $dw^R/dF < 0$. Letting q_i^d and q_i^u denote firm i's quantity under price discrimination and uniform pricing, respectively, where $i \in \{I, E\}$, this in turn implies that $q_I^d < q_I^u$ and $q_E^d > q_E^u$. Welfare under price discrimination is

$$W^d(F;k) = \int_0^{q_I^d} P(x)dx + \int_0^{q_E^d} P(x)dx - kq_E^d - F, \tag{A.74}$$

whereas welfare under uniform pricing is

$$W^u(F;k) = \int_0^{q_I^u} P(x)dx + \int_0^{q_E^u} P(x)dx - kq_E^u - F. \tag{A.75}$$

Then

$$\begin{aligned}\Delta W(F;k) = \int_{q_E^u}^{q_E^d} P(x)dx - \int_{q_I^d}^{q_I^u} P(x)dx - k(q_E^d - q_E^u) \\ < (q_E^d - q_E^u)[P(q_E^u) - k] - (q_I^u - q_I^d)P(q_I^u).\end{aligned} \tag{A.76}$$

From (A.73) we know that $P(q_E^u) - k < P(q_I^u)$. Thus, $q_E^d - q_E^u \leq q_I^u - q_I^d$, or equivalently $Q^d(F;k) = q_I^d + q_E^d \leq q_I^u + q_E^u = Q^u(F;k)$, is a sufficient condition for $\Delta W(F;k) < 0$.

Parts (ii) and (iii) follow immediately from the reasoning in the text. ■

4.2. Downstream Competition

In this appendix we provide a detailed analysis if entry into the downstream industry results in downstream competition as discussed in Section 5. The equilibrium concept employed is subgame perfect Nash equilibrium. We solve the game by backward induction, beginning in stage three.

Stage 3: For given wholesale prices and a given number of active firms in the intermediate industry, we determine the quantities produced of the final good by firms active in the downstream market. If a downstream firm with own marginal cost k_i is a downstream monopolist, its demand for the input at a wholesale price w is

$$q(w + k_i) = \begin{cases} \frac{1-w-k_i}{2} & \text{for} \quad w < 1 - k_i \\ 0 & \text{for} \quad w \geq 1 - k_i \end{cases}.$$

If two firms i and j are active in the downstream market, then firm i's best response at wholesale price w_i given that firm j produces quantity q_j is

$$q(q_j; w_i + k_i) = \max\left\{0, \frac{1 - w_i - k_i - q_j}{2}\right\} \tag{A.77}$$

For $2w_i - w_j < 1 - 2k_i + k_j$ and $2w_j - w_i < 1 - 2k_j + k_i$ the Cournot Nash equilibrium is interior with both firms producing strictly positive quantities. The equilibrium quantity of firm $i \neq j$ is

$$q(w_i + k_i, w_j + k_j) = \frac{1 - 2(w_i + k_i) + (w_j + k_j)}{3}. \tag{A.78}$$

If $2w_i - w_j < 1 - 2k_i + k_j$ and $2w_j - w_i \geq 1 - 2k_j + k_i$, then firm j produces nothing whereas firm i produces its monopoly quantity. For $2w_i - w_j \geq 1 - 2k_i + k_j$ and $2w_j - w_i \geq 1 - 2k_j + k_i$ both downstream firms produce zero quantity.

Stage 2 Given wholesale prices w_I and w_E charged to firm I and firm E, respectively, and correctly anticipating Nash equilibrium play in stage three, firm E enters the market if its profits in the resulting market outcome in stage 3 exceed the entry fee. If indifferent between entering and not entering the market, as a tie-breaking rule we assume that firm E behaves as the upstream supplier U wishes.[3] If profits for firm E in the resulting market outcome in stage three are strictly negative, then E does not enter the intermediate industry.

Stage 1 Correctly anticipating firm E's entry decision in stage two and Nash equilibrium play in stage three, U chooses wholesale prices w_I and w_E in order to maximize upstream profits. In what follows, we refer to a duopoly as a situation, in which E enters the downstream market and downstream demand is strictly positive for both firms I and E. Again, when indifferent between implementing a downstream duopoly or a downstream monopoly, the upstream supplier implements a downstream monopoly. Let $\Pi^r_{\{i\}}$ denote firm U's profit from implementing firm $i \in \{I, E\}$ as a downstream monopolist, and let $\Pi^r_{\{I,E\}}$ denote firm U's profit from implementing firms I and E as downstream duopolists. Superscript $r \in \{d, u\}$ again refers to either a discriminatory pricing regime or a uniform pricing regime. Moreover, in order not to clutter notation, we will often suppress the dependency of downstream quantity choices on effective marginal costs as well as the dependency of optimal wholesale prices and welfare on the entry fee and own marginal costs of the downstream firms.

Lemma A.6: *Under Price discrimination,*

(i) if $\sqrt{F} \leq (1/6) - (1/3)k$, then U charges wholesale prices $w^d_I = w^d(0) = 1/2$ and $w^d_E = w^d(k) = (1-k)/2$. This implements a downstream duopoly resulting in quantities $q^d_I = (1+k)/6$, $q^d_E = (1-2k)/6$, and $Q^d = (2-k)/6$;

[3] We impose this alternative tie-breaking rule for expositional purposes only. Sticking to the original tie-breaking rule, i.e., firm E enters whenever its profits are nonnegative, yields exactly the same results.

(ii) *if* $(1/6) - (1/3)k < \sqrt{F} < (1/3) - (2/3)k$*, then* U *charges wholesale prices* $w_I^d = w_I^R = 1/2$ *and* $w_E^d = w_E^R(\sqrt{F}; k) = (3/4) - k - (3/2)\sqrt{F}$*. This implements a downstream duopoly resulting in quantities* $q_I^d = (1/4) - (1/2)\sqrt{F}$*,* $q_E^d = \sqrt{F}$*, and* $Q^d = (1/4) + (1/2)\sqrt{F}$*;*

(iii) *if* $(1/3) - (2/3)k \leq \sqrt{F}$*, then* U *charges wholesale prices* $w_I^d = w_M = 1/2$ *and* $w_E^d = \infty$*. This implements a downstream monopoly resulting in quantities* $q_I^d = Q^d = 1/4$*.*

Proof:
Suppose U wants to implement a downstream duopoly. Then U chooses wholesale prices in order to solve the following problem:

Program D-PD:

$$\max_{(w_I, w_E) \in \mathbb{R}^2_{\geq 0}} w_I \frac{1 - 2w_I + (w_E + k)}{3} + w_E \frac{1 - 2(w_E + k) + w_I}{3}$$

$$\text{subject to} \quad q_I = \frac{1 - 2w_I + (w_E + k)}{3} > 0$$

$$q_E = \frac{1 - 2(w_E + k) + w_I}{3} > 0$$

$$F \leq \left[\frac{1 - 2(w_E + k) + w_I}{3}\right]^2$$

Next, we show that for a sufficiently low entry fee, the solution to Program D-PD is identical to the solution of the relaxed program, which only considers the latter two constraints.

Claim 1: *If* $\sqrt{F} \leq (1/2) - (2/3)k$*, the solution to Program R*

$$\max_{(w_I, w_E)} w_I \frac{1 - 2w_I + (w_E + k)}{3} + w_E \frac{1 - 2(w_E + k) + w_I}{3}$$

$$\textit{subject to} \quad 2w_E - w_I \leq 1 - 2k - 3\sqrt{F},$$

also solves Program D-PD.

Proof of Claim 1: First, note that the latter two constraints of Program D-PD can equivalently be replaced by the following condition:

$$2w_E - w_I \leq 1 - 2k - 3\sqrt{F}, \tag{A.79}$$

which corresponds to the one constraint in Program R. The Lagrangian associated with Program R is

$$\mathcal{L} = w_I \frac{1 - 2w_I + (w_E + k)}{3} + w_E \frac{1 - 2(w_E + k) + w_I}{3} - \lambda \left\{2w_E - w_I - (1 - 2k - 3\sqrt{F})\right\}. \tag{A.80}$$

With $\mathcal{L}$ being a strictly concave function, the associated Kuhn-Tucker conditions are sufficient for global optimality. These Kuhn-Tucker conditions are given by

$$\frac{\partial \mathcal{L}}{\partial w_I} = \frac{1 + 2w_E + k - 4w_I}{3} + \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial w_E} = \frac{1 - 4w_E - 2k + 2w_I}{3} - 2\lambda = 0$$
$$\lambda \geq 0 \ \left(= 0 \ \text{ if } \ 2w_E - w_I < 1 - 2k - 3\sqrt{F}\right)$$
$$2w_E - w_I \leq 1 - 2k - 3\sqrt{F}$$

Consider the case of $\sqrt{F} \leq (1/6) - (1/3)k$ first. Suppose the constraint is not binding, i.e., $2w_E - w_I < 1 - 2k - 3\sqrt{F}$. The complementary slackness condition then implies $\lambda = 0$. Combining the two first-order conditions yields wholesale prices $w_I = 1/2$ and $w_E = (1 - k)/2$. It is readily verified that for $\sqrt{F} \leq (1/6) - (1/3)k$, at these prices the constraint of Program R is satisfied. Moreover, under these wholesale prices, all remaining constraints of Program D-PD are also satisfied: wholesale prices are nonnegative, and associated quantities are strictly positive, $q_I = (1 + k)/6$ and $q_E = (1 - 2k)/6$. Next, consider the case $(1/6) - (1/3)k < \sqrt{F} \leq (1/2) - (2/3)k$. Suppose that the constraint is binding, i.e., $2w_E - w_I = 1 - 2k - 3\sqrt{F}$. The complementary slackness condition then implies $\lambda \geq 0$. Combining the two first-order conditions yields $w_I = 1/2$. Inserting this into the binding constraint leads to $w_E = (3/4) - k - (3/2)\sqrt{F}$. Solving for the Lagrange parameter yields $\lambda = (-1 + 2k + 6\sqrt{F})/6$, which is strictly positive for $(1/6) - (1/3)k < \sqrt{F}$. It is readily verified that for $\sqrt{F} \leq (1/2) - (2/3)k$ all remaining constraints of Program D-PD are also satisfied under these wholesale prices: wholesale prices are nonnegative, and associated quantities are strictly positive, $q_I = (1/4) - (1/2)\sqrt{F}$ and $q_E = \sqrt{F}$. This proves Claim 1. ||

Straightforward calculations show that U's profit from implementing a downstream duopoly is $\Pi^d_{\{I,E\}} = (1 - k + k^2)/6$ if $\sqrt{F} \leq (1/6) - (1/3)k$, and $\Pi^d_{\{I,E\}} = (1/8) + ((1/2) - k)\sqrt{F} - (3/2)(\sqrt{F})^2$ if $(1/6) - (1/3)k < \sqrt{F} \leq (1/2) - (2/3)k$. Note that for $\sqrt{F} > (1/2) - (2/3)k$, U's problem becomes more heavily constrained, such that U's profit cannot be larger than for $\sqrt{F} \leq (1/2) - (2/3)k$.

Next, suppose U wants to implement a downstream monopoly. Straightforward calculations show that the maximum profit U can make when facing a downstream monopolist with own marginal cost k_i, the optimal wholesale price for U to charge is $w = (1-k_i)/2$, which results in downstream demand $q = (1-k_i)/4$ and upstream profits $\Pi^d_{\{i\}} = (1-k_i)^2/8$. Since U's maximum profit decreases in the downstream monopolists own marginal cost, U always prefers I to become a monopolist over E becoming a monopolist. Since under price discrimination U can charge E a prohibitively high price which keeps E out of the downstream market without affecting the price paid by the incumbent firm I, U can always make I the downstream monopolist, resulting in upstream profits of $\Pi^d_{\{I\}} = 1/8$.

In order to conclude the proof, we have to determine when U prefers to implement a downstream duopoly over implementing a downstream monopoly. If $\sqrt{F} \leq (1/6) - (1/3)k$, $\Pi^d_{\{I,E\}} > \Pi^d_{\{I\}}$ if and only if $(1-2k)^2 > 0$. Thus, if $\sqrt{F} \leq (1/6) - (1/3)k$, U will implement a downstream duopoly resulting in quantities $q^d_I = (1+k)/6$ and $q^d_E = (1-2k)/6$. Next, if $(1/6) - (1/3)k < \sqrt{F} \leq (1/2) - (2/3)k$, $\Pi^d_{\{I,E\}} > \Pi^d_{\{I\}}$ if and only if $\sqrt{F} < (1/3) - (2/3)k$. Thus, if $(1/6) - (1/3)k < \sqrt{F} < (1/3) - (2/3)k$, U will implement a downstream duopoly resulting in quantities $q^d_I = (1/4) - (1/2)\sqrt{F}$ and $q^d_E = \sqrt{F}$, whereas for $\sqrt{F} \geq (1/3) - (2/3)k$, U will implement a downstream monopoly resulting in quantity $q^d_I = 1/4$. This establishes the desired result. ■

Lemma A.7: *Under uniform pricing,*

(i) if $k < 2 - \sqrt{3}$ *and* $\sqrt{F} \leq (1/6) - (7/12)k$, *then* U *charges a wholesale prices* $w^u = w^u(k) = (1/2) - (1/4)k$. *This implements a downstream duopoly resulting in quantities* $q^u_I = (2+5k)/12$, $q^u_E = (2-7k)/12$, *and* $Q^u = (2-k)/6$*;*

(ii) if $k < 2 - \sqrt{3}$ *and* $(1/6) - (7/12)k < \sqrt{F} < (1/6) - (7/12)k + (\sqrt{1-4k+k^2})/12$, *then* U *charges a wholesale prices* $w^u = w^{Ru}(\sqrt{F}; k) = 1 - 2k - 3\sqrt{F}$. *This implements a downstream duopoly resulting in quantities* $q^u_I = k + \sqrt{F}$, $q^u_E = \sqrt{F}$, *and* $Q^u = k + 2\sqrt{F}$*;*

(iii) if $k \geq 2 - \sqrt{3}$ *or* $\sqrt{F} \geq (1/6) - (7/12)k + (\sqrt{1-4k+k^2})/12$, *then* U *charges a wholesale price* $w^u = w_M = (1/2)$. *This implements a downstream monopoly resulting in quantities* $q^u_I = Q^u = 1/4$.

Proof:
Suppose U wants to implement a downstream duopoly. Then U chooses the uniform wholesale price in order to solve the following problem:

Program D-UNI:

$$\max_{w \in \mathbb{R}_{\geq 0}} w\frac{2-2w-k}{3}$$
$$\text{subject to} \quad q_I = \frac{1-w+k}{3} > 0$$
$$q_E = \frac{1-w-2k}{3} > 0$$
$$F \leq \left[\frac{1-w-2k}{3}\right]^2$$

First, note that if the second constraint holds also the first constraint holds with strict inequality, i.e., if E demands a nonnegative quantity at wholesale price w, $q_E \geq 0$, then I demands a strictly positive quantity, $q_I > 0$. Moreover, the second and third constraint together can equivalently be replaced by the following condition: $w \leq 1 - 2k - 3\sqrt{F}$. Thus, Program D-UNI can be equivalently rewritten as

Program D-UNI:

$$\max_{w \in \mathbb{R}_{\geq 0}} w \frac{2 - 2w - k}{3}$$
$$\text{subject to} \quad w \leq 1 - 2k - 3\sqrt{F}$$

Note that U's objective is maximizing a strictly concave function with a unique global maximum attained at $w = (2 - k)/4$. Therefore, if $(2 - k)/4 \leq 1 - 2k - 3\sqrt{F}$, or, equivalently, if $\sqrt{F} \leq (1/6) - (7/12)k$, the optimal uniform wholesale price that implements a downstream duopoly is $w = (2 - k)/4$, resulting in quantities $q_I = (2 + 5k)/12$ and $q_E = (2 - 7k)/12$. Note that $q_E > 0$—and thus also $q_I > 0$—if and only if $k < 2/7$. If $\sqrt{F} > (1/6) - (7/12)k$, the constraint becomes binding. If $\sqrt{F} \leq (1/3) - (2/3)k$, the optimal uniform wholesale price in order to implement a downstream duopoly is given by $w = 1 - 2k - 3\sqrt{F}$, resulting in quantities $q_I = k + \sqrt{F}$ and $q_E = \sqrt{F}$. If $\sqrt{F} > (1/3) - (2/3)k$, implementation of a downstream duopoly with E demanding a strictly positive quantity and making nonnegative profits is not possible with a nonnegative wholesale price.

Straightforward calculations show that U's profit from implementing a downstream duopoly is $\Pi^u_{\{I,E\}} = (2 - k)^2/24$ if $\sqrt{F} \leq (1/6) - (7/12)k$, and $\Pi^u_{\{I,E\}} = (1 - 2k - 3\sqrt{F})(k + 2\sqrt{F})$ if $(1/6) - (7/12)k < \sqrt{F} \leq (1/3) - (2/3)k$.

Next, suppose that U wants to implement a downstream monopoly. As noted above, for a given wholesale price w, if E demands a nonnegative quantity, then I demands a strictly positive quantity. Thus, under uniform pricing, the only possible form monopoly can take in the downstream market is with I as downstream monopolist. Therefore, when implementing a downstream monopoly under uniform pricing, U has to choose a wholesale price at which E does not find it profitable to enter and demand a strictly positive quantity. From above we know that this requires the wholesale price to be sufficiently high, i.e., $w > 1 - 2k - 3\sqrt{F}$. Under our tie-breaking rule that E does what U wants him to do when indifferent between entering and not entering the market, U implements a downstream monopoly whenever he chooses a wholesale price $w \geq 1 - 2k - 3\sqrt{F}$. With the quantity demanded by downstream monopolist I being $q_I = (1 - w)/2$, by the choice of the wholesale price U maximizes a strictly concave function with a unique stationary point at $w = 1/2$ subject to the afore-mentioned constraint. In consequence, if $1/2 \geq 1 - 2k - 3\sqrt{F}$, or equivalently, if $\sqrt{F} \geq (1/6) - (2/3)k$, then the optimal wholesale price to implement a downstream monopoly is $w = 1/2$ resulting in quantity $q_I = 1/4$ and upstream profit $\Pi^u_{\{I\}} = 1/8$. If $\sqrt{F} < (1/6) - (2/3)k$, then the optimal wholesale price to implement a downstream monopoly is $w = 1 - 2k - 3\sqrt{F}$ resulting in quantity $q_I = k + (3/2)\sqrt{F}$ and upstream profit $\Pi^u_{\{I\}} = (1 - 2k - 3\sqrt{F})(k + (3/2)\sqrt{F})$. Note that $w = 1 - 2k - 3\sqrt{F} \geq 0$ if and only if $\sqrt{F} \leq (1/3) - (2/3)k$, which obviously is satisfied for $\sqrt{F} < (1/6) - (2/3)k$.

In order to conclude the proof, we have to determine when U prefers to implement a downstream duopoly over implementing a downstream monopoly. Combining the observations obtained above, we have to distinguish four cases. (i) If $\sqrt{F} > (1/3) - (2/3)k$, implementation of a downstream duopoly is not feasible. Thus, U implements an unconstrained downstream monopoly resulting in quantity $q^u_I = 1/8$. (ii) If $(1/6) - (7/12)k < \sqrt{F} \leq (1/3) - (2/3)k$, then $\Pi^u_{\{I,E\}} > \Pi^u_{\{I\}}$ if and only if $(1 - 2k - 3\sqrt{F})(k + 2\sqrt{F}) > 1/8$, or, equivalently, $(\sqrt{F})^2 - ((2 - 7k)/6)\sqrt{F} + (1 - 8k + 16k^2)/48 < 0$. For $k < 2 - \sqrt{3}$, this condition implies that $\Pi^u_{\{I,E\}} > \Pi^u_{\{I\}}$ if and only if $(1/6) - (7/12)k < \sqrt{F} < (1/6) - (7/12)k + (\sqrt{1 - 4k + k^2})/12$, whereas for $k \geq 2 - \sqrt{3}$ we always have $\Pi^u_{\{I,E\}} \leq \Pi^u_{\{I\}}$. Thus, U implements a downstream duopoly resulting in quantities $q^u_I = k + \sqrt{F}$ and $q^u_E = \sqrt{F}$ if $k < 2 - \sqrt{3}$ and $(1/6) - (7/12)k < \sqrt{F} < (1/6) - (7/12)k + (\sqrt{1 - 4k + k^2})/12$, and a downstream monopoly resulting in quantity $q^u_I = 1/8$ otherwise. (iii) If $(1/6) - (2/3)k < \sqrt{F} \leq (1/6) - (7/12)k$, where the latter inequality implies $k < 2/7$, then $\Pi^u_{\{I,E\}} > \Pi^u_{\{I\}}$ if and only if $(2 - k)^2/24 > 1/8$. This latter condition implies that $\Pi^u_{\{I,E\}} > \Pi^u_{\{I\}}$ if and only if $k < 2 - \sqrt{3}$. Thus, U implements a downstream duopoly resulting in quantities $q^u_I = (2 + 5k)/12$ and $q^u_E = (2 - 7k)/12$ if $k < 2 - \sqrt{3}$ and $(1/6) - (2/3)k < \sqrt{F} \leq (1/6) - (7/12)k$, and a downstream monopoly resulting in quantity $q^u_I = 1/8$ otherwise. (iv) If $\sqrt{F} \leq (1/6) - (2/3)k$, which implies $k \leq 1/4$, then $\Pi^u_{\{I,E\}} > \Pi^u_{\{I\}}$ if and only if $(2 - k)^2/24 > (1 - 2k - 3\sqrt{F})(k + (3/2)\sqrt{F})$, or, equivalently, $(\sqrt{F})^2 + ((4k - 1)/3)\sqrt{F} + (7k - 2)^2/108 > 0$. This latter inequality always holds for $k < 2 - \sqrt{3}$, and thus is always satisfied in the case under consideration. Thus, U implements a downstream duopoly resulting in quantities $q^u_I = (2 + 5k)/12$ and $q^u_E = (2 - 7k)/12$. This establishes the desired result. ■

Proposition A.4: *(i)* $W^d > W^u$ *if and only if*

(1.) $k < 2 - \sqrt{3}$ *and* $(1/6) - (7/12)k + (\sqrt{1 - 4k - k^2})/12 \leq \sqrt{F} < (1/3) - (8/9)k$, *or*

(2.) $2 - \sqrt{3} \leq k \leq 3/10$ *and* $\sqrt{F} < (1/3) - (8/9)k$, *or*

(3.) $3/10 < k < 17/46$ *and* $\sqrt{F} < \sqrt{(23/72)k^2 - (5/18)k + (17/288)}$.

(ii) $W^d < W^u$ *if and only if*

(1.) $k < 2 - \sqrt{3}$ *and* $\sqrt{F} < (1/6) - (7/12)k + (\sqrt{1 - 4k - k^2})/12$*, or*

(2.) $\sqrt{F} > (1/3) - (8/9)k$ *for* $k < 3/10$ *or* $\sqrt{F} > \sqrt{(23/72)k^2 - (5/18)k + (17/288)}$ *for* $k \geq 3/10$*, and* $\sqrt{F} < (1/3) - (2/3)k$*.*

(iii) If $\sqrt{F} \geq (1/3) - (2/3)k$*, then* $W^d = W^u$*.*

Proof:

First, note that for $k \in (0, 2 - \sqrt{3}]$, $(1/6) - (7/12)k + (\sqrt{1 - 4k + k^2})/12 < (1/3) - (2/3)k$, $(1/6) - (7/12)k + (\sqrt{1 - 4k + k^2})/12 = (1/6) - (1/3)k$ if and only if $k = (\sqrt{3} - 1)/4$, and $(1/6) - (7/12)k + (\sqrt{1 - 4k + k^2})/12 = (1/6) - (7/12)k$ if and only if $k = 2 - \sqrt{3}$. These observations together with Lemmas A.6 and A.7 imply that there are five cases to consider that we labeled with Roman numerals in Figure IV.3.

(I) $k < 2 - \sqrt{3}$ and $\sqrt{F} \leq (1/6) - (7/12)k$:
Under both pricing regimes, U implements an unconstrained downstream duopoly, resulting in the same aggregate output, $Q^d = Q^u = (2 - k)/6$. Under price discrimination, however, the less efficient firm E produces a higher share of output, $q_E^d = (1 - 2k)/6 > (2 - 7k)/12 = q_E^u$. Thus, welfare is strictly lower under price discrimination than under uniform pricing, $W^d < W^u$.

(II) $k < 2 - \sqrt{3}$, $(1/6) - (7/12)k < \sqrt{F} < (1/6) - (7/12)k + (\sqrt{1 - 4k + k^2})/12$, and $\sqrt{F} \leq (1/6) - (1/3)k$:
Under price discrimination, U implements an unconstrained duopoly resulting in quantities $q_I^d = (1 + k)/6$, $q_E^d = (1 - 2k)/6$, and $Q^d = (2 - k)/6$, whereas under uniform pricing, U implements a constrained duopoly, resulting in quantities $q_I^u = k + \sqrt{F}$, $q_E^u = \sqrt{F}$, and $Q^u = k + 2\sqrt{F}$. $(1/6) - (7/12)k < \sqrt{F}$ implies that aggregate output is larger under uniform pricing than under price discrimination, $Q^d < Q^u$. $\sqrt{F} \leq (1/6) - (1/3)k$, on the other hand, implies, that the less efficient firm's output is (at least weakly) lower under uniform pricing than under price discrimination. Together, these observations imply that welfare under uniform pricing exceeds welfare under price discrimination, $W^d < W^u$.

(III) $k < (\sqrt{3} - 1)/4$ and $(1/6) - (1/3)k < \sqrt{F} < (1/6) - (7/12)k + (\sqrt{1 - 4k + k^2})/12$:
Under both pricing regimes, U implements a constrained duopoly. Under price discrimination, this results in in quantities $q_I^d = (1/4) - (1/2)\sqrt{F}$, $q_E^d = \sqrt{F}$, and $Q^d = (1/4) + (1/2)$. Under uniform pricing, the resulting quantities are $q_I^u = k + \sqrt{F}$, $q_E^u = \sqrt{F}$, and $Q^u = k + 2\sqrt{F}$. While the less efficient firm's output being identical under both pricing regimes, $q_E^d = q_E^u = \sqrt{F}$, $(1/6) - (1/3)k \leq \sqrt{F}$

implies that aggregate output is higher under uniform pricing than under price discrimination, $Q^d < Q^u$. This, in turn, implies that welfare under uniform pricing exceeds welfare under price discrimination, $W^d < W^u$.

(IV) $(1/6)-(1/3)k < \sqrt{F} < (1/3)-(2/3)k$ and $(1/6)-(7/12)k+(\sqrt{1-4k+k^2})/12 \leq \sqrt{F}$:

Under price discrimination, U implements a constrained downstream duopoly, resulting in quantities $q_I^d = (1/4)-(1/2)\sqrt{F}$, $q_E^d = \sqrt{F}$, and $Q^d = (1/4)+(1/2)\sqrt{F}$. Welfare under this pricing regime then is given by

$$W^d = \int_0^{Q^d} (1-x)dx - kq_E^d - F = \frac{7}{32} + \left(\frac{3}{8}-k\right)\sqrt{F} - \frac{9}{8}(\sqrt{F})^2. \quad \text{(A.81)}$$

Under uniform pricing, on the other hand, U implements an unconstrained downstream monopoly with I as the downstream monopoly firm, resulting in quantity $q_I^u = Q^u = 1/4$. Welfare under this pricing regime then is given by

$$W^u = \int_0^{Q^u} (1-x)dx = \frac{7}{32}. \quad \text{(A.82)}$$

With $F > 0$, $W^d > W^u$ if and only if $\sqrt{F} < (1/3)-(8/9)k$. Obviously, for all $k \in (0,1/2)$ we have $(1/3)-(8/9)k < (1/3)-(2/3)k$. Moreover, for $k \in (0, 2-\sqrt{3}]$, $(1/3)-(8/9)k > (1/6)-(7/12)k+(\sqrt{1-4k+k^2})/12$. Last, note that $(1/3)-(8/9)k$ and $(1/6)-(1/3)k$ intersect at $k = 0.3$. Thus, $W^d > W^u$ if and only if $k < 0.3$ and $(1/6)-(1/3)k < \sqrt{F}$, $(1/6)-(7/12)k+(\sqrt{1-4k+k^2})/12 \leq \sqrt{F}$, and $\sqrt{F} < (1/3)-(8/9)k$.

(V) $k \geq (\sqrt{3}-1)/4$, $\sqrt{F} \leq (1/6)-(1/3)k$, and $\sqrt{F} \geq (1/6)-(7/12)k+(\sqrt{1-4k+k^2})/12$ for $k \in [(\sqrt{3}-1)/4, 2-\sqrt{3})$:

Under price discrimination, U implements an unconstrained duopoly resulting in quantities $q_I^d = (1+k)/6$, $q_E^d = (1-2k)/6$, and $Q^d = (2-k)/6$. Welfare under this pricing regime then is given by

$$W^d = \int_0^{Q^d} (1-x)dx - kq_E^d - F = \frac{20-20k+23k^2}{72} - F. \quad \text{(A.83)}$$

Under uniform pricing, on the other hand, U implements an unconstrained downstream monopoly with I as the downstream monopoly firm, resulting in quantity $q_I^u = Q^u = 1/4$. Welfare under this pricing regime then is given by

$$W^u = \int_0^{Q^u} (1-x)dx = \frac{7}{32}. \quad \text{(A.84)}$$

Thus, $W^d > W^u$ if and only if $F < (23/72)k^2 - (5/18)k + (17/288) =: F_W(k)$. Note that $F_W(k) > 0$ for $k < 17/46$ and $F_W(k) \leq 0$ for $k \in [17/46, 1/2]$. With

$F_W(k) > 0$ for $k < 17/46$, it is readily verified that $d\sqrt{F_W(k)}/dk < 0$ for $k \leq 17/46$. Moreover, $\sqrt{F_W(k)} = (1/6) - (1/3)k$ if and only if $k = 0.3$. Thus, $W^d \leq W^u$ if and only if $k \geq 0.3$ and $\sqrt{F_W(k)} \leq \sqrt{F} \leq (1/6) - (1/3)k$.

Last, for $\sqrt{F} \geq (1/3) - (2/3)k$, U implements a downstream monopoly with I as the downstream monopoly firm under both pricing regimes, resulting in quantity $q_I^d = q_I^u = Q^d = Q^u = 1/4$. Thus, there is no difference in welfare under both pricing regimes, $W^d = W^u$. Combining these observations establishes the desired result. ■

Bibliography

[1] **Abeler, J.**, **A. Falk**, **L. Götte**, and **D. Huffman (forthcoming):** Reference Points and Effort Provision, *American Economic Review*.

[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 (2006):** Commitment vs. Flexibility, *Econometrica*, Vol. 74, 365-396.

[5] **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.

[6] **Ariely, D.** and **K. Wertenbroch (2002):** Procrastination, Deadlines, and Performance: Self-Control by Precommitment, *Psychological Science*, Vol. 13, 219-224.

[7] **Armstrong, M. (2007):** Recent Developments in the Economics of Price Discrimination, in R. Blundell, W. Newey and T. Persson (eds.), *Advances in Economics and Econometrics: Theory and Applications: Ninth World Congress*, Camebridge University Press, Camebridge.

[8] **Arya, A.** and **B. Mittendorf (2004):** Using Job Rotation to Extract Employee Information, *Journal of Law, Economics, and Organization*, Vol. 20, 400-414.

[9] **Arya, A.** and **B. Mittendorf (2006):** Using Optional Job Rotation to Gauge On-the-Job Learning, *Journal of Institutional and Theoretical Economics*, Vol. 162, 505-515.

[10] **Barberis, N.**, **M. Huang** and **T. Santos (2001):** Prospect Theory and Asset Prices, *Quarterly Journal of Economics*, Vol. 116, 1-53.

[11] **Baron, J. (1991):** Beliefs about Thinking, in J.F. Voss, D.N. Perkins, and J.W. Segal (eds.), *Informal Reasoning and Education*, Hillsdale, NJ, Erlbaum, 169-186.

[12] **Baron, J. (1995):** Myside Bias in Thinking about Abortion, *Thinking and Reasoning*, Vol. 7, 221-235.

[13] **Bell, D.E. (1985):** Disappointment in Decision Making under Uncertainty, *Operations Research*, Vol. 33, 1-27.

[14] **Bénabou, R.** and **J. Tirole (2002):** Self-Confidence and Personal Motivation, *Quarterly Journal of Economics*, Vol. 117, 871-915.

[15] **Bertsekas, D. (1974):** Necessary and Sufficient Conditions for Existence of an Optimal Portfolio, *Journal of Economic Theory*, Vol. 8, 235-247.

[16] **Blackwell, D. (1951):** Comparison of Experiments, in J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley.

[17] **Blackwell, D. (1953):** Equivalent Comparison of Experiments, *Annals of Mathematics and Statistics*, Vol. 24, 265-272.

[18] **Bork, R. (1978):** *The Antitrust Paradox*, New York, Basic Books.

[19] **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.

[20] **Burke, L.A.** and **J.E. Moore (2000):** The Reverberating Effects of Job Rotation: A Theoretical Exploration of Nonrotaters' Fairness Perceptions, *Human Resource Management Review*, Vol. 10, 127-152.

[21] **Camerer, C.F.** and **U. Malmendier (2007):** Behavioral Economics of Organizations, in P. Diamond and H. Vartiainen (eds.), *Behavioral Economics and its Applications*, Princeton University Press, Princeton, 235-281.

[22] **Campion, M.A.**, **L. Cheraskin**, and **M.J. Stevens (1994):** Career-Related Antecedents and Outcomes of Job Rotation, *Academy of Management Journal*, Vol. 37, 1518-1542.

[23] **Caprice, S. (2006):** Multilateral Vertical Contracting with an Alternative Supply: The Welfare Effects of a Ban on Price Discrimination, *Review of Industrial Organization*, Vol. 28, 63-80.

[24] **Carmichael, L.** and **W.B. MacLeod (1993):** Multiskilling, Technical Change and the Japanese Firm, *Economic Journal*, Vol. 103, 142-160.

[25] **Carrillo, J.D.**, and **T. Mariotti (2000):** Strategic Ignorance as Self-Disciplining Device, *Review of Economic Studies*, Vol. 67, 529-544.

[26] **Churchill, G.A.**, **N.M. Ford**, and **O.C. Walker (1993):** *Sales Force Management*, Irwin, Homewood.

[27] **Cosgel, M.M.** and **T.J. Miceli (1999):** Job Rotation: Cost, Benefits, and Stylized Facts, *Journal of Theoretical and Institutional Economics*, Vol. 155, 301-320.

[28] **Crawford, V.P.** and **J. Meng (2009):** New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income, *Working Paper*, University of California, San Diego.

[29] **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.

[30] **Darley, J.M.** and **P.H. Gross (1983):** A Hypothesis-Confirming Bias in Labelling Effects, *Journal of Personality and Social Psychology*, Vol. 44, 20-33.

[31] **de Meza, D.** and **D.C. Webb (2007):** Incentive Design Under Loss Aversion, *Journal of the European Economic Association*, Vol. 5, 66-92.

[32] **DeGraba, P. (1990):** Input Market Price Discrimination and the Choice of Technology, *American Economic Review*, Vol. 80, 1246-1253.

[33] **DellaVigna, S. (2009):** Psychology and Economics: Evidence from the Field, *Journal of Economic Literature*, Vol. 47, 315-372.

[34] **DellaVigna, S.** and **U. Malmendier (2004):** Contract Design and Self-Control: Theory and Evidence, *Quartely Journal of Economics*, Vol. 119, 353-402.

[35] **DellaVigna, S.** and **U. Malmendier (2006):** Paying Not to Go to the Gym, *American Economic Review*, Vol. 96, 694-719.

[36] **DellaVigna, S.** and **M.D. Paserman (2005):** Job Search and Impatience, *Journal of Labor Economics*, Vol. 23, 527-588.

[37] **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.

[38] **Diamond, P.** and **B. Kőszegi (2003):** Quasi-Hyperbolic Discounting and Retirement, *Journal of Public Economics*, Vol. 87, 1839-1872.

[39] **Dittmann, I.**, **E. Maug** and **O.G. Spalt (forthcoming):** Sticks or Carrots? Optimal CEO Compensation when Managers are Loss Averse, *Journal of Finance.*

[40] **Eguchi, K. (2005):** Job Transfer and Influence Activity, *Journal of Economic Behavior and Organization*, Vol. 56, 187-197.

[41] **Eliaz, K.** and **R. Spiegler (2006):** Contracting with Diversely Naive Agents, *Review of Economic Studies*, Vol. 73, 689-714.

[42] **Ellison, G. (2006):** Bounded Rationality in Industrial Organization, in R. Blundell, W. Newey and T. Persson (eds.),*Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Cambridge University Press.

[43] **Eriksson, T** and **J. Ortega (2006):** The Adoption of Job Rotation: Testing the Theories, *Industrial and Labor Relations Review*, Vol. 59, 653-666.

[44] **Fischer, C. (1999):** Read this Paper Even Later: Procrastination with Time-Inconsistent Preferences, *Resources for the Future discussion paper 99-20.*

[45] **Fischer, C. (2001):** Read this Paper Later: Procrastination with Time-Consistent Preferences, *Journal of Economic Behavior and Organization*, Vol. 46, 249-269.

[46] **Fischhoff, B. (1977):** Perceived Informativeness of Facts, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 3, 349-358.

[47] **Frederick, S.**, **G. Loewenstein** and **T. O'Donoghue (2002):** Time Discounting and Time Preference, *Journal of Economic Literature*, Vol. 40, 351-401.

[48] **Gabaix, X.** and **D. Laibson (2006):** Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets, *Quarterly Journal of Economics*, Vol. 121, 505-540.

[49] **Gilpatric, S.M. (2008):** Present-Biased Preferences, Self-Awareness and Shirking, *Journal of Economic Behavior and Organization*, Vol. 67, 735-754.

[50] **Gittleman, M.**, **M. Horrigan** and **M. Joyce (1998):** Flexible' Workplace Practices: Evidence from a Nationally Representative Survey, *Industrial and Labor Relations Review*, Vol. 52 , 99-115.

[51] **Gjesdal, F. (1982):** Information and Incentives: The Agency Information Problem, *Review of Economic Studies*, Vol. 49, 373-390.

[52] **Grossman, S.J.** and **O.D. Hart (1983):** An Analysis of the Principal-Agent Problem, *Econometrica*, Vol. 51, 7-45.

[53] **Gruber, J.** and **B. Kőszegi (2001):** Is Addiction Rational? Theory and Evidence, *Quarterly Journal of Economics*, Vol. 116, 1261-1303.

[54] **Gul, F. (1991):** A Theory of Disappointment Aversion, *Econometrica*, Vol. 59, 667-686.

[55] **Haller, H. (1985):** The Principal-Agent Problem with a Satisficing Agent, *Journal of Economic Behavior and Organization*, Vol. 6, 359-379.

[56] **Haucap, J.** and **C. Wey (2007):** Input Price Discrimination (Bans), Entry and Welfare, *Working Paper*, DIW, Berlin.

[57] **Hausman, J.A.** and **J.K. MacKie-Mason (1988):** Price Discrimination and Patent Policy, *Rand Journal of Economics*, Vol. 19, 253-265.

[58] **Heidhues, P.** and **B. Kőszegi (2005):** The Impact of Consumer Loss Aversion on Pricing, *CEPR Discussion Paper*, No. 4849, http://ssrn.com/abstract=717706.

[59] **Heidhues, P.** and **B. Kőszegi (2008):** Competition and Price Variation when Consumers are Loss Averse, *American Economic Review*, Vol. 98, 1245–1268.

[60] **Henrion, M.** and **B. Fischhoff (1986):** Assessing Uncertainty in Physical Constants, *American Journal of Physics*, Vol. 54, 791-798.

[61] **Hoch, S.J. (1984):** Availability and Inference in Predictive Judgment, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 10, 649-662.

[62] **Hoch, S.J. (1985):** Counterfactual Reasoning and Accuracy in Predicting Personal Events, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 11, 719-731.

[63] **Holmström, B. (1979):** Moral Hazard and Observability, *Bell Journal Of Economics*, Vol. 10, 74-91.

[64] **Iantchev, E.P. (2009):** Risk or Loss Aversion? Evidence from Personnel Records, *Working Paper*, Syracuse University.

[65] **Ickes, B.W.** and **L. Samuelson (1978):** Job Transfers and Incentives in Complex Organizations: Thwarting the Ratchet Effect, *RAND Journal of Economics*, Vol. 18, 275-286.

[66] **Inderst, R.** and **G. Shaffer (2009):** Market Power, Price Discrimination, and Allocative Efficiency in Intermediate-Good Markets, *Rand Journal of Economics*, Vol. 40, 658-672.

[67] **Inderst, R.** and **T. Valetti (2009):** Price Discrimination in Input Markets, *RAND Journal of Economics*, VOL. 40, 1-19.

[68] **Inderst, R.** and **T. Valetti (forthcoming):** Buyer Power and the "Waterbed Effect", Journal of Industrial Economics.

[69] **Iyer, G.** and **J.M. Villas-Boas (2003):** A Bargaining Theory of Distribution Channels, *Journal of Marketing Research*, Vol. 40, 80-100.

[70] **Jewitt, I.**, **O. Kadan**, and **J.M. Swinkels (2008):** Moral Hazard with Bounded Payments, *Journal of Economic Theory*, Vol. 143, 59-82.

[71] **Joseph, K.** and **M.U. Kalwani (1998):** The Role of Bonus Pay in Salesforce Compensation Plans, *Industrial Marketing Management*, Vol. 27, 147-159.

[72] **Kahneman, D.** and **A. Tversky (1979):** Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, Vol. 47, 263-291.

[73] **Katz, M.L. (1987):** The Welfare Effects of Third-Degree Price Discrimination in Intermediate Good Markets, *American Economic Review*, Vol. 77, 154-167.

[74] **Katz, M.L. (1989):** Vertical Contractual Relationships, in R. Schmalensee and R.D. Willig (eds.), *The Handbook of Industrial Organization*, Amsterdam: North Holland Publishing.

[75] **Kidd, J.B. (1970):** The Utilization of Subjective Probabilities in Production Planning, *Acta Psychologica*, Vol. 34, 338-347.

[76] **Kim, S.K. (1995):** Efficiency of an Information System in an Agency Model, *Econometrica*, Vol. 63, 89-102.

[77] **Kim, S.K. (1997):** Limited Liability and Bonus Contracts, *Journal of Economics & Management Strategy*, Vol. 6, 899-913.

[78] **Koriat, A.**, **S. Lichtenstein** and **B. Fischhoff (1980):** Reasons for Confidence, *Journal of Experimental Psychology: Human Learning and Memory*, Vol. 6, 107-118.

[79] **Kőszegi, B.** and **M. Rabin (2006):** A Model of Reference-Dependent Preferences, *Quarterly Journal of Economics*, Vol. 121, 1133-1165.

[80] **Kőszegi, B.** and **M. Rabin (2007):** Reference-Dependent Risk Preferences, *American Economic Review*, Vol. 97, 1047-1073.

[81] **Kőszegi, B.** and **M. Rabin (2009):** Reference-Dependent Consumption Plans, *American Economic Review*, 99, 909-936.

[82] **Kuhn, D. (1989):** Children and Adults as Intuitive Scientists, *Psychological Review*, Vol. 96, 674-689.

[83] **Laibson, D. (1996):** Hyperbolic Discount Functions, Undersaving, and Savings Policy, *NBER Working Paper Series*, No.5635, Cambridge, MA.

[84] **Laibson, D. (1997):** Golden Eggs and Hyperbolic Discounting, *Quarterly Journal of Economics*, Vol. 112, 443-477.

[85] **Laibson, D. (1998):** Life-Cycle Consumption and Hyperbolic Discount Functions, *European Economic Review*, Vol. 42, 861-871.

[86] **Laibson, D.**, **A. Repetto** and **J. Tobacman (1998):** Self-Control and Saving for Retirement, *Brookings Papers on Economic Activity*, Vol. 1, 91-196.

[87] **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.

[88] **Lazear, E.P.** and **P. Oyer (2007):** Personnel Economics, *NBER Working Paper 13480*, http:// www.nber.org/papers/w13480.

[89] **Lingle, J.H.** and **T.M. Ostrom (1981):** Principles of Memory and Cognition in Attitude Formation, in R.E. Petty, T.M. Ostrom, and T.C. Brock (eds.), *Cognitive Responses in Persuasive Communications: A Text in Attitude Change*, Hillsdale, NJ, Erlbaum, 399-420.

[90] **Loftus, E.F.** and **W.A. Wagenaar (1988):** Lawyers' Predictions of Success, *Jurimetrics Journal*, Vol. 28, 437-453.

[91] **Loomes, G.** and **R. Sugden (1986):** Disappointment and Dynamic Consistency in Choice under Uncertainty, *Review of Economic Studies*, Vol. 53, 271-282.

[92] **Lord, C.**, **L. Ross** and **M.R. Lepper (1979):** Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence, *Journal of Personality and Social Psychology*, Vol. 37, 2098-2109.

[93] **Ma, C.A. (1988):** Implementation in Dynamic Job Transfers, *Economics Letters*, Vol. 28, 391-395.

[94] **MacLeod, W.B. (2003):** Optimal Contracting with Subjective Evaluation, *American Economic Review*, Vol. 93, 216-240.

[95] **MacLeod, W.B. (2007):** Reputations, Relationships, and Contract Enforcement, *Journal of Economic Literature*, Vol. 45, 595-628.

[96] **Malueg, D.A. (1993):** Bounding the Welfare Effects of Third-Degree Price Discrimination, *American Economic Review*, Vol. 83, 1011-1021.

[97] **Mellers, B.**, **A. Schwartz** and **H. Ritov (1999):** Emotion-Based Choice, *Journal of Experimental Psychology: General*, Vol. 128, 332-345.

[98] **Milliou, C.**, **E. Petrakis** and **N. Vettas (2004):** (In)efficirent Trading Forms in Competing Vertical Chains, mimeo.

[99] **Moynahan, J.K. (1980):** *Designing an Effective Sales Compensation Program*, AMACOM, New York.

[100] **Mynatt, C.R.**, **M.E. Doherty** and **R.D. Tweney (1977):** Confirmation Bias in a Simulated Research Environment: An Experimental Study of Scientific Inferences, *Quarterly Journal of Experimental Psychology*, Vol. 29, 85-95.

[101] **Nickerson, R.S. (1998):** Confirmation Bias: A Ubiquitous Phenomenon in Many Guises, *Review of General Psychology*, Vol. 2, 175-220.

[102] **Nisbett, R.E.** and **L. Ross (1980):** Human Inference: Strategies and Shortcomings of Social Judgment, Englewood Cliffs, NJ, Prentice Hall.

[103] **Nocke, V.** and **M. Peitz (2003):** Hyperbolic Discounting and Secondary Markets, *Games and Economic Behavior*, Vol. 44, 77-97.

[104] **O'Brien, D.P. (forthcoming):** The Welfare Effects of Third-Degree Price Discrimination in Intermediate Good Markets: The Case of Bargaining, *RAND Journal of Economics.*

[105] **O'Brien, D.P.** and **G. Shaffer (1994):** The Welfare Effects of Forbidding Discriminatory Discounts: A Secondary-Line Analysis of Robinson-Patman, *Journal of Law, Economics, and Organization*, Vol. 10, 296-318.

[106] **O'Donoghue, T.** and **M. Rabin (1999a):** Addiction and Self-Control, in *Addiction: Entry and Exits*, J. Elster, editor, Russel Sage Foundation.

[107] **O'Donoghue, T.** and **M. Rabin (1999b):** Doing It Now Or Later, *American Economic Review*, Vol. 89, 103-124.

[108] **O'Donoghue, T.** and **M. Rabin (1999c):** Incentives for Procrastinators, *Quarterly Journal of Economics*, Vol. 114, 769-816.

[109] **O'Donoghue, T.** and **M. Rabin (2001a):** Addiction and Present-Biased Preferences, *working paper*, University of California, Berkeley.

[110] **O'Donoghue, T.** and **M. Rabin (2001b):** Choice and Procrastination, *Quarterly Journal of Economics*, Vol. 116, 121-160.

[111] **O'Donoghue, T.** and **M. Rabin (2005):** Incentives and Self-Control, *working paper*, University of California, Berkeley.

[112] **O'Donoghue, T.** and **M. Rabin (2008):** Procrastination on Long-Term Projects, *Journal of Economic Behavior and Organization*, Vol. 66, 161-175.

[113] **OECD (1999):** OECD Employment Outlook, Paris, Organization for Economic Cooperation and Development.

[114] **Ortega, J. (2001):** Job Rotation as a Learning Mechanism, *Management Science*, Vol. 47, 1361-1370.

[115] **Oskamp, S. (1965):** Overconfidence in Case Study Judgments, *Journal of Consulting Psychology*, Vol. 29, 261-265.

[116] **Osterman, P. (1994):** How Common Is Workplace Transformation and Who Adopts It?, *Industrial and Labor Relations Review*, Vol. 47, 173-188.

[117] **Osterman, P. (2000):** Work Reorganization in an Era of Restructuring: Trends in Diffusion and Effects on Employee Welfare, *Industrial and Labor Relations Review*, Vol. 53, 179-196.

[118] **Oyer, P. (1998):** Fiscal Year Ends and Non-Linear Incentive Contracts: The Effect on Business Seasonality, *Quarterly Journal of Economics*, Vol. 113, 149-188.

[119] **Oyer, P. (2000):** A Theory of Sales Quotas with Limited Liability and Rent Sharing, *Journal of Labor Economics*, Vol 18, 405-426.

[120] **Park, E.-S. (1995):** Incentive Contracting under Limited Liability, *Journal of Economics & Management Strategy*, Vol. 4, 477-490.

[121] **Perkins, D.N.**, **R. Allen** and **J. Hafner (1983):** Difficulties in Everyday Reasoning, in W. Maxwell (ed.), *Thinking: The Frontier Expands*, Philadelphia, Franklin Institute Press, 83-105.

[122] **Perkins, D.N.**, **M. Farady** and **B. Bushey (1991):** Everyday Reasoning and the Roots of Intelligence, in J.F. Voss, D.N. Perkins, and J.W. Segal (eds.), *Informal Reasoning and Education*, Hillsdale, NJ, Erlbaum, 83-106.

[123] **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.

[124] **Pitz, G.F.**, **L. Downing** and **H. Reinhold (1967):** Sequential Effects in the Revision of Subjective Probabilities, *Canadian Journal of Psychology*, Vol. 21, 381-393.

[125] **Pollak, R.A. (1968):** Consistent Planning, *Review of Economic Studies*, Vol. 35, 201-208.

[126] **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.

[127] **Prendergast, C. (1999):** The Provision of Incentives in Firms, *Journal of Economic Literature*, Vol. 37, 7-63.

[128] **Prendergast, C.** and **R. Topel (1993):** Discretion and Bias in Performance Evaluation, *European Economic Review*, Vol. 37, 355-365.

[129] **Rabin, M. (1998):** Psychology and Economics, *Journal of Economic Literature*, Vol. 36, 11-46.

[130] **Rabin, M. (1998):** Risk Aversion and Expected-Utility Theory: A Calibration Theorem, *Econometrica*, Vol. 68, 1281-1292.

[131] **Rabin, M.** and **J.L. Schrag (1999):** First Impressions Matter: A Model of Confirmatory Bias, *Quarterly Journal of Economics*, Vol. 114, 37-82.

[132] **Rayo, L.** and **G.S. Becker (2007):** Evolutionary Efficiency and Happiness, *Journal of Political Economy*, Vol. 115, 302-337.

[133] **Ross, L. (1977):** The Intuitive Psychologist and his Shortcomings: Distortions in the Attribution Process, in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Orlando, FL, Academic Press, Vol. 10, 174-221.

[134] **Ross, L.** and **C. Anderson (1982):** Shortcomings in the Attribution Process: On the Origins and Maintenance of Erroneous Social Assessments, in D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgments under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press, 129-152.

[135] **Ross, L.** and **M.R. Lepper (1980):** The Perseverance of Beliefs: Empirical and Normative Considerations, in R. Shweder and D. Fiske (eds.), *New Directions for Methodology of Social and Behavioral Science: Fallible Judgment in Behavioral Research*, San Francisco, Jossey-Bass, Vol. 4, 17-36.

[136] **Ross, L.**, **M. Lepper** and **M. Hubbard (1975):** Perseverance in Self Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm, *Journal of Personality and Social Psychology*, Vol. 32, 880-892.

[137] **Salanié, B. (2003):** Testing Contract Theory, *CESifo Economic Studies*, Vol. 49, 461-477.

[138] **Schaeffer, R.G. (1983):** Staffing Systems: Managerial and Professional Jobs, Elsevier Science Publisher, Amsterdam.

[139] **Schmalensee, R. (1981):** Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination, *American Economic Review*, Vol. 71, 242-247.

[140] **Schmidt, U. (1999):** Moral Hazard and First-Order Risk Aversion, *Journal of Economics*, Supplement 8, 167-179.

[141] **Schwartz, B. (1982):** Reinforcement-Induced Behavioral Stereotype: How Not to Teach People to Discover Rules, *Journal of Experimental Psychology: General*, Vol. 111, 23-59.

[142] **Schwartz, M. (1990):** Third-Degree Price Discrimination and Output: Generalizing a Welfare Result, *American Economic Review*, Vol. 80, 1259-1262.

[143] **Sherman, S.J.**, **K.S. Zehner**, **J. Johnson** and **E.R. Hirt (1983):** Social Explanation: The Role of Timing, Set, and Recall on Subjective Likelihood Estimates, *Journal of Personality and Social Psychology*, Vol. 44, 1127-1143.

[144] **Skiba, P.M.** and **J. Tobacman (2008):** Payday Loans, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default, *working paper*, University of Pennsylvania.

[145] **Steenburgh, T.J. (2008):** Effort or Timing: The Effect of Lump-Sum Bonuses, *Quantitative Marketing and Economics*, Vol. 6, 235-256.

[146] **Stole, L. (2007):** Price Discrimination and Imperfect Competition, in M. Armstrong and R. Porter (eds.), *Handbook of Industrial Organization : Volume III* , North-Holland, Amsterdam.

[147] **Strausz, R. (2006):** Deterministic Versus Stochastic Mechanisms in Principal-Agent Models, *Journal of Economic Theory*, Vol. 128, 306-314.

[148] **Strotz, R.H. (1956):** Myopia and Inconsistency in Dynamic Utility Maximization, *Review of Economic Studies*, Vol. 23, 156-180.

[149] **Tetlock, P.E.** and **J.I. Kim (1987):** Accountability and Judgment Processes in a Personality Prediction Task, *Journal of Personality and Social Psychology*, Vol. 52, 700-709.

[150] **Tobacman, J. (2007):** The Partial Naivete Euler Equation, *working paper*, University of Pennsylvania.

[151] **Tversky, A.** and **D. Kahneman (1991):** Loss Aversion in Riskless Choice: A Reference-Dependent Model, *Quarterly Journal of Economics*, Vol. 106, 1039-1061.

[152] **Valetti, T.M. (2003):** Input Price Discrimination with Downstream Cournot Competitors, *International Journal of Industrial Organization*, Vol. 21, 969-988.

[153] **Varian, H.R. (1985):** Price Discrimination and Social Welfare, *American Economic Review*, Vol. 75, 870-875.

[154] **Vives, X. (1999):** *Oligopoly Pricing: Old Ideas and New Tools*, Camebridge, Mass.: MIT Press.

[155] **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.

[156] **Yaari, M.E. (1987):** The Dual Theory of Choice under Risk, *Econometrica*, Vol. 55, 95-115.

[157] **Yoshida, Y. (2000):** Third-Degree Price Discrimination in Input Market: Output and Welfare, *American Economic Review*, Vol. 90, 240-246.

[158] **Zábojník, J. (2002):** The Employment Contract as a Lottery, *Working Paper*, University of Southern California.

[159] **Zuckerman, M.**, **R. Knee**, **H.S. Hodgins** and **K. Miyake (1995):** Hypothesis Confirmation: The Joint Effect of Positive Test Strategy and Acquiescence Response Set, *Journal of Personality and Social Psychology*, Vol. 68, 52-60.