Inequity Aversion and Incentives

Three Essays in Microeconomic Theory

Inaugural-Dissertation
zur Erlangung des Grades Doctor oecomomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2004

vorgelegt von
Ferdinand von Siemens

Referent: Prof. Sven Rady, Ph.D.
Korreferent: Prof. Dr. Klaus M. Schmidt
Promotionsabschlussberatung: 21. Juli 2004

Acknowledgements

First and foremost, I would like to thank my supervisor Sven Rady for his superb research support, encouragement and patience. I am also indebted to Klaus Schmidt and my co-author Björn Bartling, who were and hopefully continue to be a great source of inspiration and motivation.

Equally, I would like thank my colleagues Manuela Beckstein, Georg Gebhardt, Florian Herold, and Markus Reisinger, who provided me with a very pleasurable work and research environment at the Ludwig-Maimilians-Universität in Munich. Many other friends and colleagues contributed comments and suggestions. Too numerous to be mentioned by name I thank all of them.

Furthermore, I would like to thank the Faculty of Economics and Politics and Darwin College at Cambridge University for the hospitality they offered me as a visiting Ph.D. student in the academic year 2001-2002. Financial support from the European Commission, grant no. HMPT-CT-2000-00056, is gratefully acknowledged.

Last but not least, I dedicate my thesis to my wife Nicole. Only her loving support and inexhaustible patience made all of this possible.

Contents

Preface

In recent years an overwhelming experimental and empirical evidence has put mounting pressure on the economic paradigm that individuals should be modelled as exclusively interested in their own material well-being. By now many economists would agree that apart from monetary concerns individuals may also be driven by altruism, reciprocity, a desire for equal or fair distributions of resources, or status considerations. This has prompted the development of numerous theories aiming at a better understanding of individual conduct. Appreciation for these new psychological approaches has cumulated in the award of the 2003 Nobel price to Daniel Kahneman and Vernon Smith. Behavioral and experimental economics have thus reached the mainstream. At the same time the theory of incentives has been continuously deepening our understanding of contracts and institutions battling the unfavorable effects of asymmetric information. However, only relatively few theoretical models investigate the impact of psychology on incentives and the design of institutions. The present thesis adds to this literature by looking at the interaction of a particular behavioral trait, inequity aversion, with asymmetric information and incentives in the context of bilateral trade of a customized good, employment decisions and contracts, and team incentives.

The theory of inequity aversion developed by Fehr and Schmidt (1999) has attracted enormous attention. The authors essentially assume that although most individuals are 'selfish' and only care for their own payoff, at least some individuals are 'inequity averse' and suffer a utility loss if their payoff is not equal to the payoffs of the other individuals in their reference group. Moreover, individuals differ with respect to their concerns for equity and individual preferences are private information. Yet apart from adding some realism to economic modelling, will Fehr and Schmidt (1999) have a lasting impact on economics? Moreover, should their theory of inequity aversion be adopted and incorporated into economic models?

Whereas only time can respond to the first, there are criteria to be met for an affirmative answer to the second question. Economics is an empirical social science. A new theory should thus be adopted if and only if it better explains the empirical evidence than the existing theories. Fehr and Schmidt (1999) pass this test as the generated predictions are consistent with an almost surprisingly large number of different experiments. This stems from a subtle influence of the strategic environment on the consequences of inequity aversion.

Inequity averse individuals want to reduce inequity. If benefits exceed costs, an inequity averse individual is thus willing to sacrifice some of his payoff in order to help another individual with a lower payoff. Yet if another individual gets a higher payoff, an inequity averse individual may be willing to take costly actions if this decreases the other's payoff by even more. Finally, if an inequity averse individual can hardly influence the overall inequality of the payoff distribution, the best he can do is maximize his own payoff. Inequity averse individuals can thus exhibit altruistic, envious or selfish behavior. Specifically, inequity aversion is consistent with the empirical evidence on experiments concerning competition and markets. As summarized by Smith (1982) and Davis and Holt (1993) markets tend to work astonishingly well under even the most adverse conditions. For a long time these experiments were seen as support for the assumption of rational, purely self-interested individuals. Yet Fehr and Schmidt (1999) can also explain this empirical evidence while providing a better fit to experiments refuting the self-interest hypothesis. Thanks to the above characterized dependency on the strategic environment, Fehr and Schmidt (1999) qualifies for an interesting methodological augmentation of microeconomic theory.

As part of the strategic situation asymmetric information can strongly influence and amplify the effect of inequity aversion. For example, inequity averse and selfish individuals might contribute to the provision of a public good if and only if there are some inequity averse individuals around willing to inflict costly punishments on free-riders.[1] If preferences are not observable, the suspicion that some individuals might be inequity averse can be sufficient to sustain cooperation in a group actually consisting only of selfish individuals. Such an interaction of inequity aversion and asymmetric information is especially interesting as the latter is often prevalent in situations in which inequity aversion should have the largest impact. Asym-

[1] See Fehr and Gächter (2000) for an experimental investigation.

metric information can give individuals some - possibly undesirable - freedom of action which they can use to 'get even'. However, even absent inequity aversion asymmetric information and diverging interests can generate efficiency losses. The corresponding contractual remedies are investigated by the theory of incentives. Together with the development of highly powerful theoretical tools, this has generated profound insights into the functioning and internal organization of institutions. Yet as frequently noted by Ernst Fehr, Klaus Schmidt, and their co-authors, the application of inequity aversion promises to be especially fruitful if individual decisions - for example effort choices or the revelation of important information - have a strong impact on payoff distributions. The present thesis takes up this point of view and explores the interaction of inequity aversion with asymmetric information and incentives.

Chapter 1 investigates the impact of inequity aversion on the efficiency of incomplete contracts. The model is set in the context of the famous hold-up problem pioneered by Grout (1982), Grossman and Hart (1986), and Hart and Moore (1990). Consider a buyer and a seller who can trade some good. The good has been customized by costly relation-specific investments of the traders so that gains of trade are especially high if that particular buyer trades with that particular seller. If the trade contract signed by buyer and seller is incomplete and therefore leaves important contractual details unspecified, both traders will try to exploit this leeway by bargaining over terms of trade after the relation-specific investments are sunk. Anticipating this buyer and seller know that they will get only part of the proceeds from their investments but incur all individual investment costs. This impairs ex ante incentives to maximize gains from trade, which constitutes the hold-up problem. However, although most contracts are highly incomplete in reality, bilateral trade is often accomplished in a reasonably efficient way. This contradicts the theory of incentives.

As the anticipated bargaining outcome determines investment incentives, bargaining plays a crucial role in the hold-up problem. Since bargaining is also important in many other areas, it has been subjected to numerous experimental tests. Indeed, the 'ultimatum game' is probably the best known experiment in economics. However, the empirical results summarized by Güth, Schmittberger, and Tietz (1990), Kagel and Roth (1995) or Camerer and Thaler (1997) strongly refute the self-interest hypothesis and call for an inclusion of psychology into economics. Yet if bargaining outcomes are affected by the psychological characteristics of the negotiants, these characteristics influence investment decisions in the hold-up problem.

Consequently, Chapter 1 is based on the assumption that individuals might be inequity averse and - equally important - individual preferences are unobservable. Inequity averse individuals prefer a bargaining breakdown as compared to trade at inequitable, unfair conditions. If preferences are observable, their opponents account for this and inequity averse individuals ultimately get a better deal than a selfish individual. This improves the formers' investment incentives, but as long as they do not have all the bargaining power even inequity averse individuals do not reap all proceeds from their investment and thus invest inefficiently little. Moreover, the insufficient investment incentives of selfish individuals are not affected.

If preferences are private information, the effect of inequity aversion changes. If a trader is sufficiently convinced that his bargaining opponent is inequity averse, he behaves as if the opponent was inequity averse. Appearing to be inequity averse is thus sufficient to increase a trader's bargaining power, and inequity aversion can affect investment incentives for selfish and inequity averse traders. Additionally, bargaining in the hold-up problem is preceded by the individuals' relation-specific investments. If preferences are private information, these investments can be a signalling device. Chapter 1 determines the conditions under which there exists a pooling equilibrium in which both inequity averse and selfish traders invest efficiently in order not to appear selfish and thus loose bargaining power in the ensuing bargaining. Inequity aversion and asymmetric information about preferences might therefore explain why in reality trade is often accomplished efficiently even though contracts are incomplete.

Whereas Chapter 1 looks at bilateral trade, Chapters 2 and 3 investigate the influence of inequity aversion on employment decisions, wage schemes, and the organization of the firm. Although involuntary unemployment is one of the most important economic problems, economists still have a hard time trying to explain why wages do not adjust as to equate supply and demand in the labor market. Blinder and Choi (1990), Bewley (1995), Campbell and Kamlani (1997), and Bewley (1999) investigate what determines a firm's employment decisions. They find that employers believe that remuneration schemes deteriorate working morale if workers perceive them as unfair. Thus, employers try to avoid 'unfair' behavior, especially treating people in similar occupations differently by paying them very different wages. This nicely fits the empirical observation that wages are often compressed as compared to differences in productivity. Moreover, the business literature on 'organizational behavior' has collected a multitude of case studies and empirical investigations suggesting that fairness considerations matter at the workplace.[2] Yet, if cooperation amongst workers, and if the workers' relationship with their employers is influenced by social comparisons, then firms should account for this in their employment decisions, wage schemes, and internal organization.

Chapter 2 studies a single firm's optimal employment decision and optimal employment contracts if workers are inequity averse. Further, it investigates the interaction of inequity aversion with incomplete information about the workers' productivity. It turns out that the impact of inequity aversion delicately depends on what workers compare.

If workers compare only their income, inequity aversion causes an income compression and - possibly - an exclusion of low-skilled workers. Employed workers differing in their productivity are usually given different income levels. If the firm wants to employ the workers with the initially lower income, these workers have to be compensated for their suffering. These additional costs can only be reduced by reducing the income differences. Inequity aversion thus compresses income levels. Moreover, adjusting employment contracts to the reduced income differences creates distortions affecting the entire workforce. If these distortions are sufficiently high, the firm might decide not to employ low-skilled workers. Inequity aversion therefore offers an explanation for an unwillingness of firms to employ low-skilled workers. Note that the above argument hinges on initial differences in income reflecting the workers' differences in productivity. As this is independent of whether productivity is observable or not, so is the impact of inequity aversion on employment decisions and contracts. The findings are consistent with the existing theoretical literature incorporating fairness concerns and social comparisons into the labor market, most notably Akerlof (1982), Frank (1984a), Frank (1984b), Akerlof and Yellen (1988), and Akerlof and Yellen (1990).

[2]See Rotemberg (2002) for a survey on this literature.

However, Chapter 2 continues to show that the above characterized consequences of inequity aversion are reduced or even reversed if workers start to incorporate production costs in their comparisons. Suppose workers perfectly account for production costs and compare rents defined as income minus production costs. Contrary to the above argument asymmetric information then interacts strongly with inequity aversion. If the firm can observe its prospective workers' productivity, it extracts all rents. Since thus all workers get an even rent of zero, inequity aversion is irrelevant. However, if each worker's productivity is private information, workers with high productivity must get an informational rent to induce truthful revelation of their type. Asymmetric information thus generates a rent inequality which causes additional costs if all types of workers are to be employed. The rent inequality can only be reduced by lowering the informational rent of the highly productive workers. As this informational rent depends on what a highly productive worker can get by pretending to be less productive, production of the less productive workers must be lowered. Since these workers then incur lower production costs, the firm may reduce their income by more than the decrease in the informational rent of the highly skilled workforce. Inequity aversion thus increases income differences. Moreover, the impact of inequity aversion can be made arbitrarily small while at the same time diminishing the distortions imposed on highly productive workers. Eventually the increase in profit from production of the less productive workers exceeds the costs of asymmetric information and inequity aversion. Hence, it is always optimal to keep the low-skilled workers in the firm. These results stand in a stark contrast to the impact of inequity aversion if workers compare only income, and thus contradict the existing theoretical literature cited above.

In Chapter 2 inequity aversion influences a worker's decision whether to accept a certain employment contract. However, once employed the worker must produce a certain quantity and has no leeway to align his actions with his fairness concerns. Fehr and Schmidt (1999) indicate that inequity aversion has a larger impact if individuals can 'get even' on their own account, for example by reducing their unobservable effort provision in a problem of moral hazard. Consequently, Chapter 3 explores incentive provision in teams if each team member's effort choice is not verifiable and individuals are inequity averse.

Very broadly speaking a team can be defined as a group of individuals whose individual payoffs are tied to the group's joint performance. Teams can emerge in many different situations, for example a firm collectively owned by its workers, or a group of employees in a large public company using team remuneration schemes. In any case, if the team's output depends on its members' effort contributions and individual effort choices are not observable - and thus cannot be sanctioned individually - some individuals might try to free-ride on the effort provision of the others. This can severely impair incentives to contribute effort or cooperate. Alchian and Demsetz (1972) and Holmström (1982) show that the contractual counter-measures influence the organization of the firm. However, the use of teams with comparatively simple contracts is more widely spread and successful as to be expected from the contract theoretical literature ensuing the above seminal articles.

Inequity aversion might offer an explanation. If individuals within a team compare their payoff with the payoffs of the other team members, this can have a positive or a negative effect on cooperation. If all the others provide effort, an individual might also contribute effort although the increase in his monetary payoff does not cover the effort costs. The reason for this is that he dislikes getting more than the others, but this is exactly what happens when he cheats the team by shirking. Effort provision may thus be facilitated. However, an inequity averse individual strongly resents the others free-riding on his own effort provision. He might thus decide not to work if all the other team members shirk although the increase in his monetary payoff would then exceed his effort costs. Social comparisons can consequently aggravate incentive provision.

Which effect prevails depends on the contracts connecting each individual's payment scheme to the joint performance of the group. Chapter 3 derives the optimal contracts accounting for inequity aversion within a team of any size. It turns out that if agents are sufficiently inequity averse, efficient effort choices can be implemented with relatively simple, budget-balancing sharing rules. Building upon this result, the consequences of inequity aversion can be further analyzed.

First, if the individual impact on joint performance decreases with team size whereas effort costs remain constant, free-riding becomes generally more attractive the larger the team.

What prevents individuals from shirking is the behavioral trait of shame from cheating. If team size increases, more people are cheated but each individual counts less in the comparisons. Consequently shame from cheating stays constant, and cooperation becomes harder to sustain with increasing team size. This fits the common observation that small teams often work well whereas larger ones suffer from free-riding.

Second, even though inequity aversion reduces the free-rider effect restricting firm size, it can have a negative impact on the optimal size of worker-owned firms. If a new inequity averse worker is employed in a worker-owner firm, he compares himself with the incumbent owners. In order to ensure participation he must receive a fair share of the joint profit in excess of his effort costs. If a firm is owned by an outsider not incorporated in the team members' comparisons, he can extract all rents without causing inequality. A new worker then accepts an employment contract granting him zero rent. Thus, a worker-owned firm has comparatively higher costs of employing an additional worker. Chapter 3 investigates this argument and presents a simple example in which inequity aversion prevents the efficient hiring of an additional worker. Summarizing, the tools developed in Chapter 3 may be helpful to further increase our understanding of the internal organization of the firm. If working in a team has a psychological impact on incentives and effort provision, this should be accounted for by firms when deciding on the composition and size of teams, divisions, and hierarchies.

Chapter 1

Incomplete Contracts and the Hold-Up Problem

"Almost every economist would agree that actual contracts are or appear quite incomplete."
Tirole (1999), page 741.

1.1 Introduction

The study of incentive problems caused by asymmetric information has generated important and interesting insights into the functioning of institutions. As one of the main conclusions contracts designed to minimize efficiency losses should be 'complete' in the sense of using all relevant and contractible information.[1] According to this definition contracts in the real world are often highly incomplete. Surprisingly these incomplete contracts work reasonably well. Does this observation contradict incentive theory? The present paper argues that the answer to this question is 'No' if social preferences and - equally important - incomplete information about an individual's preferences are taken into account. If preferences are unobservable and information about the latter influences strategic behavior, individuals might optimally choose certain - efficient - actions in order not to reveal unfavorable information. However, adding asymmetric information can generate new inefficiencies.

[1] See, for example, Holmström (1982) or Laffont and Tirole (1993).

The hold-up problem - first developed by Grout (1982), Grossman and Hart (1986), and Hart and Moore (1990) - is the traditional tool for investigating the efficiency of incomplete contracts. In a simplified version consider a buyer and a seller who can trade one unit of a good. The good's quality is determined by a costly investment of the seller. This investment is relation-specific and increases only this particular buyer's value from purchasing the good. Suppose the contract governing the traders' relationship is incomplete so that the buyer can renegotiate terms of trade once the seller has sunk his investment. The seller knows that the buyer will exploit the situation and thus anticipates that he cannot reap all proceeds from his investment. Consequently he reduces his effort and provides a good of inefficiently low quality. This constitutes the hold-up problem. However, in reality incomplete contracts do not always have such inefficient consequences: the buyer often pays the promised price, and trusting in this the seller provides good quality.

The present paper offers an explanation based on the following three assumptions. First, individuals differ in some characteristics affecting their bargaining behavior. Second, these heterogeneous characteristics are unobservable. And thirdly, individuals are aware of the above and account for it when determining their own bargaining behavior.

All three assumptions are confirmed by numerous experiments investigating simple bargaining games, and they form the basic ingredients of an enormous and influential literature on bargaining under incomplete information including such pioneering articles include Crawford (1982) and Fudenberg and Tirole (1983). Yet the most prominent bargaining experiment, the famous 'ultimatum game', has received enormous attention for an additional reason. Apart from providing supplementary empirical evidence for the above assumption, it seems to refute the standard economic paradigm that all individuals exclusively care for their own payoff.[2] Amongst other experiments the ultimatum game has thus prompted the development of a new branch of economics. The associated theories, termed behavioral economics in a broader context, aim at a better understanding of the motivations underlying individual conduct. The present paper directly draws upon this literature for motivating the heterogeneity in the traders' bargaining behavior.

[2]See Güth, Schmittberger, and Tietz (1990).

Amongst the emerging plethora of models in behavioral economics[3], the theory of inequity aversion developed by Fehr and Schmidt (1999) is one of the most successful exponents. In their very simple and applicable model these authors argue that although most individuals are 'selfish' and exclusively interested in their own payoff, some individuals are 'inequity averse' and also care for the equality of payoff distributions. Combined with private information about these preferences, the impact of their theory on individual behavior depends delicately on the strategic environment. Hence, the empirical predictions of their theory fit the empirical evidence of a wide class of experiments surprisingly well.

The present paper invokes Fehr and Schmidt (1999) in order to model the traders' bargaining behavior in the hold-up problem. There are the following results. Inequity averse sellers reject proposals that result in inequitable distributions. As they do so even if this causes a bargaining breakdown they bargain more aggressively than selfish sellers. If preferences are observable, the other negotiant - in the present model the buyer - accounts for this and inequity averse sellers get a higher share of the trade surplus. Thus, they have better investment incentives. Yet since not even inequity averse sellers get the entire trade surplus they still under-invest. Investment incentives for selfish sellers are unchanged.

However, the key argument of the model stems from an interaction of inequity aversion and incomplete information about an individual's preferences. If preferences are unobservable, the implications of inequity aversion change. Since in the hold-up problem bargaining is preceded by the investment, the latter might signal the seller's preferences. Depending on the considered perfect Bayesian equilibrium both inequity averse and selfish sellers might optimally choose a certain investment in order not to 'reveal' unfavorable information about their preferences harming them in the ensuing bargaining. If the efficient investment can be supported as the equilibrium choice in such a pooling equilibrium, investment incentives are efficient for both types. Yet, signalling is no cheap talk in the model, and amongst other, highly inefficient perfect Bayesian equilibria there also exist separating equilibria.

Even though no scientific argument, the idea of the model is nicely illustrated by the following little anecdote from the author's personal experience. After moving into his new house, a

[3] For a survey of recent developments see Fehr and Schmidt (2003).

buyer commissioned a mason to embellish his entrance area. Since work was mostly carried out on Sundays and this is usually illicit in Germany, there was probably no legal contract so that taxes can be avoided. After the entrance had been completed the buyer refused to pay *anything*.[4] As the mason had no possibility to take legal action without revealing the illegal nature of his work, the story so far is in nice accordance with the standard hold-up problem. However, the mason's reaction is note-worthy and fits well into the present model. During a particulary dark night he emptied buckets full of concrete on the newly paved doorsteps. This criminal act was clearly costly even ignoring the risk of potential legal consequences. Thus, his behavior confirms that some people react strongly whenever they feel cheated. Yet, even more interesting than the story itself is how it was spread. The mason in person circulated the news amongst his colleagues and customers. Perhaps he tried to build up some reputation for being a 'tough' guy, but this in turn would suggest that individuals have a strategic interest in transmitting certain information about their type.

Several other paper look at the impact of incomplete information or social preferences on the efficiency of incomplete contracts. Gul (2001) investigates the hold-up problem if the buyer's investment and thus his valuation is private information. Moreover, the seller makes all offers in a bargaining game with infinite horizon. By an argument similar to the Coase conjecture, the buyer invests efficiently and the good is traded immediately if time between offers converges to zero. However, inefficiencies disappear only in the limit, and psychology or inequity aversion play no role.

The following articles are more closely related in as far as they analyze the interaction of inequity aversion and incomplete contracts. Fehr and Schmidt (2000) and Fehr, Klein, and Schmidt (2001) argue that since complete contracts restrict reciprocal behavior they might crowd out reciprocal incentives. In Englmaier and Wambach (2003) the agent compares himself with the principal in a moral hazard problem. As concerns for inequity become infinitely important, output must be divided equally independently of other informative signals. Contracts are thus rendered incomplete. Nevertheless, none of these articles studies inequity aversion in the context of the hold-up problem, and there is no bargaining or signalling.

[4] No, the author is not the buyer.

Finally, the following authors consider the interaction of social preferences and the hold-up problem. Carmichael and MacLeod (2003) start off by investigating the evolution of bargaining conventions and social preferences with the help of the hold-up problem. Yet, in their model sunk costs directly influence bargaining outcomes and preferences are observable. Asymmetric information, however, lies at the heart of Ellingsen and Johannesson (2002b). In their model unobservable investment costs combined with inequity aversion can generate inefficiencies, thus worsening the hold-up problem. The same authors demonstrate in Ellingsen and Johannesson (2002a) that instead of aggravating incentive problems, fairness can improve the situation by providing a focal point for the division of the trade surplus, thus diminishing coordination failures during the bargaining. Last but not least, Fehr, Kremhelmer, and Schmidt (2003) look at the impact of inequity aversion on the optimal allocation of ownership rights. However, none of the above models captures the idea that investments might signal a trader's preferences, and that traders might invest efficiently in order not to appear to be selfish and thus loose bargaining power in the ensuing division of the trade surplus.

The paper is organized as follows. Section 1.2 presents the basic model, introduces inequity aversion, characterizes the bargaining game and sequence of actions, and defines the efficient investment. Section 1.3 describes the main results. First, it shows that a complete contract can achieve efficiency. It then analyzes the optimal bargaining behavior of buyer and seller in any perfect Bayesian equilibrium. Next it explains the impact of the equilibrium refinements 'equilibrium dominance' and 'intuitive criterion' on the set of possible pure strategy perfect Bayesian equilibria. The following subsection demonstrates the existence of pooling equilibria and elaborates on the impact of incomplete information by investigating the unique subgame-perfect equilibrium if the seller's preferences are observable. Section 1.3 closes by studying the set of existing separating equilibria thus showing that the signalling game allows for a separation of types and signalling is no cheap talk. Section 1.4 summarizes the results.

1.2 The Model

1.2.1 Investment, Bargaining, and Sequence of Actions

Consider a buyer and a seller who can trade one unit of a good. The good's quality is determined by a costly investment of the seller. The ensuing model analyzes the seller's investment incentives if final terms of trade are not characterized in a trade contract but instead determined by bargaining after investment costs have been sunk. In order to study the interaction of investment and bargaining, the bargaining process is modeled explicitly as a two-stage alternating-offer bargaining game. Together with the investment stage the model consists of three periods.

Period 1: The seller chooses the quality i of the good. Producing a good of quality $i \in [0, \infty[$ causes the seller investment costs i. The buyer makes no investment.

Period 2: The seller starts the bargaining by proposing - and thus claiming - a share s_1 of the trade surplus. The trade surplus characterized below depends on the seller's prior investment. The buyer may accept or reject the seller's offer. If he accepts, the good is traded, the seller receives the share s_1 of the trade surplus, and the buyer consumes the good. If the buyer rejects s_1, the good is not traded and buyer and seller proceed to the next period.

Period 3: The buyer proposes a share s_2 of the trade surplus to be given to the seller. The seller may accept or reject. If he accepts, the good is traded, the seller receives the share s_2 of the trade surplus, and the buyer consumes the good. If the seller rejects s_2, bargaining breaks down and the seller consumes the good himself.

1.2.2 Buyer's and Seller's Preferences

Let $\phi(i)$ denote the buyer's value from purchasing the good, and let $\psi(i)$ denote the seller's value from consuming the good himself. Alternatively $\psi(i)$ may be regarded as the price the seller receives when selling the good at an external market and not to the buyer. Moreover,

Assumption 1.1

1. *ϕ and ψ are differentiable and strictly increasing in i.*

2. *ϕ and ψ are strictly concave, and $\phi''(i) \geq \psi''(i)$ for all i.*

3. *$\phi(0) = \psi(0) = 0$ and $\phi(i) > \psi(i)$ for all $i > 0$.*

4. *$\psi'(0) > 1/\delta$ and $\lim_{i\to\infty}\phi'(i) < 1$.*

5. *$\phi'(i) \geq \psi'(i)$ and for all i.*

Part 1 and 2 of the above assumption render all maximization problems well behaved. As seen at a later stage, this requires $\phi''(i) \geq \psi''(i)$ to guarantee the concavity of the seller's objective functions which can depend negatively on the trade surplus $\phi(i) - \psi(i)$. Part 3 captures the idea that the seller's investment is relation-specific. Trading the good generates a positive trade surplus $\phi(i) - \psi(i)$ which is lost if the seller consumes the good himself. Part 4 implies that most maximization problems have strictly positive solutions. The parameter $\delta \in]0, 1[$ is a discount factor introduced below to capture costs of bargaining. Finally, Part 5 epitomizes the hold-up problem. According to the notion of efficiency defined below a seller chooses an inefficiently low investment unless he receives the entire trade surplus.

The payoffs of buyer and seller are defined as follows. If there is trade, denote by s the share of the trade surplus $\phi(i) - \psi(i)$ the seller receives in excess of his outside option $\psi(i)$. There are no restrictions on s which can thus characterize any terms of trade. The buyer's preferences are defined as follows.

Assumption 1.2 *Consider a good of quality i. If the good is traded and the seller receives share s of the trade surplus, the utility of the buyer is given by*

$$v(i, s) := (1 - s)\,[\phi(i) - \psi(i)].$$

If the good is not traded, the buyer receives a utility normalized at zero.

The buyer's utility exclusively depends on whether the good is traded, and if yes, according to what terms of trade.

Concerning the seller this paper invokes Fehr and Schmidt (1999). According to their theory of inequity aversion some individuals care for relative payoffs. However, as a key element of their model they assume that individuals have heterogeneous preferences and that individual preferences are unobservable. In the present model this is captured by the assumption that the seller can be of two types. There is an *inequity averse seller* who compares himself with the buyer whenever the good is traded, and there is a *selfish seller* who is exclusively interested in his own well-being. More precisely,

Assumption 1.3 *Consider a good of quality i. If the good is traded and the seller receives share s of the trade surplus, the utility of the selfish seller is given by*

$$u_s(i,s) := s\,[\phi(i) - \psi(i)] + \psi(i),$$

whereas the utility of the inequity averse seller is given by

$$u_a(i,s) := u_s(i,s) - \alpha \max[(1-2s)(\phi(i) - \psi(i)), 0]$$

with $\alpha > 0$. If the seller consumes the good, his utility is $\psi(i)$ independent of his type.

The subindices s and a stand for selfish and inequity averse, correspondingly. Preferences are unobservable but the commonly known prior probability for a seller to be inequity averse is $\pi \in]0,1[$. The precise functional form (linearity and kink) is taken from the theory of inequity aversion by Fehr and Schmidt (1999). The parameter α measures the importance of inequity concerns of the inequity averse seller. In the present model he wants the trade surplus, whose realization he facilitates by refraining from consuming the good himself, to be divided equally. If the good is traded and the buyer's payoff is higher than the seller's allotment of the trade surplus, an inequity averse seller suffers an utility loss. The selfish seller has no concerns for inequity. This application of inequity aversion is based upon two implicit assumptions. First, an inequity averse seller does not suffer if his allotment of the trade surplus is higher than the buyer's payoff. And second, even an inequity averse seller does not care for the buyer's payoff if the good is not traded.

There are proportional costs of bargaining, buyer and seller discount when progressing from Period 2 to Period 3. For notational simplicity, there is no discounting between Period 1 and 2. Buyer and seller share a common discount factor $\delta \in]0,1[$ and maximize discounted utility.

Finally, define

$$u_0(i) := \delta\,\psi(i) - i \quad \text{and} \quad U_0 := \max_i\{u_0(i)\} \tag{1.1}$$

as a seller's utility from himself consuming a good of quality i at the end of Period 3, and U_0 is the corresponding maximum utility after choosing the optimal investments. These utilities do not depend on the seller's type. Note that U_0 must be strictly positive as $u_0(0)$ is zero and $u_0'(0) > 0$ by Assumption 1.1.

1.2.3 Perfect Bayesian Equilibrium and Signalling

As shown below the seller's bargaining behavior depends on his preferences. As preferences are private information, bargaining takes place under incomplete information. The relevant equilibrium notion is the concept of *perfect Bayesian equilibrium*. In a perfect Bayesian equilibrium each player's strategy must be optimal given his type, his beliefs about the other player's types, and the other player's type-dependent strategy. Whenever possible beliefs have to be consistent with the other player's equilibrium strategy. Thus, the buyer's belief must be determined by Bayes' rule after observing an action played with positive probability in equilibrium. However, Bayes' rule cannot be applied when observing an out-of-equilibrium action. Beliefs may then be chosen at will.

The present model is a two-dimensional signalling game. The seller's signal is his investment i in Period 1 and his proposal s_1 at the beginning of Period 2. Let $\mu(i, s_1)$ denote the probability with which the buyer believes the seller to be inequity averse after observing (i, s_1). The buyer's response is his acceptance decision in Period 2 and his proposal s_2 in Period 3. As the seller may not take any action in the mean time the buyer's belief remains unchanged.[5] The next action of the seller is his acceptance decision at the end of Period 3 but this action ends the game. Consequently, only the seller's investment and proposal are a signal, and only the buyer's prior π and (possibly) updated belief $\mu(i, s_1)$ are important in any perfect Bayesian equilibrium. Note that since the seller's signal is two-dimensional and the buyer's belief $\mu(i, s_1)$ thus depends on two actions, the seller might change the buyer's belief by changing his proposal while keeping the investment fixed.

[5]Especially, the seller may not not consume the good in Period 2.

1.2.4 Efficient Investment

Carmichael and MacLeod (2002) point out that concerns for relative payoffs might cause conceptual problems in a welfare analysis. In the present model efficiency is defined in the following way. First, since the trade surplus is positive for all investment levels the good should always be traded. Second, as buyer and both types of seller loose from delaying trade the good should be traded at the earliest possible moment, and therefore in Period 2. Building upon these characteristics the *efficient investment* i_e is defined as

$$i_e := \arg\max_i \{\phi(i) - i\}. \tag{1.2}$$

Thus, the efficient investment maximizes the 'net pie' generated if the good is traded at the earliest moment. This captures the idea that an allocation should be considered efficient if it maximizes the pie to be distributed. However, if the seller is inequity averse and gets less than half the trade surplus, the above defined efficient investment i_e is not Pareto optimal in the usual sense. Indeed, increasing the investment increases the trade surplus. Keeping the buyer's utility constant his share in the trade surplus can be reduced whereas the seller's share increases. Thus, inequity and the thereby caused suffering is reduced. This utility increase has a first order effect on the inequity averse seller whereas the increase in investment costs causes only a second order loss in his 'standard utility'. Therefore, the inequity averse seller's utility increases without reducing the utility if the buyer.

Abstracting from these complications suppose the good is always traded in Period 2 and the seller receives share s_1 of the trade surplus. If s_1 is constant, the inequity averse seller maximizes

$$s_1[\phi(i) - \psi(i)] + \psi(i) - i - \alpha \max[1 - 2s_1, 0]\,[\phi(i) - \psi(i)] \tag{1.3}$$

with respect to his investment. By Assumption 1.1 he chooses an inefficiently low investment unless he gets the entire trade surplus. The same holds true if the seller is selfish. This constitutes the hold-up problem.

1.2.5 Complete Contracts

As a benchmark case suppose buyer and seller can sign a complete trade contract conditioning terms of trade on the seller's investment. Not surprisingly,

Proposition 1.1 *There exist a complete contract so that both types of seller invest efficiently and the good is always traded in Period 2.*

Proof: Consider the following complete contract conditioning on the seller's investment. "If the seller chooses the efficient investment i_e, seller and buyer trade in Period 2 and the seller gets the share

$$s_c := \min \left\{ s : s \geq 1/2 \ \wedge \ s\,[\phi(i_e) - \psi(i_e)] + \psi(i_e) - i_e \geq \max_i \{\delta\psi(i) - i\} \right\} \tag{1.4}$$

of the trade surplus. If the seller chooses any other investment, $i \neq i_e$, seller and buyer may not trade." This contract has the following impact on the seller's incentives. As $s_c \geq 1/2$, both types of seller receive the same utility $s_c[\phi(i_e) - \psi(i_e)] + \psi(i_e) - i_e$ if the efficient investment i_e is chosen. For any other investment i, both types of seller receive an utility of $\delta\psi(i) - i$. If an s_c satisfying the above property exists, the above contract gives both types of seller the right investment incentives. It further implies that the good is always traded in Period 2. The remainder of the proof shows the existence of s_c.

The utility both types of seller get when investing efficiently is strictly increasing in s_c and converges to $\phi(i_e) - i_e$ as s_c converges to 1. By definition of i_e and $\delta < 1$

$$\phi(i_e) - i_e > \phi(i) - i > \delta\psi(i) - i \quad \forall\, i \neq i_e. \tag{1.5}$$

By continuity and strict monotonicity of the seller's utility function there thus exists a unique $s_c \in [1/2\,, 1[$ satisfying the property required in the contract. *Q.E.D.*

A complete contract can forbid trade if the seller's investment is inefficient. This restricts the seller to a binary choice. Either he invests efficiently, the good is traded, and he gets a (possible very large) share of the trade surplus. Or he invest inefficiently, the good is not traded, and he receives nothing of the trade surplus. By definition the efficient investment maximizes the net pie which can be distributed between buyer and seller. Therefore there must exist a share of the trade surplus so that the seller receives at least half the trade surplus and investment incentives are efficient.

Proposition 1.1 implies that voluntarily leaving a contract incomplete can never improve incentives. The present model thus argues that buyer and seller should write a complete contract whenever this is possible. Yet, in some situations this might not be the case. For example, renegotiation might render the above complete contract not credible. If the seller has invested inefficiently, both buyer and seller could agree to trade instead of foregoing the trade surplus as prescribed by the contract. The remainder of the paper concentrates on the cases where buyer and seller have, for whatever reason, not signed a complete contract.

1.3 Results

1.3.1 Equilibrium Bargaining Behavior

As the bargaining game is finite it can be analyzed by backwards induction. In any perfect Bayesian equilibrium buyer and seller behave in the following way. If the seller rejects the buyer's proposal s_2 in Period 3, bargaining breaks down and he receives nothing in excess of his utility $\psi(i)$ from consumption. A selfish seller thus accepts all weakly positive proposals by the buyer, but an inequity averse seller prefers a bargaining breakdown to trade at very inequitable conditions. Therefore, he rejects proposals that are strictly positive but smaller than a certain cutoff. Accounting for the seller's behavior the buyer has two potentially optimal strategies. Either he bargains aggressively and proposes $s_2 = 0$. In this case he reaps the entire trade surplus if the seller accepts but trades only if the seller is selfish. Alternatively the buyer chooses the proposal making the inequity averse seller exactly indifferent between accepting and rejecting. If the inequity averse seller always accepts in case of difference, the buyer then reaps only part of the trade surplus but trades with certainty. Any other proposal cannot be optimal as the buyer could then increase his proposal and get a higher utility if the good is traded without reducing the probability of trade. Which strategy is optimal for the buyer depends on his belief. If the buyer is sufficiently convinced to be facing an inequity averse seller, he should make the high proposal resulting in certain trade. Otherwise he should claim the entire trade surplus even though trade then occurs only if the seller is selfish.

The anticipated bargaining behavior in Period 3 determines the buyer's acceptance decision in Period 2. After observing the seller's investment i and proposal s_1 the buyer forms his belief $\mu(i, s_1)$. This belief does not change in case of rejecting the seller's proposal. Moreover, the buyer's expected utility from himself proposing in Period 3 is weakly decreasing in his belief. If $\mu(i, s_1)$ is high, the buyer makes the high proposal and thus gets little of the trade surplus. If $\mu(i, s_1)$ is low, the buyer bargains aggressively but his expected utility decreases with the probability $\mu(i, s_1)$ with which he expects the seller to be inequity averse and reject. Therefore, the maximum proposal s_1 acceptable to the buyer is increasing in his belief $\mu(i, s_1)$. Including the precise cutoffs the above argument is summarized in the following lemma.

Lemma 1.1 *In any perfect Bayesian equilibrium it is optimal*

1. *for a selfish seller to accept all* $s_2 \geq 0$, *and*

2. *for an inequity averse seller to accept all* $s_2 \geq \alpha/(1 + 2\alpha)$.

Defining $\tilde{\mu} := \alpha/(1 + 2\alpha)$, *it is optimal for the buyer*

3. *iff* $\mu(i, s_1) \geq \tilde{\mu}$, *to set* $s_2 = \alpha/(1 + 2\alpha)$ *and accept* $s_1 \leq 1 - \delta(1 + \alpha)/(1 + 2\alpha)$,

4. *iff* $\mu(i, s_1) < \tilde{\mu}$, *to set* $s_2 = 0$ *and accept* $s_1 \leq 1 - \delta[1 - \mu(i, s_1)]$,

if the seller always accepts s_2 *in case of indifference.*

Proof: The game can be analyzed by backwards induction. First, consider the seller in Period 3. The selfish seller accepts the buyer's proposal s_2 if and only if $s[\phi(i) - \psi(i)] + \psi(i) \geq \psi(i)$. Thus, he accepts all $s_2 \geq 0$. The inequity averse seller accepts s_2 if and only if $s_2[\phi(i) - \psi(i)] + \psi(i) - \alpha \max[(1 - 2s_2)[\phi(i) - \psi(i)], 0] \geq \psi(i)$. Thus, he accepts s_2 if and only if $s_2 \geq \alpha/(1 + 2\alpha)$ where this cutoff strictly increases in α and lies in the interval $]0, 1/2[$ for all $\alpha \geq 0$.

Second, consider the buyer holding belief $\mu(i, s_1)$ in Period 3. Suppose the seller accepts if he is indifferent.[6] Thus, if the buyer proposes $s_2 = \alpha/(1 + 2\alpha)$, he trades with certainty. His payoff is then $[\phi(i) - \psi(i)](1 + \alpha)/(1 + 2\alpha)$. If the buyer proposes $s_2 = 0$, he gets the entire trade surplus in case of trade, but trades only if the seller is selfish. His expected payoff is

[6]Otherwise, there might not exist an optimal proposal of the buyer.

then $[1-\mu(i,s_1)][\phi(i)-\psi(i)]$. Any proposal $s_2 \in]\alpha/(1+2\alpha),1]$ or $s_2 \in]0,\alpha/(1+2\alpha))]$ cannot be optimal as the buyer can then increase his payoff by lowering s_2 without decreasing the probability of trade. Choosing between two alternatives only, it is optimal for the buyer to propose $s_2 = \alpha/(1+2\alpha)$ if and only if $[\phi(i)-\psi(i)]\cdot(1+\alpha)/(1+2\alpha) \geq [1-\mu(i,s_1)]\cdot[\phi(i)-\psi(i)]$, and there exists a unique cutoff $\tilde{\mu} = \alpha/(1+2\alpha)$ strictly increasing in α and lying in the interval $[0,1/2[$ making the buyer indifferent between the two proposals. If $\mu(i,s_1) > \tilde{\mu}$, the buyer should propose $s_2 = \alpha/(1+2\alpha)$, and if $\mu(i,s_1) < \tilde{\mu}$, he should propose $s_2 = 0$.

After analyzing Period 3, consider the buyer holding belief $\mu(i,s_1)$ in Period 2. Whether the buyer accepts the seller's proposal s_1 depends on what he expects to get in Period 3 by rejecting. This ultimately depends on his belief $\mu(i,s_1)$. First, suppose $\mu(i,s_1) > \tilde{\mu}$. After rejecting s_1, the buyer proposes $\alpha/(1+2\alpha)$ in Period 3 which yields payoff $[\phi(i)-\psi(i)]\cdot(1+\alpha)/(1+2\alpha)$. Therefore, the buyer accepts if and only if $(1-s_1)\cdot[\phi(i)-\psi(i)] \geq \delta[\phi(i)-\psi(i)]\cdot(1+\alpha)/(1+2\alpha)$, or equivalently $s_1 \leq 1-\delta(1+\alpha)/(1+2\alpha)$. Second, suppose $\mu(i,s_1) < \tilde{\mu}$. After rejecting s_1, the buyer proposes $s_2 = 0$ in Period 3 which yields an expected payoff $[1-\mu(i,s_1)]\cdot[\phi(i)-\psi(i)]$. Therefore, the buyer accepts if and only if $(1-s_1)\cdot[\phi(i)-\psi(i)] \geq \delta[1-\mu(i,s_1)]\cdot[\phi(i)-\psi(i)]$, or equivalently $s_1 \leq 1-\delta[1-\mu(i,s_1)]$. Finally, if $\mu(i,s_1) = \tilde{\mu}$, both above cutoffs are identical, $1-\delta(1+\alpha)/(1+2\alpha) = 1-\delta[1-\mu(i,s_1)]$. *Q.E.D.*

The seller can thus get a larger share of the trade surplus if he appears to be inequity averse. However, even if the buyer believes the seller to be inequity averse, the seller might not be able to get more than half the trade surplus. In this case the ensuing equilibrium analysis is greatly complicated without adding new insights or changing the idea of the argument. The following assumption averts tedious case distinctions.

Assumption 1.4 $1-\delta(1+\alpha)/(1+2\alpha) > 1/2$.

Thus, it is assumed that the seller can get more than half the trade surplus if the buyer believes him to be inequity averse with certainty. Note that the sunk investment costs i do not influence the suffering from inequity and thus have no impact on the seller's acceptance decision in Period 3. Consequently they do not alter any of the cutoffs in Lemma 1.1. Apart from possibly signalling the seller's preferences and thereby changing the buyer's belief, "sunk costs are sunk".

The remainder of this section describes some general characteristics of perfect Bayesian equilibria of the considered bargaining game. In any such equilibrium the seller must be discouraged to deviate and choose an out-of-equilibrium investment. Given any deviation the least favorable response of the buyer is rejecting the seller's proposal and claiming the entire trade surplus in Period 3. However, in the spirit of subgame perfection the concept of perfect Bayesian equilibrium requires the buyer's equilibrium response to comply with Lemma 1.1.

This imposes the following restrictions. Define a seller's *minimum deviation utility* as the supremum of utility levels he can get by choosing an out-of-equilibrium investment i under the following three conditions. First, the buyer holds the least favorable belief and believes the seller to be selfish with certainty. Second, given this belief the buyer chooses the least favorable equilibrium strategy as characterized in Lemma 1.1. He thus always rejects the proposal s_1 in case of indifference. Thirdly, given the buyer's unfavorable belief and strategy the seller chooses the optimal proposal s_1 and acceptance decision in Period 3.

According to this definition the seller's minimum deviation utility is only a function of the investment and does not depend on the buyer's belief, the buyer's strategy, and the seller's proposal and acceptance decision. Out-of-equilibrium there is no reason why the buyer should not reject the proposal s_1 in case of indifference. Thus, the minimum deviation utilities can only be approximated arbitrarily closely but never fully attained.

Lemma 1.2 *In any perfect Bayesian equilibrium*

1. *the* **selfish seller's minimum deviation utility** *is given by*

   $u_s^m(i) := (1-\delta)[\phi(i)-\psi(i)]+\psi(i)-i$ *with* u_s^m *strictly concave, and*

2. *the* **inequity averse seller's minimum deviation utility** *is given by*

   $u_a^m(i) := \max\left[\, u_0(i),\, u_s^m(i) - \alpha \max[2\delta-1,0]\,[\phi(i)-\psi(i)\,]\right],$

for given investment i.

Proof: For given investment i consider the following strategy of the seller: "Propose $s_1 = 1$ in Period 3 and reject any proposal s_2 in Period 3." If the buyer accepts, the seller receives $\phi(i) - i > u_0(i)$. If the buyer rejects, the seller rejects all proposals s_2 and gets utility $u_0(i)$. As $s_1 = 1 > 1/2$ and by Assumption 1.3 the seller's utility is independent of his type. The seller can thus always get at least utility $u_0(i)$ for any investment i.

However, given any investment i the selfish seller can set $s_1 = 1-\delta-\epsilon$ with ϵ strictly positive but arbitrarily close to zero. In any perfect Bayesian equilibrium the buyer's behavior is characterized by Lemma 1.1. The buyer thus accepts $s_1 = 1-\delta-\epsilon$ for all beliefs $\mu(i, s_1) \geq 0$. As ϵ goes to zero, $\phi(i) > \psi(i)$ for all $i > 0$ implies $(1-\delta)[\phi(i) - \psi(i)] + \psi(i) - i > \delta\psi(i) - i$. Furthermore, $u(0, s_1) = 0 = u_0(0)$ for all s_1 including $s_1 = 1-\delta-\epsilon$. Thus, for very small ϵ the selfish seller always weakly prefers to propose $1-\delta-\epsilon$ and trade in Period 2 as compared to consuming the good in Period 3. As ϵ goes to zero, his utility converges to $u_s^m(i)$.

Finally, consider an inequity averse seller. By the previous argument the seller can trade at $s_1 = 1-\delta-\epsilon$ giving him a utility of $u_s^m(i) - \alpha \max[2\delta + 2\epsilon - 1, 0]\,[\phi(i) - \psi(i)]$. Alternatively he can consume the good and receive $u_0(i)$. With ϵ going to zero his minimum deviation utility is the maximum of these two alternatives. *Q.E.D.*

The seller can always consume the good himself or trade by proposing s_1 marginally less than $1-\delta$. The selfish seller always prefers to trade at these conditions, but the inequity averse might prefer not to trade if his share $1-\delta$ of the trade surplus is smaller than one half and the associated suffering from inequity sufficiently large. The last lemma of this section concerns the inequity averse seller's equilibrium utility.

Lemma 1.3 *In any perfect Bayesian equilibrium the inequity averse seller receives a strictly positive but weakly smaller equilibrium utility than the selfish seller.*

Lemma 1.3 follows directly from the definition of the seller's utility functions. In any perfect Bayesian equilibrium the equilibrium strategies of the inequity averse seller and the buyer result in a - possibly degenerate - lottery over outcomes. If the selfish seller mimics the inequity averse seller, he receives the same lottery. However, given any outcome he enjoys a weakly higher expected utility as he gets the same rent but never suffers from inequity. In any perfect Bayesian equilibrium the selfish seller's equilibrium utility must be weakly higher than what he can get by deviating. Thus, the selfish seller's equilibrium utility must be weakly higher than the inequity averse seller's equilibrium utility. Moreover, Lemma 1.2 implies that the inequity averse seller can always get U_0, which is strictly larger than zero.

1.3.2 Pure Strategy Equilibria and Equilibrium Dominance

Although the results in the previous section hold for mixed strategy perfect Bayesian equilibria the remainder of the paper concentrates on pure strategies. The following notation and definitions will be used frequently. In a generic pure strategy perfect Bayesian equilibrium let (i^*, s_1^*) denote the inequity averse seller's, and let (i^{**}, s_1^{**}) denote the selfish seller's equilibrium investment and proposal. Given the yet unspecified equilibrium strategy and beliefs of the buyer the inequity averse seller receives equilibrium utility U_a^* and the selfish seller gets equilibrium utility U_s^*.

With pure strategies there are only two classes of equilibria, *pooling equilibria* and *separating equilibria*. Define a pooling equilibrium as a pure strategy perfect Bayesian equilibrium in which both types of seller choose the same equilibrium investment and proposal which, for notational simplicity, is denoted by the inequity averse seller's equilibrium choice (i^*, s_1^*). Bayes rule implies that the buyer does not learn anything about the seller's type in equilibrium, $\mu(i^*, s_1^*) = \pi$. Next, define a separating equilibrium as a pure strategy perfect Bayesian equilibrium in which the inequity averse and selfish seller choose different equilibrium actions, that is, either $i^* \neq i^{**}$, or $s_1^* \neq s_1^{**}$, or both. The buyer then learns the seller's type in equilibrium, $\mu(i^*, s_1^*) = 1$ and $\mu(i^{**}, s_1^{**}) = 0$.

The freedom in setting out-of-equilibrium beliefs usually results in a multiplicity of perfect Bayesian equilibria. As a response there exists a large literature proposing equilibrium refinements. The most commonly accepted equilibrium refinement is the *intuitive criterion* which corresponds to the concept of *equilibrium dominance* if there are only two types.[7] In the context of the present model equilibrium dominance is defined as follows. After out-of-equilibrium action (i, s_1) the buyer's *response* is his acceptance decision and proposal s_2 in Period 3. An *equilibrium response* is a response optimal in the sense of Lemma 1.1 for at least one belief. Action (i, s_1) is *equilibrium dominated* for a certain type of seller if he receives strictly less than his equilibrium utility for all the buyer's equilibrium responses. Given any out-of-equilibrium action a buyer's belief is *reasonable* if he assigns a type of seller zero probability if the action is equilibrium dominated for this but not for the other type. If

[7] For a more general exposition see Fudenberg and Tirole (1991) or Mas-Colell, Whinston, and Green (1995).

the action is equilibrium dominated for both or no type of seller, all beliefs are reasonable. A perfect Bayesian equilibrium complies with equilibrium dominance - and thus satisfies the intuitive criterion - if all out-of-equilibrium beliefs are reasonable.

The following examples might serve as an illustration. First, rejecting a proposal s_1 strictly smaller then $1 - \delta$ is not an equilibrium response as it is not optimal for all possible beliefs. Second, suppose that given the most favorable equilibrium responses by the buyer the inequity averse seller gets more whereas the selfish seller gets less than his equilibrium utility. In this case the buyer's belief is reasonable if he believes the seller to be inequity averse with certainty. Equilibrium dominance has the following implications.

Lemma 1.4 *In any pure strategy perfect Bayesian equilibrium with reasonable beliefs both types of seller receive the same equilibrium utility.*

Proof: The proof works by contradiction. Consider a pure strategy perfect Bayesian equilibrium with reasonable beliefs in which the selfish seller receives an equilibrium utility strictly higher than the inequity averse seller, $U_s^* > U_a^*$.

In the following it is shown that there then exists a deviation (i', s_1') so that given the most favorable equilibrium response of the buyer the selfish seller receives strictly less whereas the inequity averse seller receives strictly more than his equilibrium utility. Given reasonable beliefs the buyer must believe the seller to be inequity averse with certainty after observing (i', s_1'). The deviation is chosen so that given any equilibrium response to this belief the inequity averse seller profits from deviating. This completes the contradiction, which thus implies $U_a^* \geq U_s^*$. Together with $U_s^* \geq U_a^*$ from Lemma 1.3 this shows that both types of seller must receive the same equilibrium utility.

Consider a pure strategy perfect Bayesian equilibrium in which the inequity averse seller chooses (i^*, s_1^*) and $U_s^* > U_a^*$. Consider the following deviation (i', s_1') with

$$s_1' = 1 - \delta(1 + \alpha)/(1 + 2\alpha) - \epsilon, \tag{1.6}$$

where ϵ is strictly positive but sufficiently small so that $s_1' > 1/2$. The investment i' is left unspecified for the time being.

Accepting upon observing (i', s_1') is an equilibrium response as it is optimal for $\mu(i', s_1') = 1$ by Lemma 1.1. As $s_1' > 1/2$ both types of seller then get the same utility denoted by

$$u^d(i') := s_1'[\phi(i') - \psi(i')] + \psi(i') - i'. \tag{1.7}$$

In case of rejecting s_1' Lemma 1.1 implies that the buyer's most generous equilibrium proposal is $s_2' = \alpha/(1+2\alpha)$. Consider the inequity averse and the selfish seller in turn. First, the inequity averse seller then gets $u_0(i')$ which is strictly smaller than $u^d(i')$ as $s_1' > 1/2$ and $\delta < 1$. Second, the selfish seller then gets $\delta\, u_s(i', \alpha/(1+2\alpha)) - i$ which is strictly smaller than $u^d(i')$ for sufficiently small ϵ. Choose ϵ sufficiently small in that sense. Note that $\delta < 1$ then implies $s_1' > \delta\alpha/(1+2\alpha)$. Thus, the most favorable equilibrium response of the buyer after observing the deviation is accepting, giving both types of seller a utility of $u^d(i')$.

The function u^d has the following properties. By Assumption 1.1 it is strictly concave with unique maximizer $i^d := \arg\max\{u^d(i)\}$. Note that i^d depends on ϵ. The maximum equilibrium proposal acceptable to the buyer is given by $s_1' + \epsilon$. Thus, $\lim_{\epsilon\to 0} u^d(i^d) \geq U_s^*$ if the selfish seller trades the good in Period 2. If the buyer rejects the selfish seller's proposal in Period 2, his maximum offer in Period 3 is $s_2 = \alpha/(1+\alpha)$ yielding the selfish seller a utility of $\delta\, u_s(i, \alpha/(1+\alpha)) - i$ strictly smaller than $u^d(i)$ for all investments i and sufficiently small ϵ. As consequently $\lim_{\epsilon\to 0} u^d(i^d) \geq U_s^*$, one can find an ϵ sufficiently small so that $u^d(i^d) > U_s^* - \epsilon'$ for all $\epsilon' > 0$. For the remainder of the proof choose ϵ and ϵ' sufficiently small so that $u^d(i^d) > U_s^* - \epsilon'$ and $U_s^* - \epsilon' > U_a^*$.

It is next shown that there then exists a non-empty, open interval $\mathcal{A} :=]\underline{i}, \bar{i}[$ with strictly positive lower boundary $\underline{i}$ so that either $u^d(\underline{i}) = U_a^*$ and $u^d(\bar{i}) = U_s^* - \epsilon'$, or alternatively $u^d(\underline{i}) = U_s^* - \epsilon'$ and $u^d(\bar{i}) = U_a^*$. Given an interval with the above properties, continuity of $u^d(i)$ implies that for all $C \in]U_a^*, U_s^* - \epsilon'[$ there exists at least one $i' \in \mathcal{A}$ so that $u^d(i') = C$ and therefore

$$U_a^* < u^d(i') < U_s^* - \epsilon' < U_s^*. \tag{1.8}$$

This interval can be constructed as follows. Note that $u^d(0) = 0 < U_0 \leq U_a^* < U_s^* - \epsilon'$ and $u^d(i^d) \geq U_s^* > U_a^*$. By the continuity of u^d there thus exist $x \in]0, i^d[$ and $y \in]0, i^d[$ so that $u^d(x) = U_a^*$ and $u^d(y) = U_s^* - \epsilon'$. These boundaries x and y must be different as otherwise their definition implies $U_a^* = u^d(x) = u^d(y) = U_s^* - \epsilon'$ which contradicts $U_a^* < U_s^* - \epsilon'$. Defining $\mathcal{A} :=]x, y[$ if $x < y$ and $\mathcal{A} :=]y, x[$ if $x > y$, there thus exists a non-empty interval $\mathcal{A}$ with strictly positive lower boundary and the above properties. This in turn shows the existence of a deviation investment i' with the desired properties.

If the considered Bayesian equilibrium has reasonable beliefs, the buyer must belief the seller to be inequity averse with certainty after observing (i', s_1'). By Lemma 1.1 he always accepts. Given this belief and strategy the inequity averse seller deviates to (i', s_1'), and the above cannot form an equilibrium. *Q.E.D.*

Key to Lemma 1.4 is the dimensionality of the seller's signal. The proof shows that in any pure strategy perfect Bayesian equilibrium with reasonable beliefs there must exist a profitable deviation for the inequity averse seller if he receives a lower equilibrium utility than the selfish seller. Consider a deviation with the maximum proposal so that acceptance is an equilibrium response of the buyer. The most favorable equilibrium response of the buyer is acceptance, and - as the deviation proposal is larger than one half - the corresponding maximum deviation utility is the same for both types. Keeping the deviation proposal fixed the maximum deviation utility can be adjusted by changing the deviation investment. The proof shows that there exists an investment so that the maximum deviation utility lies in between the two equilibrium utilities of the two types of seller. If the buyer's beliefs are reasonable he must believe the seller to be inequity averse after observing this deviation, and accept. This makes the deviation profitable. As by Lemma 1.3 the inequity averse seller receives a weakly smaller equilibrium utility than the seller seller, this completes the proof. Lemma 1.4 has a number of important implications.

Lemma 1.5 *In any pure strategy perfect Bayesian equilibrium with reasonable beliefs the buyer may believe the seller to be selfish with certainty after observing an out-of-equilibrium action.*

Consider any out-of-equilibrium action and equilibrium response of the buyer. As the selfish seller never suffers from inequity, he ultimately receives a weakly higher maximum deviation utility than the inequity averse seller. Since both types get the same equilibrium utility it is thus impossible that the inequity averse seller gets strictly more whereas the selfish seller gets strictly less than his equilibrium utility. Thus, the buyer may always assign positive probability to the seller being selfish whenever observing an out-of-equilibrium action. He may therefore also reasonably believe the seller to be selfish with certainty. This belief is especially important as it allows the buyer to play his most unfavorable equilibrium response. This out-of-equilibrium belief will thus be used frequently as to afflict 'maximum punishment' for deviations. Lemma 1.4 further implies

Lemma 1.6 *In any pure strategy perfect Bayesian equilibrium with reasonable beliefs the good is always traded in Period 2.*

Proof: Consider a pure strategy perfect Bayesian equilibrium with reasonable beliefs in which the buyer does not always accept after observing the inequity averse seller's equilibrium action (i^*, s_1^*). Given pure strategies the good is never traded in Period 2. By Lemma 1.1 the buyer never sets $s_2^* > \alpha/(1+2\alpha)$ and the inequity averse seller's equilibrium utility U_a^* is given by $u_0(i^*)$. By Lemma 1.2 the selfish seller can essentially get $u_s^m(i^*) > u_0(i^*)$ by choosing investment i^* and a proposal s_1 marginally below $1-\delta$. His equilibrium utility must be weakly higher than what he can get by deviating. Therefore, $U_s^* \geq u_s^m(i^*)$ which implies $U_s^* > U_a^*$. This contradicts Lemma 1.4.

However, if in all pure strategy perfect Bayesian equilibria with reasonable beliefs the buyer always accepts after observing the inequity averse seller's equilibrium action (i^*, s_1^*), the good must always be traded in Period 2 after observing the selfish seller' equilibrium action (i^{**}, s_1^{**}). There are two cases.

First, consider a pooling equilibrium in which the inequity averse and the selfish seller choose the same equilibrium action, $(i^*, s_1^*) = (i^{**}, s_1^{**})$. As the buyer cannot differentiate between the seller's types he must always accept. Second, consider a separating equilibrium in which the buyer does not always accept after observing the selfish seller's equilibrium action (i^{**}, s_1^{**}). Given pure strategies the buyer never accepts in Period 2. Since action (i^{**}, s_1^{**}) is only chosen by the selfish seller, the buyer learns the seller's type and his belief is given by $\mu(i^{**}, s_1^{**}) = 0$. He then proposes $s_2^* = 0$ by Lemma 1.2 and the selfish seller's equilibrium utility U_s^* is given by $u_0(i^{**})$. As shown for Lemma 1.2 the selfish seller then profits from deviating to $(i^{**}, 1-\delta-\epsilon)$ and the above cannot form a perfect Bayesian Nash equilibrium. *Q.E.D.*

Thus, inequity aversion and incomplete information does not cause an inefficient delay of trade when considering pure strategy equilibria. Intuitively this has the following reason. If the buyer always rejects after observing the inequity averse seller's equilibrium investment and proposal, the inequity averse seller's equilibrium utility is equal to his outside option by Lemma 1.1. As given the inequity averse seller's investment the selfish seller can already get a higher minimum deviation utility characterized in Lemma 1.2, he must receive an equilibrium utility higher than the inequity averse seller. This contradicts Lemma 1.4.

Consequently, the good must always be traded in Period 2 in any pooling equilibrium where the selfish seller and the inequity averse seller choose the same equilibrium action. Finally, consider separating equilibria. In equilibrium the buyer learns the selfish seller's type after observing his equilibrium action. If he rejects, he claims the entire trade surplus in Period 3 and the selfish seller receives only his outside option. Yet, by Lemma 1.2 he profits from choosing the same equilibrium investment but proposing marginally less than $1-\delta$ in Period 2. Thus, the good must always be traded in Period 2, which together with Lemma 1.4 has the following implication.

Lemma 1.7 *In any pure strategy perfect Bayesian equilibrium with reasonable beliefs the inequity averse seller receives at least half the trade surplus.*

Proof: By Lemma 1.6 the buyer always accepts after observing the inequity averse seller's equilibrium action (i^*, s_1^*). Suppose $s_1^* < 1/2$. The inequity averse seller's equilibrium utility is then given by $U_a^* = u_a(i^*, s_1^*) - i^*$. But $s_1^* < 1/2$ implies $u_a(i^*, s_1^*) - i^* < u_s(i^*, s_1^*) - i^* \leq U_s^*$ so that $U_s^* > U_a^*$. This contradicts $U_a^* = U_s^*$. *Q.E.D.*

If the inequity averse seller receives less than half the trade surplus, he suffers from inequity in equilibrium. Since the selfish seller never suffers from inequity he could mimic the inequity averse seller and get a strictly larger utility. He thus receives an equilibrium utility strictly larger than the inequity averse seller, which contradicts Lemma 1.4.

1.3.3 Pooling and the Role of Incomplete Information

Building upon the previous sections the following proposition characterizes the set of investments supportable as equilibrium choice in a pooling equilibrium with reasonable beliefs.

Proposition 1.2 *If and only if $i^* \in \mathcal{I}^P$, there exists a pooling equilibrium with reasonable beliefs in which both types of seller invest i^* and the good is always traded in Period 2. The set $\mathcal{I}^P$ is defined as follows.*

1. *If $\pi < \tilde{\mu}$ and $1 - \delta(1-\pi) < 1/2$, then $\mathcal{I}^P$ is empty and there exists no pooling equilibrium.*

2. *If $\pi < \tilde{\mu}$ and $1 - \delta(1-\pi) \geq 1/2$, then*
   $\mathcal{I}^P := \big\{\, i : u_s\big(i,\, 1-\delta(1-\pi)\big) - i \geq \max_k\{u_s^m(k)\}\big\}.$

3. *If $\pi \geq \tilde{\mu}$, then*
   $\mathcal{I}^P := \big\{\, i : u_s\big(\, i, 1-\delta(1+\alpha)/(1+2\alpha)\big) - i \geq \max_k\{u_s^m(k)\}\big\}.$

Unless $\pi < \tilde{\mu}$ and $1 - \delta(1-\pi) < 1/2$, the set $\mathcal{I}^P$ is non-empty, compact, and convex with strictly positive lower boundary. If $i_e \in \mathcal{I}^P$, there exists an efficient pooling equilibrium with reasonable beliefs.

Proof: Lemma 1.6 and Lemma 1.7 imply that in any pooling equilibrium with reasonable beliefs the buyer always accepts after observing the equilibrium action (i^*, s_1^*) and the seller receives at least half the trade surplus, $s_1^* \geq 1/2$. As (i^*, s_1^*) is always chosen by both types of seller, the buyer keeps his prior belief $\mu(i^*, s_1^*) = \pi$. According to Lemma 1.1 this belief π imposes an upper bound on s_1^* and thus restricts the seller's maximum equilibrium utility hereby denoted by $U^P(i)$ for given equilibrium investment i. The remains of the proof characterize the set $\mathcal{I}^P$ of investments so that $s_1^* \geq 1/2$ and the maximum equilibrium utility $U^P(i)$ is higher than the maximum deviation payoffs of both types of seller.

The maximum acceptable equilibrium proposal s_1^* the buyer is willing to accept depends on α, on δ, and on $\pi = \mu(i^*, s_1^*)$ as characterized in Lemma 1.1. Note that for $\pi < \tilde{\mu}$ and $1 - \delta(1-\pi) < 1/2$ the maximum acceptable equilibrium proposal s_1^* is smaller than 1/2. In this case there exists no pooling equilibrium with reasonable beliefs conforming with Lemma 1.7. For all the other cases the maximum s_1^* follows directly from Lemma 1.1.

Consider next the maximum deviation utility. According to Lemma 1.5 the buyer may reasonably believe $\mu(i, s_1) = 0$ for all out-of-equilibrium actions $(i, s_1) \neq (i^*, s_1^*)$. For given deviation investment i the deviation utility is the minimum deviation utility as in Lemma 1.2. The selfish seller does not suffer from inequity and thus receives a weakly higher deviation payoff than the inequity averse seller, $u_s^m(i) \geq u_a^m(i)$, and he receives a weakly higher maximum deviation utility than the inequity averse seller, $\max_i\{u_s^m(i)\} \geq \max_i\{u_a^m(i)\}$. An investment i can thus be supported as equilibrium choice in a pooling equilibrium with reasonable beliefs if and only if $U^P(i^*) \geq \max_i\{u_s^m(i)\}$.

The remainder of the proof characterizes the properties of $\mathcal{I}^P$ for the cases other than $\pi < \tilde{\mu}$ and $1 - \delta(1-\pi) < 1/2$. According to Lemma 1.1 the seller then receives an equilibrium share s_1^* strictly larger than $1 - \delta$. As thus $u_s(i, s_1^*) > u_s^m(i)$ for all i the set $\mathcal{I}^P$ must contain $\tilde{i} := \arg\max_i\{u_S^m(i)\}$, and $\mathcal{I}^P$ is non-empty. The maximum equilibrium utility $U^P(i)$ is strictly concave by Assumption 1.1. $\mathcal{I}^P$ is thus an upper contour set of a strictly concave function, and therefore convex. As $\mathcal{I}^P$ is one-dimensional it forms an interval. As $u_s^m(0) = 0$ and $\partial/\partial i\,\{u_s^m(0)\}$ is strictly positive, $\max_i\{u_s^m(i)\} > 0$. As $U^P(0) = 0$, $\mathcal{I}_S^P$ does not contain 0, is bounded below, and its lower boundary is strictly positive. As $U^P(i)$ is strict concave, $U^P(i') < \partial/\partial i\{U^P(i')\}(i' - i)$ for any i and $i' > i$. As $\lim_{i\to\infty} \partial/\partial i\{U^P(i)\} < 0$ there exists a i with $\partial/\partial i\{U^P(i)\} < 0$, and $U_S^P(i')$ eventually falls below $\max_k\{u_s^m(k)\} > 0$ for sufficiently large i' and i. Consequently $\mathcal{I}_S^P$ is bounded above and therefore bounded. Due to the weak inequality in its definition the boundaries of $\mathcal{I}_S^P$ are included. $\mathcal{I}_S^P$ is thus closed, and therefore compact. *Q.E.D.*

Proposition 1.2 and the following argument are illustrated in Figure 1.1. If the seller chooses the pooling investment, the buyer learns nothing about the seller's type. His prior belief then determines the seller's maximum equilibrium proposal just accepted by the buyer according to Lemma 1.1. Trade at this maximum proposal marks down the seller's maximum equilibrium utility. If the seller chooses any other investment, the buyer may reasonable believe the seller to be selfish with certainty by Lemma 1.5. The deviation utility is then characterized by Lemma 1.2.

As the selfish seller never suffers from inequity he receives a deviation utility weakly higher than the selfish seller. Consequently, an investments is supportable as equilibrium choice in a pooling equilibrium with reasonable beliefs if and only if the associated maximum equilibrium utility exceeds the maximum deviation utility of the selfish seller.

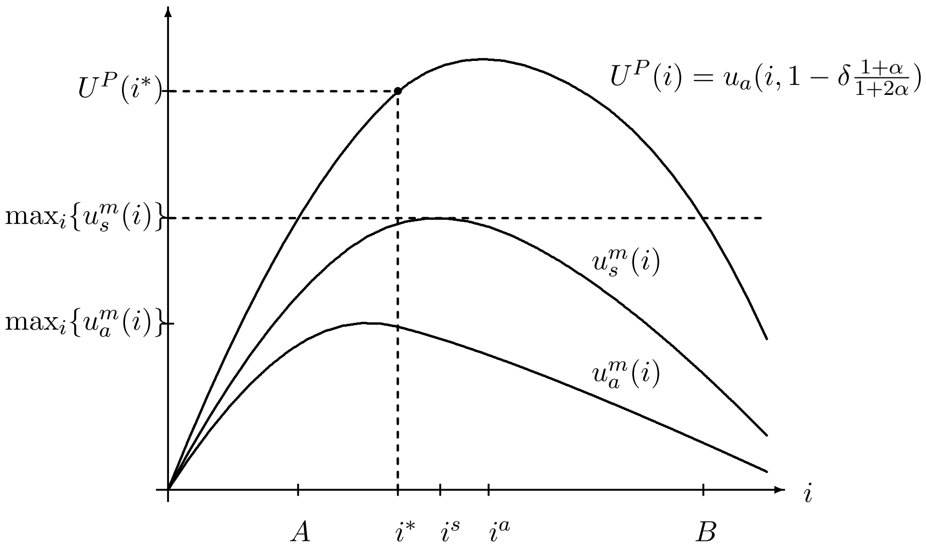

Figure 1.1: Pooling equilibria for $\pi > \tilde{\mu}$ and $1-\delta < 1/2$. Given prior π the maximum accepted equilibrium proposal is $s_1^* = 1-\delta(1+\alpha)/(1+2\alpha)$ giving both types of seller equilibrium utility $U^P(i)$ for equilibrium investment i. The buyer holds out-of-equilibrium belief $\mu(\cdot,\cdot) = 0$ and Lemma 1.2 determines deviation utilities $u_s^m(i)$ and $u_a^m(i)$ with $u_s^m(i) > u_a^m(i)$ as $1-\delta < 1/2$. The set $\mathcal{I}^P$ of equilibrium investments satisfying $U^P(i) \geq \max_k\{u_s^m(k)\}$ is thus given by the interval $[A, B]$. If preferences are observable, the inequity averse seller gets $U^P(i)$ and the selfish seller gets $u_s^m(i)$ for any investment. The unique equilibrium investments are i_a and i_s.

Although the investment choice is continuous, the buyer's equilibrium beliefs create a discontinuity in the seller's objective function crucial for investment incentives. If the seller chooses an investment just marginally lower than the equilibrium investment, the buyer's belief changes. This affects the accepted proposal and thus causes a discrete change in utility. The above mentioned discontinuity is created by the interplay between inequity aversion and incomplete information about preferences. To further highlight this crucial element of the model the following proposition summarizes the effect of inequity aversion on investment incentives if preferences are observable.

Proposition 1.3 *If the seller's preferences are observable, there exists a unique subgame-perfect equilibrium in which*

1. *the good is always traded in Period 2,*
2. *the inequity averse seller proposes* $s_1^* = 1 - \delta(1+\alpha)/(1+2\alpha)$ *and invests* $i^* = \arg\max_i \{u_a(i, s_1^*) - i\}$, *and*
3. *the selfish seller proposes* $s_1^{**} = 1 - \delta$ *and invests* $i^{**} = \arg\max_i\{u_s(i, s_1^{**}) - i\}$.

Moreover, $i_e > i^* > i^{**} > 0$.

Proof: As the seller's type is observable the buyer's belief is independent of the seller's investment and proposal. Thus $\mu(i, s_1) = 1$ for all (i, s_1) if the seller is inequity averse, and $\mu(i, s_1) = 0$ for all (i, s_1) if the seller is selfish. For any investment i both types receive their outside option $u_0(i)$ if they do not trade in Period 2. Instead both types of seller prefer to trade by choosing proposals marginally below the maximum proposals accepted by the buyer according to Lemma 1.1. The good is thus always traded in Period 2 in equilibrium. By a shaving argument the inequity averse seller proposes exactly $s_1^* = 1 - \delta(1+\alpha)/(1+2\alpha)$, the selfish seller proposes exactly $s_2^{**} = 1 - \delta$, and the buyer accepts. Furthermore, the buyer's acceptance decision is independent of the chosen investment. The equilibrium proposals thus imply the equilibrium investments. They are unique and strictly positive by Assumption 1.1. *Q.E.D.*

If preferences are observable, the buyer's belief is independent of the seller's action. Applying Lemma 1.1 with degenerate beliefs the inequity averse seller gets a larger share of the trade surplus than the selfish seller. However, not even the inequity averse seller receives the entire trade surplus. Although the inequity averse seller's investment incentives are improved, both types of seller under-invest. This is illustrated in Figure 1.1.

Incomplete information thus has the following impact on investment incentives. If preferences are observable, the buyer exploits the hold-up situation less - bargains less aggressively - if and only if the seller is inequity averse. This changes if information is incomplete. In a pooling equilibrium the buyer behaves in the same way independent of the seller's type.

Thus, investment incentives for both types of seller's may improve. However, investment incentives are not always ameliorated as they depend on the considered pooling equilibrium. For example, in the equilibrium in Figure 1.1 both types of seller choose the same equilibrium investment i^* smaller than even the selfish seller's investment if preferences are observable.

1.3.4 Separation and the Costs of Signalling

It should be strongly emphasized that the signalling in the present model is no 'cheap talk'. Since the seller's signal is two-dimensional, signalling costs depend on the proposal if the latter is less than one half. If the buyer accepts and the good is traded, the inequity averse seller suffers from inequity. Both investment and proposal then influence this suffering and thus have a type-dependent impact on the seller's utility. Increasing the proposal while keeping the investment fixed increases the utility of both types of seller by increasing his share of the trade surplus. The inequity averse seller's utility rises further as unfavorable inequity is reduced. Moreover, investing more while keeping the proposal fixed increases the trade surplus and thus the seller's allotment of the latter. As the inequity averse seller's suffering increases in the trade surplus, this reduces his utility. Thus, there is scope for separating equilibria in which the selfish seller invests a lot and receives a small share of the large trade surplus, whereas the inequity averse seller invests a little and receives a more equitable share of the small trade surplus. Yet, equilibrium dominance has a strong impact.

Proposition 1.4 *If and only if the inequity averse seller's equilibrium investment i^* is element of $\mathcal{I}^S := \mathcal{A} \setminus \mathcal{B}$ with*

$$\mathcal{A} := \left\{ i : u_a\big(i, 1-\delta(1+\alpha)/(1+2\alpha)\big) - i \geq \max_i\{u_s(i,1-\delta)-i\} \right\}, \text{ and}$$

$$\mathcal{B} := \left\{ i : u_a(i,1/2) - i > \max_i\{u_s(i,1-\delta)-i\} \right\},$$

there exists a separating equilibrium with reasonable beliefs and the following characteristics. The good is always traded in Period 2. The selfish seller invests $i^{**} = \arg\max_k\{u_s^m(k)\}$, proposes $s_1^{**} = 1-\delta$, and receive equilibrium utility $U_s^* = u_s(i^{**}, s_1^{**})$. The inequity averse seller invests i^* and proposes s_1^* so that $U_a^* = u_s(i^*, s_1^*) = U_s^*$.
Moreover, $\mathcal{A}$ is non-empty, compact, and convex with strictly positive lower boundary. $\mathcal{B}$ is either empty or a real subset of $\mathcal{A}$, open, and convex. $\mathcal{A}\backslash\mathcal{B}$ is non-empty and compact, it is convex if $\mathcal{B}$ is empty.

Proof: The set of separating equilibria with reasonable beliefs is almost entirely determined by Lemma 1.4 to 1.7: the good is always traded in Period 2, the inequity averse seller receives at least half the trade surplus, $s_1^* \geq 1/2$, and both types of seller must receive the same equilibrium utility so that $U_s^* = U_a^*$. This has the following implications.

The proof first shows that if there exists no profitable deviation for the selfish seller, the same holds true for the inequity averse seller. Consider first deviations to any out-of-equilibrium action (i, s_1). By Lemma 1.5 the buyer may hold the reasonable belief $\mu(i, s_1) = 0$. A seller's deviation utility is then given by $u_a^m(i)$ and $u_s^m(i)$ as characterized in Lemma 1.2. Since the selfish seller never suffers from inequity, $u_s^m(i) \geq u_a^m(i)$ holds for all i. In equilibrium the selfish seller may not profit from choosing an out-of-equilibrium action, $U_s^* \geq \max_i\{u_s^m(i)\}$. This implies $U_a^* \geq \max_i\{u_a^m(i)\}$: if the selfish seller cannot profit from choosing an out-of-equilibrium action, the same holds true for the inequity averse seller.

Moreover, no type of seller profits from exactly mimicking the other type. According to Lemma 1.6 the good is always traded in Period 2. Therefore, $U_s^* = u_s(i^{**}, s_1^{**}) - i^{**}$ and $U_a^* = u_a(i^*, s_1^*) - i^*$. First, consider the selfish seller mimicking the inequity averse seller. Since By Lemma 1.7 the inequity averse seller receives at least half the trade surplus, $s_1^* \geq 1/2$, the selfish seller then gets exactly the inequity averse seller's equilibrium utility, $u_s(i^*, s_1^*) - i^* = U_a^*$. As $U_a^* = U_s^*$ by Lemma 1.4, the selfish seller does not profit from mimicking the inequity averse seller. Second, consider the inequity averse seller mimicking the selfish seller. As the selfish seller's equilibrium proposal can be smaller than 1/2 the inequity averse seller receives weakly less than the selfish seller's equilibrium utility when mimicking the selfish seller, $u_a(i^{**}, s_1^{**}) - i^{**} \leq u_s(i^{**}, s_1^{**}) - i^{**} = U_s^*$. As both types receive the same equilibrium utility, the inequity averse seller thus receives weakly less than his own equilibrium utility and does not profit from mimicking the selfish seller.

So far results can be summarized as follows. First, no type profits from mimicking the other type. Second, if the selfish seller weakly prefers to choose his own equilibrium action (i^{**}, s_1^{**}) to any out-of-equilibrium action, the inequity averse seller at least weakly prefers to choose his own equilibrium action (i^*, s_1^*).

As the buyer learns the seller's type in equilibrium, $\mu(i^*, s_1^*) = 1$ and $\mu(i^{**}, s_1^{**}) = 0$, the selfish seller's equilibrium utility U_s^* is essentially predetermined. By Lemma 1.6 the good must always be traded in Period 2. As $\mu(i^{**}, s_1^{**}) = 0$ the buyer only accepts if $s_1^{**} \leq 1 - \delta$ by Lemma 1.1. By a shaving argument the selfish seller's equilibrium proposal is thus given by $s_1^{**} = 1 - \delta$. Mimicking the inequity averse seller can never be profitable. However, the selfish seller can get the same share $s_1 = 1 - \delta = s_1^{**}$ when choosing any out-of-equilibrium investment. His equilibrium investment must thus be characterized by $i^{**} = \arg\max_i \{u_s^m(i)\}$ yielding equilibrium utility $U_s^* = u_s(i^{**}, 1 - \delta) - i^{**}$.

Given the fixed equilibrium utility U_s^* of the selfish seller all those investments can be supported as the inequity averse seller's equilibrium choice i^* for which there exists a proposal s_1^* equating the equilibrium utilities of both types. Yet there are restrictions on s_1^*. The remainder of the proof first describes the restrictions on s_1^* and then characterizes the set $\mathcal{I}^S$ of investments supportable as equilibrium choices of the inequity averse seller.

There are the following restrictions on the inequity averse seller's equilibrium proposal s_1^*. First, note that the buyer learns the inequity averse seller's type after observing his equilibrium action, $\mu(i^*, s_1^*) = 1$. According to Lemma 1.6 the good must always be traded in Period 2. Thus, s_1^* may not exceed $1 - \delta(1 + \alpha)/(1 + 2\alpha)$ by Lemma 1.1. Second, Lemma 1.7 requires $s_1^* \geq 1/2$. Thirdly, s_1^* must exceed $1 - \delta = s_1^{**}$. If $s_1^* < 1 - \delta$, then $u_s(i^*, s_1^*) - i^* < u_s^m(i^*) < U_s^*$ for all i^*. This contradicts $U_a^* = U_s^*$ and thus $s_1^* \geq 1 - \delta$. Note that if $s_1^* = 1 - \delta = s_1^{**}$, equality of utilities implies $i^* = i^{**}$ and the separating equilibrium is in fact a pooling equilibrium. Summarizing the inequity averse seller's equilibrium proposal s_1^* must lie in the interval $[\underline{s}, \overline{s}]$ where $\underline{s} := \max[1/2,\, 1 - \delta]$ and $\overline{s} := 1 - \delta(1 + \alpha)/(1 + 2\alpha)$. This interval is non-empty.

The remainder of the proof characterizes $\mathcal{I}^S$, the set of investments supportable as equilibrium choice of the inequity averse seller. Note that $U_a^* = u_a(i^*, s_1^*) - i^*$. The set $\mathcal{I}^S$ is defined so that for all $i \in \mathcal{I}^S$ the following holds: $u_a(i, \underline{s}) - i \leq U_s^*$ and $u_a(i, \overline{s}) - i \geq U_s^*$ with at least one inequality strict. The continuity of $u_a(i, s_1)$ in s_1 then implies that for all $i \in \mathcal{I}^S$ there exists an admissible s_1 equating the equilibrium utilities of both types.

Define $\mathcal{I}^S := \mathcal{A} \setminus \mathcal{B}$ where

$$\mathcal{A} := \left\{ i : u_a(i, \bar{s}) - i \geq U_s^* \right\}, \quad \text{and} \quad \mathcal{B} := \left\{ i : u_a(i, 1/2) - i > U_s^* \} \right\}.$$

With the corresponding definitions of $\underline{s}$ and $\bar{s}$ these sets are identical to the sets defined in the proposition. All $i \in \mathcal{I}^S$ are element of $\mathcal{A}$ as $u_a(i, \bar{s}) - i \geq U_s^*$ by definition. Otherwise there are two cases. First, suppose $1 - \delta \geq 1/2$. In this case $\mathcal{B}$ is empty and $\underline{s} = 1 - \delta$. Since $u_a(i, \underline{s}) = u_s^m(i)$ and $u_s^m(i) \leq \max_k\{u_s^m(k)\} = U_s^*$, this implies $u_a(i, \underline{s}) \leq U_s^*$ for all $i \in \mathcal{A} = \mathcal{I}^S$. Second, suppose $1 - \delta < 1/2$. In this case $\mathcal{B}$ is non-empty and $\underline{s} = 1/2$. However, all $i \in \mathcal{I}^S$ are not element of $\mathcal{B}$ and thus $u_a(i, \bar{s}) \geq U_s^*$. Taken together the set $\mathcal{I}^S = \mathcal{A} \setminus \mathcal{B}$ satisfies the above mentioned properties.

This last part of the proof characterizes the properties of $\mathcal{A}$, $\mathcal{B}$, and thus $\mathcal{I}^S$. By definition of $\bar{s}$ the set $\mathcal{A}$ is identical to the set $\mathcal{I}^P$ for $\pi \geq \tilde{\mu}$. It thus forms a non-empty, compact interval with strictly positive lower boundary. Due to the same logic $\mathcal{B}$ is either empty or a bounded, open interval with strictly positive upper lower bound. Even if $\mathcal{B}$ is non-empty, the boundaries of $\mathcal{A}$ are not included in $\mathcal{B}$. Consider a boundary i' of $\mathcal{A}$. By definition $u_a(i', \bar{s}) - i' = U_s^*$. Since $\bar{s} > 1/2 = \underline{s}$ this implies $U_s^* = u_a(i', \bar{s}) - i' > u_s(i', \underline{s}) - i'$. Thus, $i' \notin \mathcal{B}$. If $\mathcal{B}$ is non-empty, it is thus a real subset of $\mathcal{A}$. This further implies that $\mathcal{A} \setminus \mathcal{B}$ is non-empty. If $\mathcal{B}$ is empty, $\mathcal{A} \setminus \mathcal{B}$ forms an interval and is therefore convex. Otherwise $\mathcal{A} \setminus \mathcal{B}$ is an interval with a 'open hole' around the selfish seller's equilibrium investment i^{**}. *Q.E.D.*

Proposition 1.4 and the following argument are illustrated in Figure 1.2. In a separating equilibrium the buyer holds the following beliefs. First, he learns the seller's type after observing the equilibrium actions. Second, he believes the seller to be selfish after observing any out-of-equilibrium action. Unless precisely mimicking the inequity averse seller, the selfish seller can thus always get share $1 - \delta$ of the trade surplus. As shown below mimicking the inequity averse seller is not profitable for the selfish seller. Thus, the selfish seller's unique equilibrium investment maximizes his utility given he receives share $1 - \delta$ of the trade surplus. This determines his equilibrium utility.

According to Lemma 1.4 an investments and proposal is supportable as the inequity averse seller's equilibrium action if it equates both type of seller's equilibrium utility. However, the inequity averse seller's proposal must satisfy the following conditions. First, the good must always be traded according to Lemma 1.6, which imposes an upper bound by Lemma 1.1. Second, the proposal must exceed one half by Lemma 1.7. These restrictions on the inequity averse seller's equilibrium proposal define the set of possible equilibrium investments.

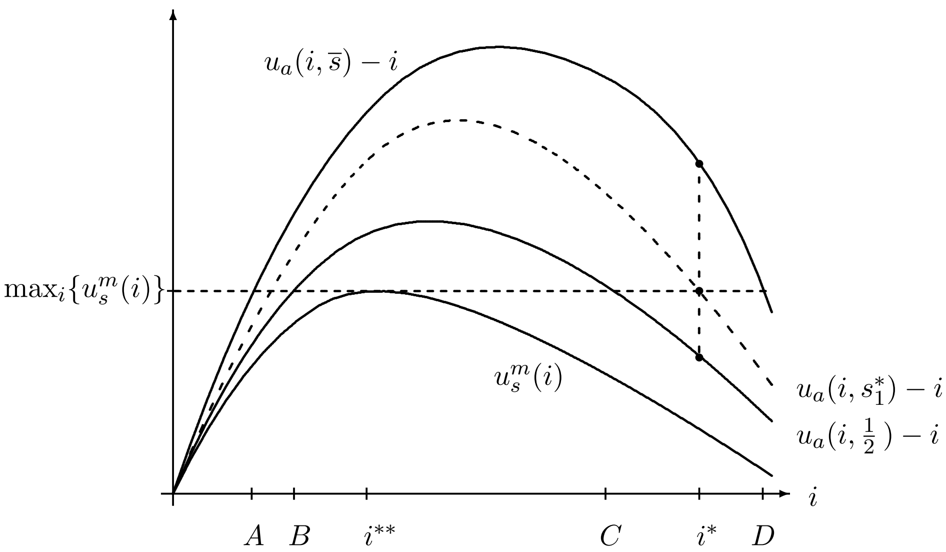

Figure 1.2: Separating equilibria for $1-\delta < 1/2$. The buyer learns the seller's type in equilibrium. He otherwise believes the seller to be selfish. The selfish seller gets share $1-\delta$ and utility $u_s^m(i)$ for all investments i, chooses equilibrium investment i^{**}, and gets equilibrium utility $\max_i\{u_s^m(i)\}$. By Lemma 1.6 the good is always traded, by Lemma 1.7 the inequity averse seller gets at least half the trade surplus. Thus, his equilibrium proposal s_1^* must be weakly smaller than $\overline{s} = 1-\delta(1+\alpha)/(1+2\alpha)$ and weakly larger than $1/2$. In both extreme cases he gets equilibrium utility $u_a(i,\overline{s}) - i$ or $u_a(i, 1/2) - i$. Lemma 1.4 requires equilibrium utilities to be equal. The set $\mathcal{I}^S$ of the inequity averse seller's equilibrium investments contains all investments i for which there exists $s_1^* \in [1/2\,, \overline{s}\,]$ so that $u_a(i, s_1^*) - i = \max_i\{u_s^m(i)\}$. Thus, $\mathcal{I}^S = [A,B] \cup [B,C]$. It does not contain $]B,C[$ as given the minimum admissible proposal $1/2$ the inequity averse seller gets more than the selfish seller.

Finally, no type of seller has incentives to deviate. By construction of the equilibrium the selfish seller is exactly indifferent between choosing his equilibrium action and mimicking the inequity averse seller. By choice of his equilibrium investment he strictly prefers his equilibrium action to all out-of-equilibrium actions. Furthermore, the inequity averse receives a utility weakly smaller than the selfish seller if he deviates or mimics the selfish seller, thus has a weakly smaller deviation utility, and both types of seller at least weakly prefer their own equilibrium action to any possible deviations.

1.4 Conclusion

Although inequity aversion can improve investment incentives in the hold-up problem, the interaction of inequity aversion and incomplete information about an individual's preferences is the key element of the present model. If preferences are observable and the seller always gets the same share of the trade surplus, changing the investment has a continuous impact on the trade surplus and the seller's utility. If preferences are unobservable, changing the investment might change the buyer's belief about the seller's type and thus the buyer's bargaining behavior. Depending on these changes, the signalling character of the seller's investment creates a discontinuity generating very strong investment incentives.

As with most bargaining games under incomplete information, the game form strongly influences equilibrium behavior. The quality of the results should be unaffected as long as both types of seller receive some rent in equilibrium, a rent they can loose by appearing to be of a certain type.[8] In the present model discounting grants both types a first mover advantage. However, if at least one type of seller receives no rent, the model breaks down.

Suppose, for example, buyer and seller play an 'ultimatum game' after the seller has sunk his investment. In this game the buyer proposes how to split the trade surplus, and the seller may accept or reject. If the seller accepts the good is traded according to the proposed term of trade. If the seller rejects, bargaining breaks down and the game ends. The maximum equilibrium proposal in any perfect Bayesian equilibrium never exceeds the proposal rendering the inequity averse seller just indifferent between accepting and rejecting. The buyer thus extracts all rents from the inequity averse seller independently of the buyer's belief after observing the seller's investment. Given that the inequity averse seller always gets his outside option, convincing the buyer that he is inequity averse has no value to him. He therefore chooses the investment maximizing his outside option. The same argument holds true if the bargaining game has any number of periods but the buyer makes all proposals.

[8] Incorporating two-sided investments by buyer and seller should thus cause no problems as long as a suitable bargaining game is considered in which both traders get a rent by appearing to be of a certain type.

Despite the use of the equilibrium refinement 'equilibrium dominance', the multiplicity of perfect Bayesian equilibria does not admit unique empirical predictions. There is little hope that other refinements can improve the situation. However, this apparent weakness makes the model consistent with the empirical evidence that efficiency problems caused by incomplete contracts are not always equally severe. Furthermore, the model does not argue that incomplete contracts are efficient, nor does it recommend that contracts should be voluntarily left incomplete. Yet it explains why the consequences of contractual incompleteness are not necessarily catastrophic if buyer and seller might be inequity averse and - equally important - preferences are private information.

Chapter 2

Adverse Selection and Employment Contracts

2.1 Introduction

Numerous surveys, empirical studies and anecdotes suggest that fairness and social comparisons are of great concern for employees. Moreover, investigations by Blinder and Choi (1990), Bewley (1995), and Campbell and Kamlani (1997) find convincing evidence that employers account for these concerns. Fairness considerations have thus become an increasingly accepted explanation for wage stickiness, wage compression, and unemployment of low-skilled workers. These intuitive implications of fairness and social comparisons are supported and further deepened by a relatively small but influential collection of theoretical articles. The present paper adds to this theoretical discussion by investigating a firm's optimal employment decision and employment contracts if workers have private information about their productivity and are inequity averse in the spirit of Fehr and Schmidt (1999). It finds that the impact of fairness and inequity aversion depends on what the workers compare. Particulary, if production costs are incorporated in the workers' comparisons, fairness causes an increasing income difference and can never account for an unwillingness of firms to employ low-skilled workers. This contradicts the existing literature incorporating fairness into the labor market.

Consider a single firm intending to employ workers with private information about their productivity. There are two types of workers, workers with low production costs and workers with high production costs. For simplicity these workers are called productive and unproductive workers throughout the introduction. All workers are taken to be inequity averse, they compare themselves with their colleagues. The firm exclusively cares for its profit but knows that workers are inequity averse and takes this into account. The present paper investigates the impact of inequity aversion on the firm's employment decision and employment contracts. It turns out that the results delicately depend on the workers' point of reference. Two extreme cases are considered: workers comparing only income and workers comparing income minus production costs.

There are the following results. As the two types of workers differ in their productivity they usually receive different income levels. If workers compare income, those workers getting the lower income suffer. In order to employ these workers they must receive an income premium as compensation for this suffering. The firm optimally responds to these additional costs by reducing the income difference. Inequity aversion thus causes an income compression. Moreover, adjusting employment contracts to the diminished income difference distorts production quantities of both types of workers. In order to focus on the employment effect of inequity aversion it is assumed that marginal costs of production converge to zero as production goes to zero. Thus, asymmetric information never causes an exclusion of unproductive workers. Inequity aversion, however, introduces an additional distortion on the productive workers' employment contract. If there are only few unproductive workers, it might then be optimal for the firm to exclude these workers in order to avoid these distortions. Even though the labor market is not explicitly modeled, inequity aversion might thus hint at an explanation for unemployment of unproductive workers. Note that both effects only depend on the initial income dispersion created by differences in productivity. They are therefore independent of whether the firm can observe the workers' productivity or not.

The present paper then shows that the impact of inequity aversion reverses direction if workers compare rents defined as income minus production costs. If productivity is observable, the firm can exactly compensate each worker for his production costs. Thus, both types of worker receive zero rent, there is no rent inequality, and inequity aversion is irrelevant. Contrary to the case of workers comparing income, introducing asymmetric information changes the implications of inequity aversion if workers compare rents. If productivity is unobservable, productive workers must get an informational rent to induce truthful revelation of their type. This informational rent is exactly what a productive worker can gain by pretending to be unproductive. The informational rent thus depends on the production quantity specified for unproductive workers. As unproductive workers are not given an informational rent, asymmetric information causes a rent inequality for which unproductive workers must be compensated. The associated costs can only be diminished by reducing the informational rent, and thus by reducing the unproductive workers' production quantity. However, this allows the firm to reduce the unproductive workers' compensation for both rent inequality and production costs so that their income decreases strongly. The productive workers' income is only reduced by the informational rent, and inequity aversion increases the income difference.

Although inequity aversion causes additional distortions these depend only on the rent inequality and thus on the informational rent. The compensation for rent inequality can be made arbitrarily small by reducing the unproductive workers' production quantity. At the same time this diminishes the distortions imposed on the productive workers' employment contract. As marginal costs of production converge to zero, the first units produced by the unproductive workers are essentially costless regarding production costs, informational rent, and compensation for rent inequality. Thus, if the workers are inequity averse and compare rents, it is always optimal to employ both types of workers, and both types of workers produce positive production quantities.

Summarizing, the impact of inequity aversion depends on what workers compare. Even if workers only compare income, the overall effect might be small for reasonable levels of inequity aversion. Moreover, if workers account for the other workers' production costs, the effect of inequity aversion either reverses direction or is at least diminished. This holds true for large levels of inequity aversion. Therefore, inequity aversion as such does not seem to be a strong explanation for the empirically observed phenomena of income compression and involuntary unemployment of workers with low productivity.

The most prominent theoretical work introducing fairness into the labor market was conducted by Akerlof (1982), Akerlof and Yellen (1988), and Akerlof and Yellen (1990). According to their 'fair wage effort hypothesis' workers reduce their work effort if they receive less than a ‘fair wage’ depending on other workers’ wages. The fair wage effort hypothesis is supported by a number of experiments, for example Fehr, Kirchsteiger, and Riedl (1993). Workers’ reaction to ‘unfair’ wages reduces their marginal productivity of labor, which in turn influences firms’ demand for labor. In a general equilibrium model the authors show that there exist equilibria in which wages are compressed relative to differences in productivity, and there is unemployment of low-skilled workers. Yet, firms are confined to flat wage contracts although the story is based on a moral hazard argument and firms correctly anticipate the workers’ reaction to their wage payments. The present paper explicitly investigates optimal incentive contracts. Focussing on the heterogeneity of workers, the source of incentive problems is not hidden action but hidden characteristics. Adding moral hazard might amplify the impact of inequity aversion but also complicates the analysis. However, Bartling and von Siemens (2004b), Bartling and von Siemens (2004a) and the authors cited therein look at the impact of inequity aversion in a setting of moral hazard with several, homogeneous agents. In another seminal paper, Frank (1984b) proposes status considerations as an explanation for wage compression. In his model workers enjoy additional utility if they receive a higher wage than co-workers. A productive worker might thus reject a higher wage - corresponding to his high marginal productivity - offered by another firm if by accepting he no longer enjoys the status of a high earner. Yet there is no asymmetric information in his model, and status considerations are different from inequity aversion.

The paper is organized as follows. Section 2.2 presents the model. It specifies workers’ productivity types, the informational structure, and introduces inequity aversion. Section 2.3 contains the main results of the paper. It first analyzes the informational constraints restricting the firm’s employment contracts. It then shows that inequity aversion causes an income compression and possibly an exclusion of unproductive workers if workers compare income. Next, it demonstrates that if workers compare rents, inequity aversion increases the income difference between productive and unproductive workers. In addition both types of workers are then always employed. Section 2.3 analyzes the impact of asymmetric information by solving the firm’s maximization problem if the workers’ types are observable. Section 2.4 discusses some extensions, and Section 2.5 concludes.

2.2 The Model

2.2.1 Production, Information and Sequence of Actions

Consider a single firm facing a continuum of workers with measure one. Once employed each worker can produce some output q with type dependent production costs $\theta\, c(q)$. There are two types of workers, θ_g and θ_b with $\theta_b > \theta_g > 0$. Since they can produce the same output at lower costs, workers of type θ_g are called 'good' whereas workers of type θ_b are called 'bad' workers. Moreover,

Assumption 2.1

1. *c is strictly increasing and strictly convex for all $q > 0$.*
2. *c satisfies the Inada conditions,* $\lim_{q\to 0} c'(q) = 0$ *and* $\lim_{q\to\infty} c'(q) = +\infty$.

An employed worker's relationship with the firm is governed by an employment contract. An employment contract consists of a production quantity q and an income t. Given employment contract (q, t) a worker has to produce q and receives income t. This generates profit $q - t$ for the firm.

Asymmetric information about workers' productivity lies at the heart of the ensuing analysis. The distribution and eventual revelation of this information is best explained by looking at the sequence of actions. The model can be roughly divided into three periods.

Period 1: At the beginning of Period 1 each worker's type is assumed to be private information and thus only known to him. However, it is common knowledge that types are independent and each worker is good with probability $\pi \in]0, 1[$ and bad otherwise. As there is an infinite number of workers the law of large numbers implies that the fraction of good workers is always π. Given its limited information about workers' productivity the firm designs a yet unspecified mechanism to be used in Period 1. This mechanism determines whether a worker gets employed or not. If he is employed, the mechanism assigns him an employment contract, otherwise the worker remains unemployed.

The present paper restricts attention to deterministic, direct revelation mechanisms.[1] Thus, the mechanism asks each worker for his productivity and uses only the obtained answers to determine the job allocation. Since each worker's type is private information and types are drawn independently, the firm cannot learn anything about one worker from the announcements of the others. Thus, it is no restriction to look only at mechanisms in which each worker's contract conditions exclusively on his own announcement, and the firm offers a maximum of two employment contracts (q_b, t_b) and (q_g, t_g), one for each type. A worker chooses contract (q_k, t_k) by announcing to be of type θ_k. Given the mechanism and the expected announcements of the other workers, each worker can decide not to participate. He then stays unemployed.

Period 2: In Period 2 workers work and the firm receives profit $\eta_k(q_k - t_k)$ if fraction η_k of the workers fulfills employment contract (q_k, t_k). Furthermore, in the course of Period 2 both the fractions of employed workers satisfying a certain employment contract and each worker's type become common knowledge.

Period 3: Thus, there is no asymmetric information at the beginning of Period 3. However, it is assumed that the information revealed 'on the job' may not be used in any way. Thus, employment contracts may not initially condition on the information revealed in Period 3, or be renegotiated. This assumption seems justifiable as, for example, workers might already have a legal claim on their income by having produced the corresponding production quantity in Period 2. The use of this seemingly unnecessary complication will become clear when defining the workers' utility function. Finally, employed workers receive their income at the end of Period 3.

The firm would like to exploit its monopolistic position by giving each worker an income just sufficient to make him accept work. Yet, each worker's acceptance decision depends on his production costs. As costs are type-dependant and types are private information, the firm faces a standard contract theoretical problem of rent extraction. The present paper

[1] Looking at the proof of the revelation principle in Fudenberg and Tirole (1991), there seems to be no obvious reason why the revelation principle should not hold if agents are inequity averse. However, this is not explicitly proven in the present paper.

differs from the usual literature on mechanism design as it investigates the firm's optimal employment decision and employment contracts assuming that workers are inequity averse.

2.2.2 Incorporating Inequity Aversion

Inequity aversion is modeled in the spirit of Fehr and Schmidt (1999). In order to apply their theory of fairness, 'who compares what with whom in which way' must be specified. There are the following general assumptions.

Assumption 2.2

1. *The firm maximizes profits.*
2. *All workers have identical preferences and are inequity averse.*
3. *Employed workers compare themselves exclusively with other employed workers.*
4. *Unemployed workers compare themselves exclusively with other unemployed workers.*
5. *Inequity averse workers do not suffer if they feel 'better off'.*

Assumption 2.2 needs some comments. Surely, some firms do care for their workers, and especially founding entrepreneurs occasionally exhibit a paternalistic trait. For example, in the 19th century the then privately owned German steel manufacturer Krupp offered his workers housing and free social services like schooling or a theater. However, this attitude seems to be vanishing as companies become public. Owners then often consider their share in the corporation as nothing but a small position in their investment portfolio. Moreover, the present paper concentrates on inequity aversion of workers and ignores other but selfish motivations of the firm.

As the main focus is on screening workers with respect to their productivity all workers are taken to have identical concerns for inequity. Section 2.5 shortly discusses the consequences if individuals differ in their degree of inequity aversion and this is private information.

Part 2 and 3 of Assumption 2.2 consider workers' reference group. Psychologists and behavioral economists seem to believe that individuals mostly compare themselves with individuals perceived as 'equal' and working or living in close proximity.[2] The present paper captures this in the following way. First, the reference group of employed workers consists only of all the other workers employed within the same firm. Second, the reference group of unemployed workers consists only of all the other unemployed workers. Thus, the utility of a worker rejecting all employment contracts offered by the firm is independent of what the employed workers get, and the outside option can be normalized to zero. Section 2.5 discusses this assumption. Thirdly, no worker compares himself with the firm in any way as the present paper concentrates on inequity aversion amongst workers.

Finally, workers only dislike being worse off but not being better off than their colleagues. Fehr and Schmidt (1999) suggest that altruistic motivations are dominated by envy. The present paper focuses on the latter aspect of inequity aversion, which seems plausible in the context of unobservable characteristics. Not many people complain that they earn more than their colleagues.

Inequity aversion is formalized as follows. In all definitions below consider a worker of type θ_i with $\hat{\theta}_k$ indicating that he has announced to be of type θ_k. Suppose the worker is then employed and has to satisfy employment contract (q_k, t_k). Let

$$u(\theta_i, \hat{\theta}_k) := t_k - \theta_i \, c(q_k) \tag{2.1}$$

denote this worker's *rent* defined as his income minus production costs. Define his overall utility as his rent minus his *suffering* $S(\theta_i, \hat{\theta}_k)$ so that

$$U(\theta_i, \hat{\theta}_k) := u(\theta_i, \hat{\theta}_k) - S(\theta_i, \hat{\theta}_k). \tag{2.2}$$

The term $S(\theta_i, \hat{\theta}_k)$ specified below captures the worker's concerns for inequity. Note that Assumption 2.2 does not pin down what workers compare. There seems to be no common agreement on this important issue amongst behavioral economists. The present paper thus explores the following two alternatives, the first of which is

[2]See, for example, Festinger (1954) and Williams (1975).

Definition 2.1 (Comparing Income) *Given a direct revelation mechanism, the suffering of a worker of type θ_i announcing to be of type θ_k is given by*

$$S(\theta_i, \hat{\theta}_k) = \alpha \sum_{j=g,b} \eta_j \max[\, t_j - t_k, 0],$$

where $\alpha \geq 0$ is the measure of inequity aversion and η_j is the fraction of workers announcing to be of type θ_j.

According to Definition 2.1 inequity averse workers exclusively compare their own income with the income of the other workers. Production quantities specified in other contracts and the associated production costs are completely ignored. Many people seem to talk a lot about their absolute income per year, and Blinder and Choi (1990) suggest that individuals tend to overemphasize nominal wages. However, ignoring the impact of production costs seems absurd when taken to extremes. As an illustration consider a worker who receives 2000 Euro per month and in turn has to spend 1 hour a day at work. If production costs were not incorporated, this worker would envy another worker who receives 2001 Euro per month but has to spend 10 hours a day at work. The second definition of inequity aversion perfectly incorporates production costs.

Definition 2.2 (Comparing Rents) *Given a direct revelation mechanism, the suffering of a worker of type θ_i announcing to be of type θ_k is given by*

$$S(\theta_i, \hat{\theta}_k) = \alpha \sum_{j=g,b} \sum_{m=g,b} \eta_{jm} \max[u(\theta_m, \hat{\theta}_j) - u(\theta_i, \hat{\theta}_k), 0],$$

where $\alpha \geq 0$ is the measure of inequity aversion and η_{jm} is the fraction of workers of type θ_m announcing to be of type θ_j.

Note that in order to allow a comparison of rents a worker must either know or have a correct expectation of the other workers' types. In the present model workers' learn this information in the course of working in the firm. Changing this assumption can cause the following problems. First, if the workers know the other workers' types from the start, the firm can devise simple mechanisms eliciting all information at no costs. Second, if the workers' types remain private information, a worker's utility depends on his belief about the other workers' types. In this case, psychological game theory must be applied, but doing this is inherently difficult.[3] These complications are avoided by the revelation of each worker's type on the job. Yet the present model degenerates to a static problem of rent extraction since the firm cannot immediately use this information to revise its employment decision by assumption.

Unfortunately, the experimental evidence cannot really help to discriminate between the two proposed definitions of inequity aversion. Quite often players' real effort does not depend on their chosen action but simply on the time spent playing the experiment. Sometimes a 'virtual' effort influences a player's payoff, but then this effort is translated into monetary terms. This might create a strong framing effect falsely suggesting that individuals compare rents. The present paper analyzes the above two extreme cases: workers comparing only income and workers perfectly incorporating production costs. This simplifies computations and disentangles the different impacts of inequity aversion.

However, individuals probably account imperfectly for production costs, either as precise effort functions are not known, or individuals are prone to a self-serving bias like 'Yes, the other worker produced more and worked harder than I did. But I also worked hard, so he should not receive a higher income'. Partially accounting for production costs could be captured by an additional parameter weighing the extent to which inequity averse workers incorporate production costs. Preliminary computations seem to suggest that the solution is continuous in this weight. Working with such a hybrid model should thus generate results lying in between the two extremes cases studied here.

[3]Note that the optimal mechanisms proposed in this paper satisfy Bayesian incentive compatibility: a worker reveals his type truthfully to the firm if he expects that all other workers reveal their type truthfully. Given the associated Bayesian equilibrium, each worker's expectation of the type (and therefore which fraction) of workers choosing a certain contract is correct. One could thus conjecture that the quality of the results is not changed by applying psychological game theory.

2.3 Results

2.3.1 Incentive and Participation Constraints

According to the specification of inequity aversion employed in this model, a worker changes reference group depending on whether he is employed or not. Therefore, not employing a worker at all is not equivalent to employing a worker and having him produce nothing while giving him no income. The firm's maximization problem must thus explicitly analyze the employment decision, and optimal mechanisms are found in the following way. First, the firm's maximum profit if both types of workers are employed are calculated. Second, the firm's maximum profit if only one type of worker is employed are derived. Finally, the firm's employment decision follows from a comparison of the maximum profits accruing in the above two alternatives.[4] The following results simplify the computation of the optimal employment contracts. They hold for both definitions of inequity aversion.

First, suppose the firm employs only one type of worker. Excluding the good workers is not possible if productivity is unobservable. The firm makes a strictly positive profit only if the offered employment contract specifies a strictly positive production quantity. If bad workers get their outside option upon accepting employment, good workers get a strictly positive rent as they then incur strictly lower production costs. A good worker thus also accepts. Moreover, contracts of the form "burn all workers in oil if the fraction of workers accepting exceeds the measure $1 - \pi$ of bad workers" cannot work as each individual worker has measure zero. A good worker can thus join the firm without changing the fraction of workers accepting and thus without activating the punishment.

The relevant case is the firm excluding the bad workers. The firm can then extract all rents from the good workers but looses profit generated by the bad workers. As only one type of workers is employed and all get the same contract, inequity aversion is irrelevant. The firm optimally offer a single employment contract (q, t) with $t = \theta_g \, c(q)$. This employment contract is just accepted by the good workers. However, if a bad worker accepts, the income does not cover his higher production costs and in addition he suffers from inequity if workers

[4]The model does not consider mechanisms employing only a fraction of the good or bad workers.

compare rents. Bad workers thus reject, and the firm's profit is given by

$$R = \pi\,[q - \theta_g\, c(q)]. \tag{2.3}$$

Maximizing with respect to q the optimal employment contract is characterized by

$$t^{NB} = \theta_g\, c(q^{NB}) \quad \text{and} \quad c'(q^{NB}) = \frac{1}{\theta_g}. \tag{2.4}$$

Superscript 'NB' stands for 'no bad workers'. Of course, the firm can also use this contract if the workers' types are observable. The optimal profit depends on the cost function and the fraction π of good workers. If this profit is higher than the maximum profit employing both types of workers, it is indeed optimal for the firm not to employ the bad workers at all.

Second, suppose the firm wants to employ both types of workers and productivity is unobservable. Given a direct revelation mechanism it then maximizes its profit

$$R = \pi[q_g - t_g] + (1-\pi)[q_b - t_b] \tag{2.5}$$

with respect to

$$\begin{array}{llcl}
(PCB) & u(\theta_b,\hat\theta_b) - S(\theta_b,\hat\theta_b) & \geq & 0, \\
(PCG) & u(\theta_g,\hat\theta_g) - S(\theta_g,\hat\theta_g) & \geq & 0, \\
(ICB) & u(\theta_b,\hat\theta_b) - S(\theta_b,\hat\theta_b) & \geq & u(\theta_b,\hat\theta_g) - S(\theta_b,\hat\theta_g), \\
(ICG) & u(\theta_g,\hat\theta_g) - S(\theta_g,\hat\theta_g) & \geq & u(\theta_g,\hat\theta_b) - S(\theta_g,\hat\theta_b).
\end{array}$$

Due to the incentive constraints (ICB) and (ICG) all workers reveal their type truthfully. The participation constraints (PCB) and (PCG) ensure that both types of workers accept. If there are only two types, usually only two constraints are binding and impose a real restriction on optimal contracts. The following lemma helps to determine which constraint may be discarded in the present setting.

Lemma 2.1

1. *$u(\theta_g,\hat\theta_k) \geq u(\theta_b,\hat\theta_k)$ and $S(\theta_g,\hat\theta_k) \leq S(\theta_b,\hat\theta_k)$ for all $k \in \{g,b\}$, and*

2. *$\frac{d}{dq_k}\{S(\theta_i,\hat\theta_k)\} \geq 0$ and $\frac{d}{dq_k}\{S(\theta_i,\hat\theta_m)\} \leq 0$ for all $i \in \{g,b\}$ and $k,m \in \{g,b\}$ with $k \neq m$.*

Proof: Choosing contract (q_k, t_k) good workers benefit equally from income but incur lower (or if $q_k = 0$, equal) production costs as compared to bad workers. Thus, their rent must be weakly higher. This weakly decreases unfavorable inequality if workers compare rents. If workers compare income, the suffering is independent of the workers' types and thus identical for good and bad workers.

Suppose q_k increases while all other production quantities are kept constant. This reduces the rent of workers choosing contract (q_k, t_k). If workers compare rents, their suffering can potentially only increase whereas the suffering of workers choosing another contract (q_m, t_m) can potentially only decrease. If workers compare income, the suffering is independent of production costs, and Part 2 is trivially satisfied with equality. *Q.E.D.*

Lemma 2.1 follows from the productivity advantage of the good workers. It implies

Lemma 2.2 *Suppose the firm employs both types of workers. If the bad workers' production quantity is strictly positive,* (PCG) *is not binding, whereas* (PCB) *and* (ICG) *are binding irrespectively of whether workers compare income or rents.*

Proof: First, (PCG) is not binding as due to

$$u(\theta_g, \hat{\theta}_g) - S(\theta_g, \hat{\theta}_g) \geq u(\theta_g, \hat{\theta}_b) - S(\theta_g, \hat{\theta}_b) > u(\theta_b, \hat{\theta}_b) - S(\theta_b, \hat{\theta}_b) \geq 0. \tag{2.6}$$

The first inequality stems from (ICG), the second from Lemma 2.1, and the third from (PCB). The second inequality is strict as a good worker has strictly lower production costs than the bad workers if their production quantity is strictly positive.

Second, (PCB) must be binding as otherwise the firm could decrease t_g and t_b by the same amount, keeping $t_g - t_b$ is constant but increasing its profit. As the production quantities are not changed, the difference in rents, all $S(\theta_i, \hat{\theta}_k)$, and consequently all incentive constraints are not affected.

Thirdly, (ICG) is binding. Otherwise the firm could increase q_g and its profit. This equalizes (ICG) as $u(\theta_g, \hat{\theta}_g)$ decreases, $S(\theta_g, \hat{\theta}_g)$ weakly increases, $u(\theta_g, \hat{\theta}_b)$ is left unchanged, and $S(\theta_g, \hat{\theta}_b)$ weakly decreases. Equally, increasing q_g leaves $u(\theta_b, \hat{\theta}_b)$ unchanged and weakly decreases $S(\theta_b, \hat{\theta}_b)$, thus causing no problems with (PCB). As further $u(\theta_b, \hat{\theta}_g)$ decreases and

$S(\theta_b, \hat{\theta})_g)$ weakly increases, (ICB) is softened. The non-binding good workers' participation constraint can be ignored. *Q.E.D.*

A good worker can always pretend to be bad. Since he then incurs lower production costs than the bad worker he has a higher rent and suffers weakly less from inequity. As the utility given to a bad worker is higher than zero in order to induce participation, a good worker receives more when pretending to be bad. Since he must get even more when announcing his type truthfully, participation of the good workers is no issue. However, the bad workers' participation constraint must be binding as the firm could otherwise increase its profit by decreasing the income of both types of workers by the same amount. This does not change the differences in income or rents so that the incentive constraints are unaffected. Finally, the good workers' incentive constraint must be binding. Otherwise the firm could raise its profit by increasing the good workers' production quantity. This makes the contract of the good workers less attractive. The bad workers' incentives constraints are further relaxed as the changes in suffering - more suffering if they lie and less if they report the truth - add to this. Participation of the bad workers is left unchanged as their contract - and therefore their rent - is not altered whereas any suffering once they have joined the firm is potentially reduced. As participation of the good workers is no issue that aspect can be ignored.

2.3.2 Workers Comparing Income

Throughout this section suppose workers compare income. Consider first the optimal employment contracts if the firm has to employ both types of workers. Due to the maximum functions in the workers' suffering their utility functions have a kink at $t_g = t_b$ and are thus not differentiable at this point. This can cause difficulties in the maximization problem, but fortunately incentive compatibility implies the following lemma. Superscript 'SB' stands for 'second-best'.

Lemma 2.3 *Suppose the firm cannot observe the workers' types, workers compare income, and the firm employs both types of workers. Then* $t_g^{SB} \geq t_b^{SB}$ *for all* $\alpha \geq 0$.

Proof: Suppose $t_b^{SB} > t_g^{SB}$ for some $\alpha \geq 0$ so that $S(\theta_g, \hat{\theta}_g) = (1-\pi)\,\alpha\,(t_b^{SB} - t_g^{SB})$ and $S(\theta_b, \hat{\theta}_b) = 0$. As (PCG) is not binding the optimal employment contracts must satisfy the following remaining constraints

$$\begin{array}{llcl} (PCB) & t_b^{SB} - \theta_b\, c(q_b^{SB}) & = & 0, \\ (ICG) & (t_g^{SB} - t_b^{SB})[1 + (1-\pi)\alpha] + \theta_g[c(q_b^{SB}) - c(q_g^{SB})] & = & 0, \\ (ICB) & (t_b^{SB} - t_g^{SB})[1 + (1-\pi)\alpha] - \theta_b[c(q_b^{SB}) - c(q_g^{SB})] & \geq & 0. \end{array}$$

Since $t_g^{SB} < t_b^{SB}$ by assumption, (ICG) implies $c(q_b^{SB}) - c(q_g^{SB}) > 0$ and

$$(t_b^{SB} - t_g^{SB})[1 + (1-\pi)\alpha] = -\frac{\theta_g[\, c(q_b^{SB}) - c(q_g^{SB})\,]}{(1-\pi)\,\alpha}.$$

After substitution of the above expression (ICB) is satisfied if and only if $(\theta_g - \theta_b)[\, c(q_b^{SB}) - c(q_g^{SB})\,] \geq 0$. As $\theta_b > \theta_g$ this contradicts $c(q_b^{SB}) > c(q_g^{SB})$ from (ICG). *Q.E.D.*

If the bad workers receive a higher income, the good workers reveal their type truthfully only if they have to produce less than the bad workers. Inequity aversion requires an even greater reduction in production as compensation for the suffering from getting a lower income. But bad workers profit more from a lower production quantity than good workers while their suffering is equal. Therefore, the bad workers must also prefer the good workers' contract to their own, and the considered contract cannot be incentive compatible. Building on Lemma 2.3 an *order constraint*

$$(OC) \quad t_g - t_b \geq 0$$

can be added to the constraint set without restricting the attainable maximum. Optimal contracts can now be derived using first order conditions. Note that all arguments for Lemma 2.2 leave the income difference $t_g - t_b$ constant. Therefore, Lemma 2.2 remains valid if the additional constraint $t_g - t_b \geq 0$ is introduced.

Consider the firm's maximization problem including the order constraint (OC). Suppose that the bad workers' production quantity is strictly positive so that Lemma 2.2 can be applied. Assume that (OC) and (ICB) are not binding. Solving the binding (PCB) and (ICG)

$$t_g^{SB} = \theta_g\, c(q_g^{SB}) + (\theta_b - \theta_g)\, c(q_b^{SB}), \tag{2.7}$$

characterizes the good worker's optimal income, whereas

$$t_b^{SB} = \theta_b\, c(q_b^{SB}) + \frac{\pi\alpha\,\theta_g}{1+\pi\alpha}\,[\, c(q_g^{SB}) - c(q_b^{SB})\,]. \tag{2.8}$$

characterizes the bad workers' optimal income.

If there is asymmetric information, the good workers must get an informational rent to make them reveal their type truthfully. Therefore, their income exceeds their production costs by $(\theta_b - \theta_g)\, c(q_b^{SB})$. Inequity aversion does not influence this informational rent. If a worker announces to be bad, he receives a lower income than the good workers truthfully announcing their type. If workers compare income, this suffering is independent of whether the worker is good or bad. In order to make a bad worker participate he must be compensated for this suffering by an income premium. The bad workers' income thus exceeds their production costs $\theta_b\, c(q_b^{SB})$. However, the income premium given to the bad workers also neutralizes the suffering of a good worker pretending to be bad. Even if workers are inequity averse, the informational rent given to the good workers thus remains $(\theta_b - \theta_g)\, c(q_b^{SB})$. As in the case absent inequity aversion, the informational rent reflects a good worker's save on production costs as compared to a bad worker when both satisfy the bad workers' employment contract.

With inequity aversion the bad workers must be compensated for the income difference. The latter is thus reduced to

$$t_g^{SB} - t_b^{SB} = \frac{\theta_g\,[\,c(q_g^{SB}) - c(q_b^{SB})\,]}{1+\pi\alpha}. \tag{2.9}$$

Consequently, the order constraint is not binding if and only if $q_g^{SB} \geq q_b^{SB}$. Substitution of t_g^{SB} and t_b^{SB} shows that (ICB) is then equally not binding.

Substitution of t_g^{SB} and t_b^{SB} turns the firm's profit into a function strictly concave in q_g^{SB} and q_b^{SB}. By Assumption 2.1 the unique optimal production quantities exist and are given by

$$c'(q_g^{SB}) = \frac{1+\pi\alpha}{\theta_g\,(1+\alpha)} \quad \text{and} \quad c'(q_b^{SB}) = \frac{1-\pi}{\theta_b - \pi\,\theta_g\,\frac{1+\alpha}{1+\pi\alpha}}. \tag{2.10}$$

Inspection of q_b^{SB} supports the initial assumption that the bad workers' production quantity is strictly positive for all α. Note that as long as (OC) is not binding, the envelope theorem implies

$$\frac{d\,R}{d\,\alpha} = -\,\frac{\pi^2\,[\,c(q_g) - c(q_b)\,]}{(1+\pi\alpha)^2} < 0 \tag{2.11}$$

and the firm's profit is strictly falling in α as the solution is valid if and only if $q_g \geq q_b$ and thus $c(q_g) \geq c(q_b)$.

The order constraint is indeed not binding if the workers are not inequity averse, $q_g^{SB} \geq q_b^{SB}$ at $\alpha = 0$. However, q_b^{SB} is strictly increasing whereas q_g^{SB} is strictly decreasing in α by the strict convexity of c. Thus, the income difference strictly decreases in α as long as the above solution remains valid. Both q_g^{SB} and q_b^{SB} converge as α goes to infinity. Comparing the corresponding limit expressions yields the following result. If $\pi \geq \theta_g/\theta_b$ then q_g^{SB} converges to something weakly larger than q_b^{SB}. In this case (OC) and (ICB) remain non-binding and the above solution valid for all α.

However, if $\pi < \theta_g/\theta_b$, there exists a cutoff $\tilde{\alpha}$ so that for all $\alpha > \tilde{\alpha}$ the order constraint and the bad worker's incentive constraint become binding. As both types of worker receive the same income there is no suffering, $S(\theta_b, \hat{\theta}_b) = S(\theta_g, \hat{\theta}_g) = 0$. A worker can only be indifferent between two contracts with equal income levels if the production quantities are identical. The binding incentive constraint of the good workers and participation constraint of the bad workers then imply

$$t^{SB} := t_g^{SB} = t_b^{SB} = \theta_b c(q_b^{SB}) \quad \text{and} \quad q^{SB} := q_g^{SB} = q_b^{SB} \tag{2.12}$$

which automatically satisfies (ICB). Substitution makes the firm's profit function strictly concave in q^{SB} and

$$c'(q^{SB}) = \frac{1}{\theta_b} \tag{2.13}$$

characterizes the optimal production quantity which is strictly positive and unique by Assumption 2.1. Note that in this case q^{SB} and thus the firm's profit is independent of the degree α of inequity aversion. These results are summarized in the following proposition.

Proposition 2.1 *Suppose the firm cannot observe the workers' types, workers compare income, and the firm employs both types of workers.*

1. *If $\pi \geq \theta_g/\theta_b$, the firm's maximum profit and the income difference are strictly decreasing in α. The income difference is converging to zero as α goes to infinity.*

2. *If $\pi < \theta_g/\theta_b$, there exists a cutoff $\tilde{\alpha} > 0$ so that for all $\alpha < \tilde{\alpha}$ the income difference and the firm's maximum profit are strictly decreasing in α, whereas for all $\alpha \geq \tilde{\alpha}$ the firm offers both types of workers a single contract no longer changing in α.*

Consider an increase of the good workers' production quantity. As the good workers incur higher production costs they must be given a higher income. This increases the income difference for which the bad workers must be compensated. Their consequent increase in income is an additional cost for the firm. Consider next an increase of the bad workers' production quantity. As the bad workers incur higher production costs they must be given a higher income. As this decreases the income difference they enjoy an additional utility and their production costs may increase by more than their income. Thus, inequity aversion makes production by the good workers comparatively more expensive, and production by the bad workers comparatively cheaper. The firm reacts to this change in 'real' production costs by increasing the production by the bad while decreasing production by the good workers. Ultimately this compresses income levels.

Building upon these results the firm's employment decision can be considered. Unfortunately, the cost function c determines absolute profits and thus strongly influences the firm's comparison between employing all workers and employing only good workers. However, there is the following result for a special cost function.

Proposition 2.2 *Suppose the firm cannot observe the workers' types, workers compare income, and* $c(q) = q^2/2$. *If* $\pi \leq \theta_g/\theta_b$, *it is always optimal for the firm to employ both types of workers. If* $\pi > \theta_g/\theta_b$, *there exists a cutoff* $\hat{\alpha}$ *so that for all* $\alpha > \hat{\alpha}$ *it is optimal to employ only the good workers.*

Proof: If the firm employs only the good worker, the optimal production quantity is given by $q^{SB} = 1/\theta_g$ yielding profit $R = \pi/(2\,\theta_g)$. If the firm employs both types of workers, Proposition 2.1 implies that its profit is continuous and falling in α. Thus, it is sufficient to compare the profit of employing only the good workers with the minimum profit of employing all workers, that is the limit of the maximum profit as α goes to infinity. Depending on π there are two cases.

First, suppose $\pi < \theta_g/\theta_b$. As α eventually exceeds the cutoff $\tilde{\alpha}$ in Proposition 2.1 the firm offer a single contract to both types of workers. Given the assumed cost function the optimal contract is characterized by $q^{SB} = 1/\theta_b$ and $t^{SB} = 1/(2\theta_b)$ yielding an limit profit of $1/(2\,\theta_b)$ as α goes to infinity. But $\pi < \theta_g/\theta_b$ implies $1/(2\,\theta_b) < \pi/(2\,\theta_g)$ so that it is never optimal to only employ the good workers for all α smaller than infinity.

Second, suppose $\pi \geq \theta_g/\theta_b$. In that case the principal offers two different contracts for all α. As α goes to infinity, the optimal production quantities converge to $q_b^{SB} = (1-\pi)/(\theta_b - \theta_g)$ and $q_g^{SB} = \pi/\theta_g$ yielding an limit profit of $[\pi^2\,\theta_b + \theta_g(1-2\,\pi)]/[2\theta_g\,(\theta_b - \theta_g)]$. Comparing profits it is optimal to employ only the good workers if and only if

$$\frac{(1-\pi)(\theta_g - \pi\,\theta_b)}{\theta_g\,(\theta_b - \theta_g)} < 0$$

which is satisfied if and only if $\pi > \theta_g/\theta_b$. In these cases there exists a cutoff $\tilde{\alpha}$ so that for all $\alpha \geq \tilde{\alpha}$ it is optimal to only employ the good workers. If $\pi = \theta_g/\theta_b$ the firm is indifferent between employing only the good workers and employing both types of workers in the limit. Thus, it strictly prefers to employ both types of workers for all α smaller than infinity. *Q.E.D.*

If both types of workers are employed and receive different income levels, those workers with the lower income must be compensated for their suffering. This cost is independent of any production quantities and can only be reduced by decreasing the income difference as such. Even after adjusting production quantities to a certain income difference this strongly distorts the good workers' contract. If types are very different in their productivity and there are many good workers, this distortion can become large enough as to make it optimal not to employ bad workers at all.

2.3.3 Workers Comparing Rents

Results so far seem to suggest that inequity aversion could be an explanation for income compression and might hint at the reasons for involuntary unemployment of workers with low productivity. However, the impact of inequity aversion reverses direction if workers compare rents. In this case inequity aversion increases the income difference while never causing an exclusion of bad workers. Analogous to the Order Constraint (OC) the following lemma eliminates the maximum functions in the workers' utility.

Lemma 2.4 *Suppose the firm cannot observe the workers' types and workers compare rents. Then $u(\theta_g, \hat{\theta}_g) < u(\theta_b, \hat{\theta}_b)$ cannot be incentive compatible.*

Proof: Suppose $u(\theta_g, \hat{\theta}_g) < u(\theta_b, \hat{\theta}_b)$ and thus $S(\theta_g, \hat{\theta}_g) > 0$ and $S(\theta_b, \hat{\theta}_b) = 0$. As $u(\theta_g, \hat{\theta}_k) > u(\theta_b, \hat{\theta}_k)$ for all $k \in \{g, b\}$ it must be that $S(\theta_g, \hat{\theta}_b) = 0$. Bringing $S(\theta_g, \hat{\theta}_g)$ on the left hand side of (ICG) this implies $u(\theta_g, \hat{\theta}_g) \geq S(\theta_g, \hat{\theta}_g) + u(\theta_g, \hat{\theta}_b) > u(\theta_b, \hat{\theta}_b)$. This contradicts $u(\theta_g, \hat{\theta}_g) < u(\theta_b, \hat{\theta}_b)$. *Q.E.D.*

Suppose the bad workers get a higher rent than the good workers. First, due to his lower production costs a good worker receives a higher rent than the bad workers if he falsely pretends to be bad. If the bad workers get a higher rent than the good worker announcing their true type, a good worker must receive a higher rent reporting to be bad as compared to announcing the truth. Second, a good worker suffers more from inequity if he reports his type truthfully as compared to pretending to be bad. Thus, inequity aversion has no positive impact on incentives, a good worker has no incentives to reveal his type truthfully, and contracts can only be incentive compatible if the good workers receive a rent at least as high as the bad workers. The impact of inequity aversion can now be summarized as follows.

Proposition 2.3 *Suppose the firm cannot observe the workers' types and workers compare rents. Then it is always optimal to employ both types of workers, and the income difference is increasing in* α.

Proposition 2.3 follows from the optimal contracts characterized as follows. If types are private information, good workers must be given an informational rent in order to induce them to reveal their true type. Since bad workers never get a rent, asymmetric information and screening cause a rent inequality.

Optimal employment contracts account for this in the following way. Lemma 2.4 implies that good workers get a weakly higher rent so that only bad workers suffer from inequity and therefore $S(\theta_b, \hat{\theta}_b) = \pi\,\alpha[t_g - \theta_g\,c(q_g) - t_b + \theta_b\,c(q_b)]$ and $S(\theta_g, \hat{\theta}_g) = 0$. Suppose that the bad workers' production quantity is strictly positive so that Lemma 2.2 is applicable. Assuming (ICB) to be non-binding, solving the binding (PCB) and (ICG) then yields

$$t_g^{SB} = \theta_g\,c(q_g^{SB}) + (1 + \pi\,\alpha)(\theta_b - \theta_g)\,c(q_b^{SB}) \tag{2.14}$$

as the good workers' optimal income and

$$t_b^{SB} = \theta_b\,c(q_b^{SB}) + \pi\,\alpha(\theta_b - \theta_g)\,c(q_b^{SB}) \tag{2.15}$$

as the bad workers' optimal income. Note that both types of workers receive an income premium in excess of their production costs. Subtracting these premia the rent inequality is given by $(\theta_b - \theta_g)\,c(q_b^{SB})$. This expression coincides with the informational rent required to induce truthful revelation of the good workers' type if there is no inequity aversion. The bad workers only participate if they receive compensation $\pi\,\alpha(\theta_b - \theta_g)\,c(q_b^{SB})$ for their suffering from inequity. As the good workers' incentive constraint is binding a good worker pretending to be bad gets the same rent as the remaining good workers revealing their type truthfully. Thus, a good worker receives the bad workers' compensation for rent inequity as an additional rent. His income premium $(1+\pi\,\alpha)\,(\theta_b-\theta_g)\,c(q_b^{SB})$ thus exceeds the traditional informational rent and increases with the degree α of inequity aversion. Consequently, inequity aversion influences the good workers' income premium if workers compare rents.

Substitution of the optimal income levels shows that (ICB) is indeed non-binding as long as $c(q_g^{SB}) - c(q_b^{SB}) > 0$. Moreover, the profit function is strictly concave in the production quantities q_g and q_b. Maximization yields

$$c'(q_g^{SB}) = \frac{1}{\theta_g} \quad \text{and} \quad c'(q_b^{SB}) = \left[\theta_b + (\theta_b - \theta_g)\left[\frac{\pi(1+\alpha)}{1-\pi}\right]\right]^{-1} \tag{2.16}$$

as the unique solution characterizing the optimal production quantities. Note that q_b^{SB} is strictly positive for all α and Lemma 2.2 may indeed be applied. Moreover, $q_g^{SB} > q_b^{SB}$ at $\alpha = 0$, and since q_g^{SB} is constant whereas q_b^{SB} is decreasing in α the solution remains valid for all $\alpha \geq 0$.

If workers compare rents, inequity aversion influences optimal employment contracts only via the informational rent given to the good workers. As the degree α of inequity aversion increases so does the compensation of the bad workers and the income premium given to the good workers. These costs caused by inequity aversion can only be reduced by lowering the bad workers' production quantity. Thus, the income of both bad and good workers may be reduced. However, the marginal decrease of the good workers' income given by $(1+\pi\alpha)\theta_b\,c'(q_b) - \theta_g(1+\pi\alpha)c'(q_b)$ is smaller than the marginal decrease of the bad workers' income given by $(1+\pi\alpha)\theta_b\,c'(q_b) - \theta_g\pi\alpha c'(q_b)$. Equivalently, using the optimal income levels the income difference is given by

$$t_g^{SB} - t_b^{SB} = \theta_g\,[\,c(q_g^{SB}) - c(q_b^{SB})\,],$$

where q_g^{SB} stays constant whereas q_b^{SB} decreases with increasing α. Summarizing, the income difference increases as inequity concerns and the associated costs gain more importance.

As the rent inequality is independent of the good workers' production quantity, the latter may be chosen as if there were no asymmetric information or inequity aversion. There is 'no distortion at the top'. Furthermore, suppose the good workers get a contract characterized by $t_g^{SB} = \theta_g c(q_g^{SB})$. As they then receive no rent the firm can employ the bad workers by offering them a contract $(q_b, t_b) = (0, 0)$ requiring them to produce nothing while giving them no income. If a bad worker accepts this contract, he gets zero rent. Since there is no rent inequity both types participate. Additionally, if a bad worker lies and pretends to be good, he gets a negative rent as his high production costs exceed the income just sufficient for good workers. Due to his negative rent he also suffers from inequity. Thus, a bad worker has no incentives to misrepresent his type. If a good worker pretends to be bad, he gets zero rent. Since this is not more than what he gets by announcing his type truthfully he reports his true type in equilibrium. Summarizing, there are no problems with respect to incentives and participation when introducing such a 'null-contract'. Moreover, such a contract causes zero costs and the firm can employ the bad workers without affecting its profit. Yet once employed it is optimal to have the bad workers produce a positive quantity. Marginally increasing the production quantity of the bad workers marginally increases the firm's profit by

$$\frac{d\,R}{d\,q_b} = 1 - \pi - c'(q_b)\,[(1-\pi)\,\theta_b + (\theta_b - \theta_g)\pi(1+\alpha)\,]. \tag{2.17}$$

As the workers' cost function satisfies the Inada conditions marginal costs converge to zero as production goes to zero. However, the marginal increase in profit of the firm is always $1 - \pi$. Since $c'(q_b)$ converges to zero as q_b goes to zero, (2.17) converges to $1 - \pi > 0$. Marginal benefits eventually dominate marginal costs for sufficiently small but strictly positive production quantities, and it is optimal to always employ both types of workers and have them produce a strictly positive quantity.

2.3.4 The Impact of Asymmetric Information

This subsection investigates the interaction of inequity aversion and asymmetric information. It characterizes the firm's optimal employment decision and employment contracts if there is no asymmetric information. It then contrasts the obtained results with the findings in the previous subsections. Before getting into the details the following general result simplifies the analysis. If the workers' types are observable, incentive constraints are obsolete and the firm is only restricted by the participation constraints

$$\begin{aligned} (PCB) \quad & u(\theta_b, \hat{\theta}_b) - S(\theta_b, \hat{\theta}_b) && \geq 0 \\ (PCH) \quad & u(\theta_g, \hat{\theta}_g) - S(\theta_g, \hat{\theta}_g) && \geq 0. \end{aligned}$$

As the firm could otherwise increase q_g or q_b, both constraints have to be binding independently of whether workers compare income or rents. Let q_g^* and q_b^* characterized by

$$c'(q_g^*) = \frac{1}{\theta_g} \quad \text{and} \quad c'(q_b^*) = \frac{1}{\theta_b}. \tag{2.18}$$

denote the optimal production quantities if there is no asymmetric information and workers are not inequity averse.

First, suppose workers compare income. As in the case of unobservable productivity, an order constraint can be introduced without affecting the attainable maximum profit of the firm. However, this time the precise form of the order constraint depends on the optimal production costs if there is no inequity aversion. The following lemma holds, where superscript 'FB' stands for 'first-best', that is, observable productivity .

Lemma 2.5 *Suppose the firm can observe the workers' types, workers compare income, and the firm employs both types of workers.*

1. *If $\theta_g\, c(q_g^*) \geq \theta_b\, c(q_b^*)$, then $t_g^{FB} \geq t_b^{FB}$ for all $\alpha \geq 0$.*

2. *if $\theta_b\, c(q_b^*) \geq \theta_g\, c(q_g^*)$, then $t_b^{FB} \geq t_g^{FB}$ for all $\alpha \geq 0$.*

Proof: Consider the case $\theta_g\, c(q_g^*) \geq \theta_b\, c(q_b^*)$. Suppose $t_b^{FB} > t_g^{FB}$. Solving the binding (PCG) and (PCB) the optimal income levels are given by

$$t_g^{FB} = \theta_g c(q_g^{FB}) + \frac{(1-\pi)\alpha}{1+(1-\pi)\alpha}\left[\theta_b\, c(q_b^{FB}) - \theta_g\, c(q_g^{FB})\right] \quad \text{and} \quad t_b^{FB} = \theta_h c(q_b^{FB}). \tag{2.19}$$

This yields an income difference of

$$t_b^{FB} - t_g^{FB} = \frac{\theta_b\, c(q_b^{FB}) - \theta_g\, c(q_g^{FB})}{1+(1-\pi)\alpha}. \tag{2.20}$$

The solution satisfies $t_b^{FB} > t_g^{FB}$ if and only if $\theta_b\, c(q_b^{FB})$ is strictly larger than $\theta_g\, c(q_g^{FB})$. After substitution of the optimal income levels the firm's profit function is strictly concave in both production quantities. Because of the limit assumptions on $c'(q)$ the first order conditions have a unique solution and the optimal production quantities are given by

$$\theta_g\, c'(q_g^{FB}) = 1 + (1-\pi)\alpha \quad \text{and} \quad \theta_b\, c'(q_b^{FB}) = \frac{1+(1-\pi)\alpha}{1+\alpha}. \tag{2.21}$$

With increasing α this implies that q_g^{FB} strictly increases, q_b^{FB} strictly decreases, and the income difference in (2.20) strictly decreases. As α goes to zero q_g^{FB} and q_b^{FB} converge to q_g^* and q_b^*. Since $\theta_g\, c(q_g^*) \geq \theta_b\, c(q_b^*)$ the income difference is weakly negative at $\alpha = 0$ and becomes strictly negative as α increases. Thus, the solution violates $t_b^{FB} > t_g^{FB}$ for all α.

Part 2: Consider next the case $\theta_g\, c(q_g^*) \leq \theta_b\, c(q_b^*)$. Suppose $t_g^{FB} > t_b^{FB}$. Solving the binding (PCG) and (PCB) for the optimal income levels yields

$$t_g^{FB} = \theta_g c(q_g^{FB}) \quad \text{and} \quad t_b = \theta_b c(q_b^{FB}) + \frac{\pi\alpha}{1+\pi\alpha}\left[\theta_g c(q_g^{FB}) - \theta_b c(q_b^{FB})\right]. \tag{2.22}$$

The income difference is then given by

$$t_g^{FB} - t_b^{FB} = \frac{\theta_g\, c(q_g^{FB}) - \theta_b\, c(q_b^{FB})}{1+\pi\alpha}. \tag{2.23}$$

Thus, this solution satisfies $t_g^{FB} > t_b^{FB}$ as long as $\theta_g\, c(q_g^{FB})$ is strictly larger than $\theta_b\, c(q_b^{FB})$. By substitution and maximization the optimal production quantities are given by

$$\theta_g\, c'(q_g^{FB}) = \frac{1+\pi\alpha}{1+\alpha} \quad \text{and} \quad \theta_b\, c'(q_b^{FB}) = 1 + \pi\alpha. \tag{2.24}$$

As q_g^{FB} is strictly decreasing and q_b^{FB} is strictly increasing in α, the income difference in (2.23) is strictly decreasing in α. As α goes to zero q_g^{FB} and q_b^{FB} converge to q_g^* and q_b^*. As $\theta_g\, c(q_g^*) \leq \theta_b\, c(q_b^*)$, the solution violates $t_g^{FB} > t_b^{FB}$ for all $\alpha \geq 0$. *Q.E.D.*

With the use of Lemma 2.5 optimal employment contracts can be easily computed. As summarized by the following proposition the impact of inequity aversion is essentially independent of whether there is asymmetric information or not.

Proposition 2.4 *Suppose the firm can observe the workers' types, workers compare income, and the firm employs both types of workers. If $\theta_g\, c(q_g^*) \neq \theta_b\, c(q_b^*)$, there exists a cutoff $\tilde{\alpha}$ so that*

1. *if $\alpha < \tilde{\alpha}$, the income difference and the firm's maximum profit are strictly decreasing in α.*

2. *if $\alpha \geq \tilde{\alpha}$, the firm offers two contracts with identical income levels but different production quantities. Both contracts no longer change in α.*

If $\theta_g\, c(q_g^*) = \theta_b\, c(q_b^*)$, inequity aversion has no impact on optimal employment contracts.

Proof: First, suppose $\theta_g\, c(q_g^*) > \theta_b\, c(q_b^*)$, the good workers receive a higher income than the bad workers if there is no inequity aversion. By Lemma 2.5 the order constraint

$$(OC) \quad t_g^{FB} - t_b^{FB} \geq 0$$

can be added to the constraint set without restricting the firm's maximum profit.

Assume that the order constraint is not binding. Solving the binding (PCG) and PCB) yields optimal income levels as characterized in (2.22). Substitution and maximization yields optimal production quantities as characterized in (2.24). Since $\theta_g\, c(q_g^*) > \theta_b\, c(q_b^*)$ the income difference is strictly positive at $\alpha = 0$ and the solution is indeed valid for small α. However, as α goes to infinity q_g^{FB} exceeds q_b^{FB} at some point. Therefore, there exists a cutoff $\tilde{\alpha}$ so that $\theta_g\, c(q_g^{FB}) < \theta_b\, c(q_b^{FB})$ for all larger $\alpha \geq \tilde{\alpha}$. In this case the solution is invalid and the order constraint must be binding.

Suppose the order constraint is binding so that the firm must set $t_g^{FB} = t_b^{FB}$. As the firm can observe the workers' types it can offer two contracts with different production quantities q_g and q_b but the same income t^{FB}. Solving the participation constraints yields

$$t^{FB} = \theta_b c(q_b^{FB}) \quad \text{and} \quad c(q_g^{FB}) = \frac{\theta_b}{\theta_h}\, c(q_b^{FB}). \tag{2.25}$$

Unfortunately, the firm's profit function is not necessarily strictly concave in q_b after substitution of t^{FB} and q_g^{FB}. As done frequently in similar situations the following trick is applied. Define $v(y)$ as the inverse function of $c(y)$, that is $v(y) := c^{-1}(y)$. As $c(q)$ is strictly increasing and convex, $v(y)$ is strictly increasing and concave. Moreover,

$$v'(y) = \frac{1}{c'(v(y))}. \tag{2.26}$$

Thus, $\lim_{y\to c(0)} v'(y) = +\infty$ and $\lim_{y\to\infty} v'(y) = 0$. Define $x := c(q_b)$. The firm's maximization problem can be restated as maximizing its profit

$$R = \pi v(\theta_b\, x/\theta_g) + (1-\pi)\, x - \theta_b\, x \tag{2.27}$$

with respect to x. The firm's profit function is now strictly concave in x, and the first order condition has a unique solution by the assumptions on $c'(q)$. Therefore,

$$\pi \frac{\theta_b}{\theta_g}\, v'(\theta_b\, x/\theta_g) + (1-\pi) v'(x) = \theta_b \tag{2.28}$$

characterizes the optimal production costs x, and thus the optimal production quantities q_b^{FB} with $c(q_b^{FB}) = x$ and $q_g^{FB} = v(\theta_b\, x/\theta_g)$.

For the case $\theta_g\, c(q_g^*) > \theta_b\, c(q_b^*)$ the first-best employment contracts can now be characterized as follows. At $\alpha = 0$ the firm offers the first-best contracts absent inequity aversion. These give the good workers a higher income than the bad workers. As inequity aversion increases the income difference is compressed. For any production quantities the income of the bad workers is strictly increasing in α whereas the income of the good workers is constant. The envelope theorem thus implies that the firm's maximum profit is strictly decreasing in α as long as the solution remains valid. As α surpasses some cutoff $\tilde{\alpha}$ the firm offers two employment contracts giving both types of workers the same income, but the good workers have to produce more. These contracts remain unchanged as α increases further.

If $\theta_g\, c(q_g^*) < \theta_b\, c(q_b^*)$, the bad workers are given the higher first-best income absent inequity aversion. In this case the order constraint is given by $t_b^{FB} - t_g^{FB} \geq 0$ and the optimal employment contracts are as characterized in (2.19) and (2.21). The remaining argument stays the same.

Finally, suppose $\theta_g\, c(q_g^*) = \theta_b\, c(q_b^*)$. In this case the firm accidentally gives both types of workers the same income even absent inequity aversion. Contracts are not changed as α increases, and inequity aversion has no impact. *Q.E.D.*

If both types of workers are paid their production costs, they usually receive different income levels. Note that contrary to the case of asymmetric information it may well happen that the bad workers receive a higher income than the good workers. Given an income difference those workers with the lower income suffer from inequity and must be compensated. The ensuing logic is identical to the case of asymmetric information and inequity aversion causes an income compression. However, if both types of workers are (accidentally) paid the same income absent inequity aversion, inequity aversion has no impact. Further, just as in the case of asymmetric information it might be optimal for the firm to exclude the bad workers. Again, results depend on the cost function, but the following proposition holds for a special cost function.

Proposition 2.5 *Suppose the firm can observe the workers' types, workers compare income, and* $c(q) := q^2/2$. *If* $\theta_g/\theta_g \geq 1/4$, *it is always optimal for the firm to employ both types of workers. If* $\theta_g/\theta_g < 1/4$, *there exist pairs of cutoffs* $(\hat{\pi}, \hat{\alpha})$ *with* $\hat{\pi} < 1$ *so that for all* $\pi > \hat{\pi}$ *and* $\alpha > \hat{\alpha}$ *it is optimal to employ only the good workers.*

Proof: If the firm employs only the good worker, the optimal production quantity is given by $q^{NB} = 1/\theta_g$ yielding profit $R^{NB} = \pi/(2\theta_g)$. Superscript 'NB' stands for 'no bad workers'. By Proposition 2.4 the firm's maximum profit is continuous and weakly decreasing in α. Thus, it is sufficient to compare the profit of employing only the good workers with the maximum profit of employing both types as α goes to infinity.

According to Proposition 2.4 the firm offers two employment contract with identical income $t^{FB} = \theta_b\,(q_b^{FB})^2/2$ and production quantity $q_g^{FB} = q_b^{FB}\,\sqrt{\theta_b/\theta_g}$. Maximization with respect to q_b yields

$$q_b^{FB} = \frac{\pi\sqrt{\theta_b/\theta_g} + 1 - \pi}{\theta_b}.$$

The maximum profit is then given by

$$R^{FB} = \frac{[\pi\sqrt{\theta_g\theta_b} + \theta_g\,(1-\pi)]^2}{2\,\theta_b\,\theta_g^2}.$$

The difference in profits between employing only the good workers and employing both types of workers is given by

$$R^{NB} - R^{FB} = \frac{(1-\pi)[\pi(\theta_b + \theta_g - 2\sqrt{\theta_g\theta_b}) - \theta_g]}{2\,\theta_g\,\theta_b}.$$

For π close to zero the above expression is negative and it is optimal to employ both types of workers. However, $R^{NB} - R^{FB}$ is strictly increasing in π since $\theta_b + \theta_g - 2\sqrt{\theta_g\theta_b}$ is strictly positive as $(\theta_b + \theta_g)/2 > \sqrt{\theta_b + \theta_g}$ for all $\theta_g \neq \theta_b$ and $\theta_b > \theta_g > 0$. As π goes to one the limit of $R^{NB} - R^{FB}$ is weakly negative if and only if $\theta_g/\theta_b \geq 1/4$.

If instead $\theta_g/\theta_b < 1/4$, there exists a cutoff $\hat{\pi} < 1$ so that $R^{NB} - R^{FB}$ is strictly positive for all $\pi > \hat{\pi}$. For every $\pi > \hat{\pi}$ there then exists a cutoff $\hat{\alpha}$ so that it is optimal to employ only the good workers for all $\alpha > \hat{\alpha}$. *Q.E.D.*

The intuition for this result is identical to the case of asymmetric information. Summarizing, the impact of inequity aversion on the firm's employment decision and employment contracts seems to be independent of whether productivity is observable or not. However, as shown in the following proposition the consequences of inequity aversion change with the informational structure if workers compare rents.

Proposition 2.6 *Suppose the firm can observe the workers' types, workers compare rents, and the firm employs both types of workers. Then inequity aversion has no impact on optimal employment contracts.*

Proof: Suppose $u(\theta_g, \hat{\theta}_g) \geq u(\theta_g, \hat{\theta}_g)$, the good workers receive a weakly higher rent than the bad workers. As then $S(\theta_g, \hat{\theta}_g) = 0$ the binding (PCG) yields $u(\theta_g, \hat{\theta}_g) = 0$ and therefore $t_g^{FB} = \theta_g\, c(q_g^{FB})$. But then $S(\theta_b, \hat{\theta}_b) = 0$, and (PCB) implies $u(\theta_g, \hat{\theta_g}) = 0$. Therefore, $t_g^{FB} = \theta_g\, c(q_g^{FB})$. Substitution of the optimal income levels and maximization with respect to q_g^{FB} and q_b^{FB} yields the same first-best contracts as if there was no inequity aversion. The same argument holds for $u(\theta_g, \hat{\theta}_g) \leq u(\theta_g, \hat{\theta}_g)$. *Q.E.D.*

If the workers' types are observable, the firm can extract all rents by paying an income just covering production costs. Both types get an equal rent of zero, and inequity aversion is irrelevant. Artificially creating a rent inequity by giving one type of workers a rent above zero makes participation of the other type more difficult and expensive. As only participation matters if types are observable, it is optimal to extract all rents from both types of workers.

If workers account imperfectly for production costs, comparing income and comparing rents generates countervailing effects. The corresponding relative magnitudes determine the net consequences of inequity aversion, which are thus influenced by the informational structure. Indeed, the effect stemming from comparing rents simply vanishes if productivity is observable. Thus, one could conjecture that inequity aversion is a better explanation for income compression and unemployment of low-skilled workers if the firm is well informed about its prospective workforce.

2.4 Discussion

2.4.1 Endogenous Outside Option

This section discusses the assumptions taken with respect to the workers' reference groups. Suppose unemployed workers compare themselves with employed workers. In this case each unemployed worker's utility depends on what fraction of workers accepts which employment contract. A worker's outside option is thus endogenous, and the firm might try to profit from this by deliberately softening the participation constraints. In the ensuing argument assume that income and rent of unemployed workers are normalized to zero. As before, the impact of inequity aversion depends on what workers compare.

If workers compare income, the firm profits from inequity aversion. Increasing the income in a certain employment contract has the following effects. First, a worker choosing the considered employment contract receives a higher utility. Thus, acceptance is made more attractive and the worker's production quantity may be increased. Second, each worker's outside option decreases. An unemployed worker receives no and therefore less income than the employed workers. As he wants to avoid this, participation is facilitated. Thirdly, the income difference between employed workers with different income levels is changed. If the income difference increases, the firm might have to raise the compensation necessary for those workers with the lower income. However, it can be shown that the positive first and second effect dominate even the negative last effect. Inequity aversion thus increases the firm's profit. This argument is essentially independent of whether the workers' types are observable or not.

The results change if workers compare rents. If the workers' types are observable, the firm can extract all rents from good and bad workers. As employed and unemployed workers receive zero rent, an unemployed worker does not suffer from inequity. Is it possible for the firm to soften the participation constraint of one type of worker by giving the other type a positive, higher rent? Without loss of generality suppose good workers receive a positive rent at least weakly larger then the rent given to bad workers. Consider a bad worker. Whether he accepts the offered employment contract or not, he always receives a lower rent than the good workers. If his rent upon accepting employment is weakly larger than zero, he suffers less if he accepts work. Therefore, a bad worker accepts if he thus receives a weakly positive rent. However, his acceptance decision is just the same as if the good workers were given no rent. As giving the good agents a positive rent causes costs without having a positive impact on participation of the bad workers, the firm should extract all rents from both types of workers, and inequity aversion is irrelevant.

Almost the same argument holds if the workers' types are unobservable. As Lemma 2.1 and 2.4 still hold if the workers' outside option is endogenous, only the bad workers' participation constraint and the good workers' incentive constraint must be binding. Moreover, good workers receive a weakly larger rent than bad workers. Due to the same argument as above, a bad worker accepts whenever his rent is weakly larger than zero, and his participation constraint is the same with or without inequity aversion. Consequently, bad workers receive no income premium for the suffering caused by the informational rent. However, if workers compare rents, the good workers' contract is only affected via this income premium of the bad workers. If the bad workers receive no such premium, inequity aversion has no effect on the good workers' contract, and inequity aversion is thus irrelevant. Since these results are fundamentally different to Proposition 2.3, this shows that changing the workers' reference group can greatly influence the consequences of inequity aversion.

2.4.2 Heterogenous and Unobservable Preferences

Clearly, not all workers have the same concerns for equity and fairness. Suppose workers are heterogeneous with respect to their preferences: some workers are inequity averse whereas some workers are 'selfish' and exclusively care for their own rent. Further, assume that all inequity averse workers have identical concerns for inequity, workers differ with respect to their productivity, and all characteristics are independent. Consequently, a worker's type depends on his productivity and his degree of inequity aversion. This additional heterogeneity in preferences might make it possible and desirable for the firm to further discriminate between the workers. For the sake of the following argument consider a selfish and an inequity averse worker with identical productivity. Suppose there are two contracts A and B, and that both types of workers get a higher rent choosing contract A. Only interested in his rent the selfish worker prefers contract A.

If workers compare rents, screening with respect to preferences is impossible. If he only cared for his rent, the inequity averse worker would also prefer contract A. Moreover, inequity aversion generates an additional preference for contract A: since the worker compares rents he can only suffer less from inequity when selecting contract A. Thus, the inequity averse worker must prefer contract A to contract B, and separation is impossible.

Yet if the workers compare income, the firm may be able to screen the workers with respect to their preferences. Suppose that the firm offers another contract C which specifies an income higher than the income levels in contract A or B, and that this contract C is accepted by a strictly positive fraction of workers. Although he then receives a lower rent, the inequity averse worker might now select contract B as he can thus reduce his suffering with respect to the workers choosing contract C.

Preliminary computations seem to hint in the following direction. First, the firm might reduce the informational rent required by the good inequity averse workers. These workers are offered an employment contract with very high income but relatively low rent. Good selfish workers do not accept this contract as they are granted a higher rent if they announce their true type. However, no single good inequity averse worker mimics the good selfish workers as he then receives a lower income than the remaining other good inequity averse workers revealing their type truthfully. All bad workers do not select any of the good workers' contracts as then their income does not cover their high production costs. In equilibrium the bad inequity averse workers suffer from inequity and must be granted an income premium as compensation. Increasing their income reduces unfavorable inequity and thus has a strong positive effect on their utility. They probably receive a relatively high income but have to produce a lot. Finally, the firm might have to give the bad selfish workers an informational rent in order to prevent them from mimicking the bad inequity averse workers.

Computations are complicated as the firm has to screen the workers in two dimensions. However, further investigations could yield interesting insights into the internal organization of the firm.

2.4.3 Optimal Income Taxation

The results of the present paper can be easily applied to a number of other areas. An obvious application is optimal income taxation if the citizens' productivity is unknown. Consider the following renaming of variables. Let $\theta_i c(q)$ denote the type-dependent costs of producing income q. If t is a citizen's net-income, $q - t$ represents the taxes collected by the state. The results of the present paper then correspond to a model of optimal income taxation in which the state is exclusively interested in its tax revenue but citizens care for equality. If citizens want net-income levels to be equal, the model predicts an income compression. Highly productive citizens then have to pay relatively high taxes. However, if citizens perfectly account for production costs, inequity aversion increases the differences in net-income. In this case, highly productive citizens have to pay enormous taxes stripping them off almost all rents, whereas less productive citizens are allowed to produce and earn almost nothing. Yet such a model is lacking in a number of important aspects. First, the state does not trade off its tax revenue with the well-being of its citizens. Second, no transfers from productive to less productive citizens are considered. And finally, the state might be able to enforce participation.

2.5 Conclusion

There is a common notion amongst labor economists that fairness concerns of workers cause an income or wage compression and offer an explanation for unemployment of low-skilled workers. The present paper finds that inequity aversion need not have this intuitive impact on a firm's employment decision and employment contracts. Although the present paper does not model the labor market, it is not clear whether this would change the general direction of the results. However, competing firms bring up a number of altogether different and interesting questions. For example, inequity aversion might then result in a segregation of types across industries or firms so that all good workers work in one whereas all bad workers work in another firm. Similarly, some firms might offer very 'aggressive' employment contracts with strong monetary incentives. These firms then attract only workers with little concerns for inequity. At the same time other firms might specialize on very inequity averse workers by paying them low but equal wages. Inequity aversion could be a helpful tool to investigate such labor market peculiarities.

Although inequity aversion promises a fruitful new perspective on the labor market and the organization of the firm, the preceding analysis sounds a note of caution. The interaction of inequity aversion and incentives delicately depends on the workers' point of reference. This result coincides with the findings of Bartling and von Siemens (2004b) and Bartling and von Siemens (2004a). Although the object of comparison seems to be crucial when incorporating inequity aversion into contract theory and applying it to situations with asymmetric information, there seems to be no common agreement on what individuals compare. Empirical investigations, field studies, and experiments designed to investigate that important issue are therefore a mandatory next step.

Chapter 3

Team Production*

3.1 Introduction

If the remuneration of a worker depends on the performance of a team of workers, and if misconduct by a single team member cannot be individually sanctioned, contract theory predicts that this opens the floodgates to free-riding and generates severe consequences for effort provision. Indeed, the thus called team production problem is so prevalent that the pioneering work by Alchian and Demsetz (1972) identifies the associated contractual counter-measures as the most important determinant of the nature and organization of the firm. However, firms often use team compensation schemes, and these teams seem to work well and even increase worker productivity although relatively simple contracts are employed. This contradicts the theory of incentives. We argue that other-regarding preferences - individuals care for other team members' effort choices when determining their own effort contribution - might offer an explanation for this observation. We assume that agents are inequity averse in the spirit of Fehr and Schmidt (1999), and determine optimal contracts accounting for inequity aversion. We then derive the conditions under which the free-rider problem can be overcome, and investigate the implications for the internal organization and boundary of the firm.

By now there is substantial evidence that many people do not exclusively pursue their material self-interest but exhibit some kind of other-regarding behavior.[1] Although it is still a matter of discussion in which economic environments 'fairness' motivations influence deci-

*The chapter is based on joint work with Björn Bartling from the University of Munich.

[1]For an overview of the literature see, for example, Fehr and Schmidt (2003)

sions we argue that team production is a natural candidate. If agents work within a team and receive a payment conditioning on team output, they know that their own effort choice affects the monetary payoffs of their team partners. Equally their own monetary payoff is influenced by the effort choices of the other agents. Thus, 'team spirit' might emerge. This can generate a positive or a negative impact on effort provision.[2] If all the other agents in the team work, an agent might provide high effort even if her effort costs then exceed the increase in her monetary payoff. The reason is that she feels bad for cheating the other agents by shirking. Yet, if all other agents shirk, an agent might shirk as well although working would increase her share of the joint output by more than her effort costs. She shirks in order to avoid that the other agents free-riding on her.

We investigate the impact of inequity aversion on optimal incentive provision in teams consisting of any number of identical agents. Agents face a binary effort choice, they either work or shirk. Working causes higher effort costs than shirking. Joint output is a deterministic function of the number of agents working but does not depend on the identity of the working or shirking agents. Only joint output is verifiable so that contracts can only condition on joint output and not on individual effort contributions. The potential free-riding constitutes the classical team production problem. We depart from the standard literature by assuming that agents are inequity averse in the sense of Fehr and Schmidt (1999). In our model agents compare their rent - monetary payoff minus effort costs - with the rents of the other team members. Whenever they receive a different rent they suffer a utility loss. We are interested in the consequences of inequity aversion in two separate settings: a worker-owned firm in which all proceeds from joint production are divided amongst the agents, and team production within a firm owned by a principal - an outsider incapable to influence the team's production. In the second case we assume that agents do not compare themselves with the principal.

We derive the following results. In order to explore the impact of inequity aversion we must compute optimal contracts accounting for inequity aversion. We impose the following restrictions on contracts. First, we want contracts to be renegotiation-proof. We thus only consider budget-balancing contracts. Second, we do not want agents or the principal to have

[2]The management literature abounds with case studies and empirical evidence. See Rotemberg (2002) for a survey on 'organizational behavior'.

an incentive to bribe one of the other agents to shirk. We thus require contracts to give all agents incentives to work so that no shirking agent may take up the role of a budget-breaker. When considering the principal we also limit the reduction in aggregate payoffs imposed on to the agents in case one agent shirks. Given these arguably reasonable limitations we show that one can restrict attention to contracts that are 'equal at the top'. In these contracts all agents get equal monetary payoffs in case all agents or all but one agent work.

Building upon our insights on optimal incentive provision we can study the interaction of inequity aversion with team size. We find that inequity aversion facilitates effort provision as agents suffer from 'shame from cheating' when shirking. This shame from cheating is essentially independent of the number of agents in the team. Still, the gain in rent attained by shirking increases with team size if the change in joint production weakly decreases. Hence, the positive influence of inequity aversion usually falls with the size of the team. Our model is therefore consistent with the common observation that small teams often work whereas larger ones suffer from free-riding.

We next apply our general results to worker-owned firms. First, the positive effect of inequity aversion on team incentives increases the maximum firm size allowing cooperation amongst the agents. Moreover, if firms consist of more than three agents and there exists a contract giving all agents sufficient incentives to work, the following contract can implement all agents working as the unique Nash equilibrium. First, the contract is equal at the top and gives all workers equal monetary payoffs if all agents or all but one agent work. Second, if more than one agent shirks then one agent receives the entire output in case an even number of agents works, and another agent receives the entire output in case an uneven number of agents works.

However, if worker-owned firms decide on their own size, inequity aversion might have negative consequences. If agents are inequity averse, a newly employed agent requires - and gets - more than just her effort costs. A team might thus decide not to employ an agent even though doing so causes no incentive problems and the increase in production exceeds the agent's effort costs. If the team is employed by a principal, the negative impact of inequity aversion vanishes although the positive consequences for incentive provision remain. Whenever the principal can provide all agents with sufficient incentives to work, he can extract all rents. Thus, all agents, including any new agents, get an identical rent of zero, and inequity aversion does not hinder participation.

The following papers investigate the repercussions of inequity aversion, social norms, or other-regarding preferences on team incentives. Kandel and Lazear (1992) assume that agents are influenced by peer pressure formalized by a peer pressure function. They then analyze which peer pressure functions generate the right incentives for cooperation. We, albeit, explore optimal incentive contracts given the empirically well founded social preferences developed by Fehr and Schmidt (1999). Rey Biel (2003) considers optimal contracts accounting for inequity aversion in teams of two agents and shows that contracts may use inequity aversion to facilitate incentive provision by punishing shirking agents with unfavorable inequity off the equilibrium path. Yet he assumes that effort is contractible and that contracts are not restricted by the agents' participation constraints. We explicitly account for participation of the agents, and individual effort choices are not contractible in our model. Moreover, in our model agents work as they suffer a utility loss from shame for cheating the other agents. Huck, Kübler, and Weibull (2003) analyze the effect of social norms on team incentives in a setting where individual effort choices cannot be inferred. Limiting attention to linear contracts they focus on the possible multiplicity of equilibria arising from the agents' desire to coordinate their effort choices. However, we show that if contracts are not required to be linear they can be designed to implement certain effort choices as the unique Nash equilibrium. Finally, Demougin and Fluet (2003), Englmaier and Wambach (2003), Itoh (2003) and Bartling and von Siemens (2004b) study the impact of other-regarding preferences in situations of moral hazard with stochastic production functions.

The paper is organized as follows. Section 3.2 presents the basic model. It explains the informational structure, defines inequity aversion, and introduces the restrictions imposed on contracts in order to capture the effects of renegotiation and bribery. Section 3.3 starts with a description of the optimal contracts accounting for inequity aversion. It then shows that inequity aversion has a positive impact on incentive provision. Section 3.4 applies the general results of the previous section to the case of a worker-owned firm. It characterizes the potential negative effect of inequity aversion on the optimal firm size. Moreover, it explores under which conditions there exists a contract implementing all agents working as the unique

Nash equilibrium. Section 3.5 considers teams employed by a principal. It demonstrates that in this case the potential negative consequences of inequity aversion on firm size vanish. Section 3.6 discusses some extensions, and Section 3.7 concludes.

3.2 The Model

3.2.1 Team Production, Effort and Information

Consider a team of N identical agents who can produce some output x. Let $\mathcal{N}$ denote the set of agents in a team of size N. Each agent i chooses an effort contribution $e_i \in \{0, 1\}$ to team production. Individual effort choices are not verifiable. Effort e_i causes costs $c(e_i)$ where $c(1) = c > 0$ and $c(0) = 0$. We say an agent 'works' if she chooses high effort and 'shirks' if she chooses low effort. Let $e = \langle e_i, e_{-i} \rangle$ be an effort vector consisting of agent i's effort e_i and the vector e_{-i} of all other agents' effort choices.

Joint output is a deterministic function x of the number of agents working. It does not depend on the identity of the working or shirking agents. Thus, output reveals the number of agents working but cannot tell whether a particular agent has worked or shirked. Let $x(K)$ denote joint output if K agents work. Define $\Delta x(K) = x(K) - x(K-1)$ as the marginal contribution of the K-th agent working. Output is observable, verifiable, and can be sold at a price normalized to unity. Thus, K agents working generates a revenue of $x(K)$.

The present paper investigates team incentives in two separate settings. First, it studies a worker-owned firm in which the firm is the team and proceeds from production are allocated amongst the agents. Second, it considers the role of a principal who employs a team within his firm. The principal is unproductive in the sense that he cannot influence joint output. In order to make the following definitions applicable to both settings they often include a principal. However, all of them are easily adaptable to the case where agents form a team on their own.

3.2.2 Contracts

The relationship between the agents - and potentially between the agents and a principal - is governed by a contract. A *contract* S is a function specifying how the revenue generated by the agents is distributed amongst agents and principal. A contract can use only verifiable information so that the distribution of the revenue can only condition on joint output. Output is a deterministic function of the total number of agents working. For each number K of agents working a contract S thus specifies a vector $S(K)$ consisting of individual monetary payoffs $s_i(K)$ for each agent $i \in \mathcal{N}$ and a monetary payoff $s_p(K)$ for the principal. In case the considered firm is worker-owned set $s_p(K) = 0$ for all K in the following definitions and expressions. Define $y(K) = \sum s_i(K)$ as the sum of monetary payoffs allocated to the agents and $\Delta y(K) = y(K) - y(K-1)$ as the change in this aggregate payment if K agents rather than $K-1$ agents work. Money can be 'burned' but a contract cannot distribute more than the entire output. Further, we assume strong limited liability so that all payments must be non-negative. This implies $y(K) + s_p(K) \leq x(K)$, $s_p(K) \geq 0$, and $s_i(K) \geq 0$ for all i. The following definitions are used frequently.

Definition 3.1 *A contract is called*

1. **budget-balancing at *K*** *if* $y(K) + s_p(K) = x(K)$,
2. **budget-balancing** *if it is budget-balancing at all* $K \in \{0, 1, ..., N\}$,
3. **equal at *K*** *if* $s_i(K) = s_j(K)$ *for all* $i,j \in \mathcal{N}$, *and*
4. **equal at the top** *if it is equal at* $K \in \{N-1, N\}$.

Therefore, a contract is budget-balancing if the entire output is distributed. It is called equal at K if all agents get the same monetary payoff in case K agents work. It is called equal at the top if all agents get the same monetary payoff in case all agents or all but one agent work.

3.2.3 Utility Functions

The principal is exclusively interested in his monetary payoff. However, in order to capture the positive and negative impact of 'team spirit' we assume that agents are inequity averse and suffer a utility loss if they are better or worse off than the other agents. We invoke the theory of inequity aversion developed by Fehr and Schmidt (1999).[3] Since we want that agents suffer from cheating the others when shirking, they incorporate effort costs in their comparisons.[4] Moreover, there seems to be the notion that individuals compare themselves with other individuals they perceive as 'equal' and who work or live in close proximity.[5] As a principal is almost by definition someone outside the team, we assume that agents do not compare themselves with the principal. Finally, agents are taken to be identical and preferences are common knowledge.

Formally we define inequity aversion in the following way. Within a team of N agents consider an effort vector e with K agents working with corresponding vector $S(K)$ of monetary payoffs. Define agent i's *rent* as her monetary payoff net of effort cost

$$u_i(e, S(K)) = s_i(K) - c(e_i). \tag{3.1}$$

Agents incorporate effort costs and compare rents. Drawing upon Fehr and Schmidt (1999) we define an agent's preferences as follows.

Assumption 3.1 *Within a team of N agents consider an effort vector e with K agents working with corresponding vector $S(K)$ of monetary payoffs. Then let*

$$\begin{aligned} v_i(e, S(K)) = u_i(e, S(K)) & \\ & - \alpha \frac{1}{N-1} \sum_{j=1,\, j \neq i}^{N} \max \left[u_j(e, S(K)) - u_i(e, S(K)), 0 \right] \\ & \qquad - \beta \frac{1}{N-1} \sum_{j=1,\, j \neq i}^{N} \max \left[u_i(e, S(K)) - u_j(e, S(K)), 0 \right] \end{aligned}$$

denote agent i's utility.

The parameters α and β measure the importance of inequity concerns for the agents. As Fehr and Schmidt (1999) we assume that an agent suffers a utility loss if she receives a rent different than other agents, but suffers more from inequity if it is not in her favor, $\alpha \geq \beta$ and $1 > \beta \geq 0$. We normalize the agents' utility to zero if they decide not to work for the principal.

[3] For a more general and detailed discussion of social preferences see Fehr and Schmidt (2003).

[4] Section 3.6 discusses the implications of inequity aversion if agents exclusively compare monetary payoffs.

[5] See, for example, Festinger (1954) and Williams (1975).

In order to compare rents agents must either know or have a belief about the other agents' effort choices. According to our understanding of team production agents work closely together and get a good impression of who puts in effort and who does not. Hence, we assume that agents can observe the other agents' effort decisions but that this information is not verifiable and cannot be used by the contract.

We know that this assumption is problematic. If the entire effort vector is publicly known to the agents, a court - or a principal - could devise a simple mechanism truthfully eliciting all information. By using such a mechanism any effort vector can be implemented at no informational costs and the team production problem vanishes.[6] Yet assuming that each agent's effort choice is private information causes new problems. In this case an agent's utility depends on her beliefs about the other agents' effort decisions, and psychological game theory in the line of Geaneakoplos, Pearce, and Stacchetti (1989) must be applied. Doing this is inherently difficult, rendering the derivation of optimal contracts a very complicated task. However, in our Nash equilibria each agent's belief is correct. We thus conjecture that - apart from unique implementation - the results of our model also hold if agents cannot observe the other agents' effort decisions.

3.2.4 Renegotiation and Bribery

Holmström (1982) and Eswaran and Kotwal (1984) point at the importance of renegotiation and bribery in limiting the scope of contracts in team production. Agent could initially agree on a contract that divides output evenly if all agents work, but 'burns' the entire output if at least one agent shirks. Since every agent's effort decision is thus pivotal all agents have incentives to work. Holmström (1982) argues that such a contract is not renegotiation-proof. Once it is clear that one agent has shirked the agents can agree to equally divide what ought to be burnt. All agents profit from this and renegotiation renders the initial contract not credible.

We try to capture renegotiation in the following way. Suppose K agents work. After the output has realized a contract S endows the agents - and the principal - with a legal claim

[6] Miller (1997) shows that in a team of at least three agents it is possible to implement efficient effort choices if one agent observes the effort choice of at least one other agent.

on monetary payoffs summarized by the monetary payoff vector $S(K)$.

Definition 3.2

1. *A monetary payoff vector* $S(K)$ *is* **renegotiation-proof** *if and only if there exists no* $S'(K)$ *strictly increasing the utility of at least one agent or the principal without reducing the utility of at least one agent or the principal.*

2. *A contract* S *is* **renegotiation-proof** *if and only if for all* $K \in \{0, 1, ..., N\}$ *the monetary payoff vector* $S(K)$ *is renegotiation-proof for all effort vectors* e *with* K *agents working.*

Note that the definition distinguishes between renegotiation-proof monetary payoff vectors $S(K)$ and renegotiation-proof contracts S. Definition 3.2 has the following implications.

Lemma 3.1 *Consider an effort vector* e *with* K *agents working.*

1. *If* $\beta < (N-1)/N$, $S(K)$ *is renegotiation-proof if and only if it is budget-balancing.*

2. *If* $\beta \geq (N-1)/N$, $S(K)$ *is renegotiation-proof if and only if it is budget-balancing and* $u_i(e, S) = u_j(e, S)$ *for all agents* $i, j \in \mathcal{N}$.

Proof: Budget-balance is necessary for a vector $S(K)$ of monetary payoffs to be renegotiation-proof independently of the level of inequity aversion. Suppose this was not the case, that is consider a $S(K)$ with $y(k) + s_p(K) < x(K)$ for some number K of agents working and associated output $x(K)$. Then $S'(K)$ with $s'_i(K) = s_i(K) + [x(K) - s_p(K) - y(K)]/N$ for all $i \in \mathcal{N}$ and $s'_p(K) = s_p(K)$ increases the monetary payoff for all agents by an identical, strictly positive amount while keeping the inequity between the agents unchanged. Therefore, all agents are strictly better off under $S'(K)$ whereas the principal is indifferent, and $S(K)$ is not renegotiation-proof.

Part 2: Budget-balance is also sufficient for $S(K)$ to be renegotiation-proof if agents are not highly inequity averse, $\beta < (N-1)/N$. Given a budget-balancing $S(K)$ consider any other $S'(K)$ with different monetary payoffs. Then there must either exist an agent i with $s'_i(K) < s_i(K)$, or $s'_p(K) < s_p(K)$, or both. If $\beta < (N-1)/N$, each agent's utility is strictly increasing in her monetary payoff even if the money taken away from her is given to those agents with lower utility thus decreasing inequity. The same argument holds for the principal who is not inequity averse. Therefore, at least agent i or the principal do not agree to $S'(K)$, and $S(K)$ is renegotiation-proof.

Part 3: If agents are highly inequity averse, $\beta \geq (N-1)/N$, then given an effort vector e with K agents working $S(K)$ is renegotiation-proof only if $u_i(e, S(K)) = u_j(e, S(K))$ for all $i, j \in \mathcal{N}$ and $S(K)$ is budget-balancing. Suppose $S(K)$ is budget-balancing but there exist $i, j \in \mathcal{N}$ with $u_i(e, S(K)) > u_j(e, S(K))$. Define $\mathcal{A} = \{i \in \mathcal{N} : u_i(e, S(K)) \geq u_j(e, S(K)) \, \forall \, j \in \mathcal{N}\}$ as the set of agents with the highest utility, and $\mathcal{A}^C = \mathcal{N} \setminus \mathcal{A}$ as its complement. Denote by $\#\mathcal{A}$ the cardinality of $\mathcal{A}$. Consider another $S'(K)$ with new monetary payoffs $s'_i(K) = s_i(K) - \epsilon$ for all $i \in \mathcal{A}$, whereas $s'_j(K) = s_j(K) + \epsilon \cdot (\#\mathcal{A}/\#\mathcal{A}^C)$ for all $j \in \mathcal{A}^C$. The principal's monetary payoff is not changed, $s'_p(K) = s_p(K)$. Thus, no money is burnt and $S'(K)$ is budget-balancing. Choose $\epsilon > 0$ sufficiently small so that for all $i \in \mathcal{A}, j \in \mathcal{A}^C$ we keep $u_i(e, S'(K)) \geq u_j(e, S'(K))$. We can now check whether $S'(K)$ is accepted by all agents. All agents $j \in \mathcal{A}^C$ receive higher monetary payoffs. Since for all agents $j \in \mathcal{A}^C$ payoffs increase equally, suffering from inequity with respect to all agents in $\mathcal{A}^C$ remains unchanged. However, the suffering with respect to all agents $i \in \mathcal{A}$ is reduced. Thus, all agents $j \in \mathcal{A}^C$ prefer $S'(K)$ to $S(K)$. Equally, for all agents $i \in \mathcal{A}$ utility is changed by

$$v_i(e, S'(K)) - v_i(e, S(K)) = -\epsilon + \beta \frac{1}{N-1} \sum_{j \in \mathcal{A}^C} \left[\epsilon + \epsilon \cdot \frac{\#\mathcal{A}}{\#\mathcal{A}^C} \right] = \epsilon \cdot \left[\beta \frac{N}{N-1} - 1 \right] \geq 0.$$

Thus, all agents $i \in \mathcal{A}$ weakly prefer $S'(K)$ to $S(K)$. As his payoff is unaffected the principal is indifferent, and $S(K)$ is not renegotiation-proof. Thus, $u_i(e, S(K)) = u_j(e, S(K))$ for all $i, j \in \mathcal{N}$ and budget-balance is necessary for a contract to be renegotiation-proof.

Part 4: If $\beta \geq (N-1)/N$, budget-balance and, given e with K agents working, $u_i(e, S) = u_j(e, S)$ for all $i, j \in \mathcal{N}$ is sufficient for a contract to be renegotiation-proof. Suppose this condition is satisfied. For any changes in monetary payoffs implied by another $S'(K)$ the monetary payoff of at least one person, either principal or agent, must be reduced. If the principal's payoff is reduced he clearly vetoes $S'(K)$. If only some agents' monetary payoff is reduced, denote by i the agent whose payoff is reduced by the largest amount. Then $u_i(e, S'(K)) < u_i(e, S(K))$ and $u_i(e, S'(K)) \leq u_j(e, S'(K))$ for all $j \in \mathcal{N}$. Thus, agent i's rent is reduced while in addition she now suffers from inequity with respect to some other agents. As $v_i(e, S'(K)) < v_i(e, S(K))$, agent i prefers $S(K)$ to $S'(K)$. Putting all together, $S(K)$ is renegotiation-proof. Note that all the above arguments hold with $s'_p(K) = s_p(K) = 0$ and thus in the absence of a principal. *Q.E.D.*

Suppose $S(K)$ is not budget-balancing, and consider the following new allocation. First, keep the principal's monetary payoff unchanged. Second, take the part of the output which according to the contract ought to be burned and divide it equally amongst the agents. As this increases the monetary payoff of all agents without changing their relative standing, all agents agree. Since the principal's payoff is unaffected he also agrees. Thus, a contract must be budget-balancing to be renegotiation-proof, and we restrict attention to contracts that satisfy

Condition 3.1 *Contracts must be budget-balancing.*

If agents are not highly inequity averse, $\beta < (N-1)/N$, every budget-balancing contract is renegotiation-proof. In this case agents do not agree to a reduction in their monetary payoffs even if this makes it possible to decrease inequity by increasing the monetary payoffs of agents being worse off. As the principal is exclusively interested in his monetary payoff he never agrees to a reduction in the latter. If a contract is budget-balancing, any renegotiation changing monetary payoffs must include a reduction in the monetary payoff of at least one agent or the principal. Since neither this agent nor the principal agree every budget-balancing contract is renegotiation-proof if the agents are not highly inequity averse.

If agents are highly inequity averse, $\beta \geq (N-1)/N$, a budget-balancing monetary payoff vector $S(K)$ might not be renegotiation-proof. Agents are so keen on diminishing inequity amongst themselves so that they hand over some of their monetary payoff to agents being worse off. $S(K)$ is thus renegotiation-proof if and only if it is budget-balancing and all agents - accidentally - receive the same rent irrespective of their effort choice. Still, given the same $S(K)$ but another effort vector with the same number of agents working, it is impossible that all agents still get the same rent. Thus, a contract S could only be renegotiation-proof by conditioning the vector of monetary payoffs not only on the number of agents working but on the entire effort vector e. This is unfeasible as individual effort choices are not contractible. Contrary to an individual monetary payoff vector $S(K)$, a contract S can thus never be renegotiation-proof if agents are highly inequity averse.

If agents are highly inequity averse, the ex post distribution of monetary payoffs is thus usually determined by renegotiation. Since ex ante incentives depend on this ex post allocation the result of renegotiation must be determined. We make the following assumption.

Assumption 3.2 *Consider an effort vector e with K agents working. If $\beta \geq (N-1)/N$ and $S(K)$ is budget-balancing but not renegotiation-proof, renegotiation results in the unique budget-balancing and renegotiation-proof $S'(K)$ with $s'_p(K) = s_p(K)$.*

If there is scope for renegotiation, agents transfer money amongst themselves until all receive the same rent. No money is burned in that process and the principal is not affected. It is thus implicitly assumed that the principal cannot exploit the process of renegotiation to increase his own material payoff.

As anticipation of ex post efficient renegotiation destroys ex ante incentives, Holmström (1982) points to the role of an 'outsider budget-breaker' defined as someone who cannot lower output on his own account. Given such an outsider consider the following contract. "If all agents work, output is divided evenly amongst the agents and the outsider receives no monetary payoff. If at least one agent shirks, output is not burned but given to the outsider." If the outsider does not agree to a reduction in his monetary payoff, he can veto renegotiation. This renders the initial contract credible.

If outside budget-breaking is possible, the team production problem vanishes and our paper - as all the other paper written on the same subject during the last decades - is superfluous. However, Eswaran and Kotwal (1984) argue that budget-breaking generates a new moral hazard problem. The outside budget-breaker actually prefers one of the agents to shirk. If bribery is possible, he will bribe one of the agents to reduce effort. Anticipating this the agents' ex ante incentives to work are destroyed.

In our model all those could be budget-breakers who cannot lower output on their own account. Apart from the principal - who cannot influence output by definition - this includes agents shirking in equilibrium, who cannot lower output any further as they already contribute the minimum effort. We incorporate Eswaran and Kotwal (1984) in the following way. When analyzing team incentives without a principal we restrict attention to those contracts for which all agents working forms a Nash equilibrium. We do not allow for budget-breaking agents. When including a principal we make the following restrictions. First, we only look at optimal - that is cost minimizing - contracts that give all agents incentives to work. We thus do not allow the principal to employ agents as budget-breakers. Second, we want these contracts to be *bribery-proof* in the following sense.

Definition 3.3 *Contract S is* **bribery-proof** *if and only if the principal has no incentives to induce one of the agents to shirk.*

This has the following implications. By Condition 3.1 we require contracts to be budget-balancing. As we only consider contracts giving all agents incentives to work the above definition boils to

Condition 3.2 *Given team size N, S must satisfy*

$$x(N) - y(N) \geq x(N-1) - y(N-1),$$

or $\Delta\, y(N) \leq \Delta\, x(N)$.

Given Condition 3.2 the principal has no monetary incentives to bribe one of the agents as he prefers all agents working to all but one agent working. This limits the change $\Delta\, y(N)$ and thus restricts how hard the principal can collectively punish all agents for one agent shirking. Note that Condition 3.2 is only sufficient but not necessary as all attempts of bribery might be rejected even if the principal prefers one of the agents to shirk.

However, Condition 3.2 does not imply that the principal has no incentives to bribe a group of agents. For a contract to be bribery-proof in a more general sense there should exist no coalition of agents - and possibly the principal - so that all within the coalition profit from concerted effort choices. Although theoretically more appealing we do not pursue such an approach as it would require the use of cooperative game theory.[7]

[7]However, doing this would impose additional structure on optimal contracts for those output realization accruing when more than one agent shirks. We conjecture that this might generate a tendency towards contracts that are equal for all realizations of joint output.

The section can be summarized as follows. As we want contracts to be renegotiation-proof we limit attention to budget-balancing contracts. If agents are highly inequity averse, we further assume that renegotiation results in a renegotiation-proof distribution of monetary payoffs without affecting the principal's payoff. As we want contracts to be immune to bribery we make the following restrictions. If there is no principal or the team is employed by a principal, we only consider contracts so that all agents within the team have incentives to work. In addition, we impose an upper bound on the collective punishment the principal can inflict upon the agents for one agent shirking.

3.3 Inequity Aversion and Incentives

3.3.1 Incentive Compatibility

In this section we derive optimal contracts taking inequity aversion into account. In general contracts giving all agents incentives to work are not unique. However, by the following proposition it is no restriction to concentrate on contracts that are 'equal at the top' and thus divide aggregate payment to agents equally if all agents or all but one agent work.

Proposition 3.1 *Given any budget-balancing contract S with aggregate payments $y(N)$ and $y(N-1)$, there exists a budget-balancing contract S' that is 'equal at the top', has equal aggregate payments, and weakly improves incentives to work for all agents.*

Proof: If agents are sufficiently inequity averse, $\beta \geq (N-1)/N$, the ex post distribution $S'(K)$ of monetary payoffs is determined by renegotiation according to Assumption 3.2. Incentives depend on the anticipated $S'(K)$. As, apart from $y(K)$, $S'(K)$ is independent of the initial contract S, replacing the initial contract S with any other contract S' being equal at the top and with the same $y(K)$ for $K \in \{N-1, N\}$ does not change incentives and Proposition 3.1 is trivially satisfied.

Part 2: For the remainder of the proof assume $\beta < (N-1)/N$. Thus, budget-balancing contracts are not renegotiated and directly determine incentives. First, we show that any budget-balancing contract S giving some agents unequal payoffs if all agents work, $s_i(N) \neq s_j(N)$ for some $i, j \in \mathcal{N}$, can be transformed into a budget-balancing contract S' with equal monetary payoffs, $s'_i(N) = s'_j(N)$ for all $i, j \in \mathcal{N}$ without impairing incentives.

Consider any budget-balancing contract S where $s_i(N) \neq s_j(N)$ for some $i, j \in \mathcal{N}$. Denote by $\mathcal{B} = \{i \in \mathcal{N} : s_i(N) \leq s_j(N)\, \forall j \in \mathcal{N}\}$ the set of agents with the lowest monetary payoff if all agents work. $\mathcal{B}$ is non-empty and a strict subset of $\mathcal{N}$. Define $\mathcal{C}$ as the subset of agents from $\mathcal{B}$ who have the lowest monetary payoff in case one agent shirks, $\mathcal{C} = \{i \in \mathcal{B} : s_i(N-1) \leq s_j(N-1)\, \forall\, j \in \mathcal{N}\}$. For any agent $i \in \mathcal{N}$ define $\mathcal{H}_i = \{j \in \mathcal{N} : s_j(N-1) - c > s_i(N-1)\}$ as the set of agents with a strictly higher monetary payoff net of effort costs than agent i if agent i shirks and all other agents work. Correspondingly, define $\mathcal{L}_i = \mathcal{H}_i^C = \mathcal{N} \setminus \mathcal{H}_i$. Finally, denote by $\langle e_i, e^*_{-i} \rangle$ an effort vector e where all agents apart from agent i work, and agent i chooses effort $e_i \in \{0, 1\}$.

Consider the following transformation of contract S resulting in contract S':

1. Whenever more than one agent shirks, contract S' and S are identical, $S'(K) = S(K)$ for all $K \in \{0, 1, .., N-2\}$.

2. If one or no agent shirks, monetary payoffs of all agents $i \in \mathcal{B}$ are increased, $s'_i(N-1) = s_i(N-1) + \epsilon(N-1)$ and $s'_i(N) = s_i(N) + \epsilon(N)$, where $\epsilon(N)$ and $\epsilon(N-1)$ are strictly positive constants.

3. If one or no agent shirks, monetary payoffs of all agents $j \in \mathcal{B}^C = \mathcal{N} \setminus \mathcal{B}$ are reduced, $s'_j(N-1) = s_j(N-1) - \gamma\, \epsilon(N-1)$ and $s'_j(N) = s_j(N) - \gamma\, \epsilon(N)$, where $\gamma = \#\mathcal{B}/\#\mathcal{B}^C$. Thus, what is given to the agents in $\mathcal{B}$ is taken from the agents in $\mathcal{B}^C$ so that $y'(N-1) = y(N-1)$ and $y'(N) = y(N)$, and S' is again budget-balancing.

4. $\epsilon(N)$ and $\epsilon(N-1)$ are chosen so that incentives for all agents $i \in \mathcal{C}$ to work if all other agents work remain constant. The consequence of this property is explained below.

5. $\epsilon(N)$ and $\epsilon(N-1)$ are chosen as large as possible but sufficiently small so that the rank order of the agents is preserved in the following sense. For all $i \in \mathcal{B}$, $j \in \mathcal{B}^C$, whenever $s_j(N) > s_i(N)$ then $s'_j(N) \geq s'_i(N)$. Further, if $s_j(N-1) - c > s_i(N-1)$ then $s'_j(N-1) - c \geq s'_i(N-1)$. Finally, if $s_j(N-1) > s_i(N-1) - c$ then $s'_j(N-1) \geq s'_i(N-1) - c$. Thus, whenever according to the initial contract S an agent $j \in \mathcal{B}^C$ is strictly better off than an agent $i \in \mathcal{B}$ if all agents work, if only agent i shirks or if only

agent j shirks, then according to the new contract S' she is not strictly worse off in the corresponding situation.

We will now show that incentives are not impaired in this process. Given the above transformation only the inequity between agents $i \in \mathcal{B}$ with respect to agents $j \in \mathcal{B}^C$ changes. The change in incentives for all agents $i \in \mathcal{C}$ is thus given by

$$\epsilon(N) \cdot \left[1 + (1+\gamma)\frac{\alpha\#\mathcal{B}^C}{N-1}\right] - \epsilon(N-1) \cdot \left[1 + (1+\gamma) \cdot \left(\frac{\alpha\#(\mathcal{H}_i \cap \mathcal{B}^C)}{N-1} - \frac{\beta\#(\mathcal{L}_i \cap \mathcal{B}^C)}{N-1}\right)\right].$$

As we are in the case where agents are not sufficiently inequity averse to agree to a reduction in their monetary payoff in the course of potential renegotiations, agent i's overall utility $v_i(e, S(K))$ is strictly increasing in her monetary payoff even if favorable inequity thus increases. More formally, as $\beta < (N-1)/N$ and $\gamma = \#\mathcal{B}/\mathcal{B}^C$, $\epsilon(N-1)$ is multiplied with a strictly positive factor in the above expression. By choice of the set $\mathcal{B}$, agents $i \in \mathcal{B}$ have the lowest possible rank when all agents, including themselves, are working. Thus, these agents can only improve in their rank by shirking. As some agents $j \in \mathcal{B}^C$ may then be in $\mathcal{L}_i$ (and thus not in $\mathcal{H}_i$), we must have $\#(\mathcal{H}_i \cap \mathcal{B}^C) \leq \#\mathcal{B}^C$ and $\#(\mathcal{L}_i \cap \mathcal{B}^C) \geq 0$. The marginal impact of an increase in monetary payoff depends negatively on an agent's rank: the lower the rank, the more unfavorable inequity is reduced, and the higher the marginal increase in utility. Due to the argument above, an increase in the monetary payoff if all agents work has a higher impact on utility than an increase in monetary payoff if one agent shirks. As $\epsilon(N)$ and $\epsilon(N-1)$ are chosen so that the above change in incentives is zero, we get

$$\epsilon(N) \leq \epsilon(N-1)$$

as $\epsilon(N)$ is multiplied with a larger factor than $\epsilon(N-1)$.

Consider now the incentives for any agent $i \in \mathcal{B} \setminus \mathcal{C}$ whenever this set is non-empty. Compared to any agent $j \in \mathcal{C}$, $s_i(N-1) \geq s_j(N-1)$ by definition of $\mathcal{C}$ and consequently $\#(H_i \cap \mathcal{B}^C) \leq \#(H_j \cap \mathcal{B}^C)$ and $\#(L_i \cap \mathcal{B}^C) \geq \#(L_j \cap \mathcal{B}^C)$. Thus, agents $i \in \mathcal{B} \setminus \mathcal{C}$ will in general improve their rank by more when shirking than agents $j \in \mathcal{C}$. Since the marginal impact of the increase $\epsilon(N-1)$ in monetary payoff if one agent shirks is lower, the incentive to work hard if all other agents work hard is at least preserved for any agent $i \in \mathcal{B} \setminus \mathcal{C}$ as it is at least as large as for any agent $j \in \mathcal{C}$.

Finally, consider any agent $i \in \mathcal{B}^C$, whose change in incentives is given by

$$-\epsilon(N)\left[\gamma-(1+\gamma)\frac{\beta\#\mathcal{B}}{N-1}\right]+\epsilon(N-1)\left[\gamma+(1+\gamma)\left(\frac{\alpha\#(\mathcal{H}_i\cap\mathcal{B})}{N-1}-\frac{\beta\#(\mathcal{L}_i\cap\mathcal{B})}{N-1}\right)\right].$$

Again, the second factor of the above expression and is strictly positive as $\gamma = \#\mathcal{B}/\#\mathcal{B}^C$ and $\beta < (N-1)/N$. Since $\#(\mathcal{H}_i \cap \mathcal{B}) \geq 0$ and $0 \leq \#(\mathcal{L}_i \cap \mathcal{B}) \leq \#\mathcal{B}$, the above expression is at least weakly positive as $\epsilon(N-1) \geq \epsilon(N)$, and all agents $i \in \mathcal{B}^C$ keep their incentives to work hard. Summarizing, the above transformation of the contract S does not harm incentives. Iterated application eventually yields a contract S with $s_i(N) = s_j(N)$ for all $i, j \in \mathcal{N}$. Iterated application of this transformation eventually results in a contract S' with $s'_i(N) = s'_j(N)\,\forall\, i, j \in \mathcal{N}$.

Part 3: However, after the above transformations S is not yet necessarily equal at the top as there might exist $s_i(N-1) \neq s_j(N-1)$ for at least some $i, j \in \mathcal{N}$. In this case define $\mathcal{D} = \{i \in \mathcal{N} : s_i(N-1) \geq s_j(N-1) \forall j \in \mathcal{N}\}$ as the set of agents with the highest monetary payoff if one agent shirks. $\mathcal{D}$ is non-empty and a strict subset of $\mathcal{N}$. As the contract is equal if all agents work, $s_i(N) = s_j(N)$ for all $i, j \in \mathcal{N}$, all agents get the same utility if all agents work. As agents $i \in \mathcal{D}$ get the highest monetary payoff if one agent shirks, and as their utility is increasing in their monetary payoff as $\beta < (N-1)/N$, these agents have the minimum incentive to work if all other agents work.

Consider the following transformation of contract S resulting in contract S'.

1. Whenever more than one agent shirks or when all agents work, the contract is unchanged, $S'(K) = S(K)$ for all $K \neq N-1$.

2. If only one agent shirks, monetary payoffs of all agents $i \in \mathcal{D}$ are reduced, $s'_i(N-1) = s_i(N-1) - \epsilon'(N-1)$, where $\epsilon'(N-1)$ is a strictly positive constant.

3. If only one agent shirks, monetary payoffs of all agents $j \in \mathcal{D}^C = \mathcal{N} \setminus \mathcal{D}$ are increased, $s'_j(N-1) = s_j(N-1) + \gamma'\epsilon'(N-1)$, where $\gamma' = \#\mathcal{D}/\#\mathcal{D}^C$. Thus, what is given to agents in $\mathcal{D}$ is taken from the agents in $\mathcal{D}^C$ so that $y'(N-1) = y(N-1)$, and S' is again budget-balancing.

4. $\epsilon'(N-1)$ is chosen as large as possible but sufficiently small so that the rank order of the agents is preserved in the following sense. For all $i \in \mathcal{D}$ and $j \in \mathcal{D}^C$, $s'_i(N-1) \geq s'_j(N-1)$.

As $\beta \leq (N-1)/N$, an agent's overall utility is increasing in her monetary payoff. Thus, for each agent $i \in \mathcal{D}$ incentives to work increase, whereas for each agent $j \in \mathcal{D}^C$ incentives to work decrease. Since the transformation preserves the rank order, $v_i(\langle 0, e^*_{-i}\rangle, S'(N-1)) \geq v_j(\langle 0, e^*_{-i}\rangle, S'(N-1))$ for all $i \in \mathcal{D}$ and $j \in \mathcal{D}^C$. Thus, minimum incentives over all agents increase.

Iterated application of the above transformation eventually results in a contract that is equal at the top, and all agents have an identical incentive to exert effort given all other agents work. Moreover, minimum incentives over all agents are at least weakly increased. Note that none of the above transformations affect the principal's monetary payoff and can thus be performed without lowering his profit. *Q.E.D.*

If agents are highly inequity averse, $\beta \geq (N-1)/N$, most monetary payoff vectors implied by the initial contract are renegotiated. Incentives are thus determined by renegotiation as characterized in Assumption 3.2. Apart from the aggregate monetary payoffs $y(K)$ and $y(K-1)$ not changed by renegotiation, initial contracts are irrelevant and Proposition 3.1 is trivially satisfied.

In the more interesting case agents are not highly inequity averse, $\beta < (N-1)/N$, and initial contracts determine incentives. For an illustration of Proposition 3.1 consider a contract that is not equal at the top. By definition there then exists an agent, say agent i, who gets the lowest monetary payoff if all agents work. Since all agents incur the same effort costs if all agents work, agent i then holds the lowest rank - the lowest relative position - with respect to her rent. Consider the following changes in the contract. Agent i's monetary payoffs if all *and* if all but one agent work are increased. These changes satisfy the following properties. First, what is given to agent i is taken from the others so that the monetary payoff vector remains budget-balancing. Second, agent i's incentives are held constant.

Agents suffer more from being worse off than from being better off than others. Therefore, the lower the rank of an agent the higher the utility gain from increasing her monetary payoff. By choice of agent i her rank cannot be lower if she is the only agent shirking as compared to the situation in which everybody works - in the latter case she already holds the lowest possible rank. To keep her incentives unchanged her monetary payoff need never be increased by a larger amount if all agents work than if only one agent shirks. This has the following implication for the incentives of the other agents. Due to budget-balance the monetary payoff of all other agents decreases weakly more if one agent shirks than if all agents work. As in the considered case agents are not highly inequity averse and hence enjoy having more monetary payoff, their incentives to work are never harmed but potentially improved. Thus, the proposed change renders the contract more equal without harming incentives or altering aggregate payments to agents. Iterated application of the above procedure finally results in a contract that is equal at the top.

Note that Proposition 3.1 does not imply that contracts must be equal at the top in order to give all agents incentives to work. Indeed, if agents are sufficiently inequity averse, unequal contracts may provide sufficient incentives even if contracts are not renegotiated. However, a contract that is equal at the top maximizes the minimum incentives of all agents. Thus, the less inequity averse the agents, the less 'unequal' monetary payoffs may be in case all agents or all but one agent work.

The impact of inequity aversion on team incentives can now be easily derived. Optimal contracts depend on the level of inequity aversion. If agents are highly inequity averse, there is the following result.

Proposition 3.2 *Suppose $\beta \geq (N-1)/N$ and consider a team of size N. If and only if*

$$\Delta\, y(N) \geq c,$$

all N agents working forms a Nash equilibrium.

Proof: Suppose agents are highly inequity averse, $\beta \geq (N-1)/N$. Assumption 3.2 implies that given any budget-balancing contract S and for any effort vector e with K agents working, the agents will renegotiate $S(K)$ so that in the end $S'(K)$ satisfies $u_i(e, S'(K)) = u_j(e, S'(K))\, \forall\, i, j \in \mathcal{N}$. The aggregate payment $y(K)$ to the agents remains unchanged. Aggregation over all agents and budget-balance imply effort dependent monetary payoffs of

$$s_i'(K) = c(e_i) + \frac{y(K) - Kc}{N}$$

for each agent i. This is anticipated by all agents. After substitution of $s_i'(K)$ into the utility function, each agent has incentives to maximize aggregate monetary payoffs to agents minus their sum of costs. *Q.E.D.*

If agents are highly inequity averse, $\beta \geq (N-1)/N$, renegotiation ensures equal rents for any initial contract. Anticipating this each agent knows that she will be compensated for the incurred effort cost and in addition receive a share of the generated surplus distributed to the agents. If N agents work, the surplus is the agents aggregate monetary payoff minus the sum of their effort costs, $y(N) - N\,c$. Thus, each agent has incentives to exert effort if and only if her effort costs are smaller than the resulting increase in aggregate payment. Note that the change in aggregate payments $\Delta y(K)$ is not altered by renegotiation but fixed by the contract. In this respect the contract determines incentives.

If agents are not highly inequity averse, $\beta < (N-1)/N$, budget-balancing contracts are not renegotiated and directly determine incentives. By Proposition 3.1 we can restrict attention to contracts that are equal at the top. Therefore, it is possible to derive the precise conditions under which all agents working can form a Nash equilibrium, and

Proposition 3.3 *Suppose $\beta < (N-1)/N$ and consider a team of size N. If and only if*

$$\Delta\, y(N) \geq (1-\beta)\, N\, c,$$

all N agents working forms a Nash equilibrium.

Proof: Suppose $\beta \geq (Nc-1)/(Nc)$. In this case consider a budget-balancing contract that is equal at the top, $s_i(N) = s_j(N) = 1$ and $s_i(N-1) = s_j(N-1) = (N-1)/N$ for all $i, j \in \mathcal{N}$. If all agents work, all receive the same monetary payoff while incurring the same effort costs, their utility is $v_i(\langle 1, e^*_{-i}\rangle, S) = 1 - c$ for all $i \in \mathcal{N}$. Suppose only agent i shirks, whereas all other agents $j \neq i$ work. In this case $u_j(\langle 0, e^*_{-i}\rangle, S) = (N-1)/N - c$ for all $j \neq i$, and $u_i(\langle 0, e^*_{-i}\rangle, S) = (N-1)/N$. Thus, agent i's incentive to exert effort if all other agents work is given by

$$v_i(\langle 1, e^*_i\rangle, S) - v_i(\langle 0, e^*_i\rangle, S) = 1 - c - \left[\frac{N-1}{N} - \beta c\right],$$

which is weakly positive iff $\beta \geq (Nc-1)/(Nc)$. Thus, all agents working forms a Nash equilibrium given a budget-balancing contract that is equal at the top.

Part 2: Suppose $\beta < (Nc-1)/(Nc)$, and there exists a contract S so that all agents working forms a Nash equilibrium. By Proposition 3.1, all agents working must form a Nash equilibrium given a budget-balancing contract S' that is equal at the top. However, this cannot be as $\beta < (Nc-1)/(Nc)$. Therefore, $\beta \geq (Nc-1)/(Nc)$ is sufficient and necessary for the existence of a budget-balancing contract according to which all agents working forms a Nash equilibrium. *Q.E.D.*

Consider a contract S that is equal at the top. If an agent shirks whereas all other agents work, her monetary payoff is reduced by her share $\Delta y(N)/N$ in the reduction of the aggregate payment to all agents but she saves c on effort costs. As the agent is inequity averse she suffers βc from cheating the other agents. Thus an agent has no incentive to shirk if and only if $\Delta y(N)/N \geq (1-\beta)\,c$ which is equivalent to the condition in Proposition 3.3.

Note that Proposition 3.1 also holds for the case that agents are not inequity averse, $\beta = 0$. Proposition 3.3 then implies that in a team of size N all agents working forms a Nash equilibrium if and only if

$$\Delta\, y(N) \geq N\, c. \tag{3.2}$$

Inequity aversion hence has an unambiguously positive effect on incentive provision: the aggregate punishment needed to induce all agents to work is smaller if agents are inequity averse.

3.4 The Worker-Owned Firm

3.4.1 The Impact of Inequity Aversion on Incentives

In this section we investigate the team production problem in the context of a worker owned firm. As there is no principal we set $s_p(K) = 0$ for all $K \in \{0, 1, .., N\}$, and the workers share the entire revenue amongst themselves, $y(K) = x(K)$.[8] The aggregate punishment inflicted upon the agents if one agent shirks is thus determined by the production technology, $\Delta y(K) = \Delta x(K)$. The results of the previous section imply that cooperation arises more

[8] The same situation arises if firms compete for workers so that no firm (or in the sense of the present model, no principal) can make positive profits.

easily if agents are inequity averse. We say *all agents working is implementable* if there exists a contract satisfying Condition 3.1 so that all agents working forms a Nash equilibrium.

Proposition 3.4 *Consider a worker-owned firm of size N.*

1. *If $\Delta\, x(N) \leq c$, all agents working is not implementable.*

2. *If $\Delta x(N) > c$ and $\Delta\, x(N) < Nc$, all agents working is implementable if and only if $\beta \geq 1 - \Delta\, x(N)/(Nc)$.*

3. *If $\Delta x(N) > c$ and $\Delta\, x(N) \geq Nc$, all agents working is implementable for all $\beta \geq 0$.*

Proposition 3.4 follows directly from Proposition 3.2 and 3.3 with $\Delta\, y(K) = \Delta\, x(k)$. It has the following implications. First, if all agents working is not efficient, it is not implementable. Thus, inequity aversion can never support inefficiently large firms. Second, inequity aversion facilitates cooperation. If all agents working is implementable in case agents are not inequity averse, it is implementable if agents are inequity averse. Moreover, there exist situations in which all agents working is implementable if and only if agents are inequity averse. Thirdly, the minimum level of inequity aversion required to sustain cooperation increases with the size N of the firm if $\Delta\, x(N)/N$ then decreases. The present paper therefore can explain why in reality small partnerships often work well whereas larger ones frequently suffer from free-riding.

The reason for exerting effort depends on the agents' degree of inequity aversion. If agents are highly inequity averse, $\beta \geq (N-1)/N$, they anticipate renegotiation and thus have an incentive to maximize joint surplus. If agents are not highly inequity averse, $\beta < (N-1)/N$, they know that there will be no renegotiation. They are thus not interested in the joint surplus. Yet if an inequity averse agent shirks whereas all other agents work, she incurs a utility loss from being better off than all other agents. If this 'shame for cheating' outweighs the - potential - increase in her rent, the agent abstains from shirking. Putting it differently, inequity averse agents overcome the team production problem if 'compassion' or 'shame for cheating' is large enough. It is this behavioral trait of 'feeling bad' when cheating the others that creates incentives to exert effort. In contrast a selfish agent does not bear these behavioral costs, which makes cooperation more difficult to sustain.

3.4.2 Inequity Aversion and the Optimal Firm Size

Contrary to the positive impact on incentives, inequity aversion can prevent employment of an additional agent even if this is efficient and causes no incentive problems absent inequity aversion. In the following numerical example consider a firm consisting of two agents, agent 1 and 2. If both agents exert effort, they each incur effort costs $c = 1$ but produce joint output $x(2) = 10$. Suppose both divide revenue evenly, and that both agents working forms a Nash equilibrium. Thus, both get a rent of 4 if firm size is 2. However, the firm has the opportunity to expand and employ an agent 3. If all three agents work, output increases to $x(3) = 13$. Agent 3 has effort costs $c = 1$. As the increase of 3 in joint output exceeds the effort costs of 1 it is efficient to expand and employ the agent.

Employing the agent causes no incentive problems if there is no inequity aversion. Suppose no agent is inequity averse and agent 3 is employed. Consider the following contract. If all agents work, agent 1 and 2 receive a monetary payoff of 6 whereas the newly employed agent 3 receives a monetary payoff of 1. If only two agents work, agent 1 and 2 receive a monetary payoff of 5 whereas agent 3 receives no monetary payoff. This contract is budget-balancing. Further, if all agents work, agent 1 and 2 get a rent of 5 whereas agent 3 is just compensated for her effort costs. If either agent 1 or 2 shirks, the shirking agent receives a rent of 5. If agent 3 shirks, she receives a rent of zero. Thus, all agents working forms a Nash equilibrium. Finally, agent 1 and 2 get a rent of 5 which exceeds the rent of 4 if agent 3 is not employed. The firm expands.

However, the incumbent owners of the firm might decide not to expand if they are inequity averse. Suppose all agents are highly inequity averse, $\beta \geq 2/3$. All agents then know that they will renegotiate and divide joint surplus evenly. All agents working forms a Nash equilibrium and all agents, including agent 3, get the same rent of 10/3. As this is less than the rent of 4 which agent 1 and 2 get on their own, agent 3 is not employed.

But even if the agents are not highly inequity averse, the firm might decide not to expand. As she then suffers from rent inequality, agent 3 does not accept work if she receives a monetary payoff of 1 in case all agents work. This holds for all positive levels of inequity aversion. But if agent 3 has to be given more than 1, agent 1 and 2 must receive less than 6 if all agents work. Depending on their degree of inequity aversion all agents working might no longer form a Nash equilibrium although inequity aversion facilitates cooperation. Summarizing, if agents are inequity averse but no highly inequity averse, the above problem - although less severe - might still exist.

If firms are predominantly viewed as a place of team production, the present paper might provide fresh insights into the labor market. Suppose there exists only a limited number of production opportunities and thus firms within an economy. Cooperation amongst the owners of a firm becomes more difficult with increasing size of the team. Proposition 3.4 implies that either the conditions on β for a given production technology increase, or equivalently the conditions on the production technology become more stringent for a given β. For either interpretation, including another individual causes incentive problems so that individuals might not find a job although joint surplus increases if they are employed.

Inequity aversion has an ambiguous impact on the situation. First, incentive provision is facilitated thus reducing the negative impact of incomplete information on firm size and employment. Second, inequity aversion causes the incumbent owners of a firm additional costs. Depending on the degree of inequity aversion either a new-employed partner has to be given an income exceeding her effort costs or renegotiation directly results in a new distribution of the joint surplus. Inequity aversion might thus aggravate the situation on the labor market. Note that the interaction of incentive problems, inequity aversion and team size is determined by contracts. Thus, the above conclusions can only be drawn by referring to our results in the previous section.

3.4.3 Multiple Equilibria

If agents are highly inequity averse, $\beta \geq (N-1)/N$, initial contracts are renegotiated so that all agents receive the same rent. As this is anticipated each agent has incentives to locally maximize joint surplus. Thus, K agents working forms a Nash equilibrium if and only if $x(K-1)-(K-1)\,c \leq x(K)-Kc$ and $x(K)-Kc \leq x(K+1)-(K+1)\,c$. Whether there exist multiple Nash equilibria depends on the production technology. If there are several equilibria, some of them are undesirable because they do not maximize joint surplus. Contrary to the multiple equilibria in Huck, Kübler, and Weibull (2003) agents do not 'coordinate' on these equilibria with low aggregate effort in order to synchronize their actions as such. Instead agents might end up in these undesirable Nash equilibria as their effort choice is optimal given the effort choices of the other agents.

However, if agents are not highly inequity averse, $\beta < (N-1)/N$, contracts determine incentives and can thus be designed to destroy multiple Nash equilibria. We say a contract *uniquely implements all agents working* if all agents working forms the unique Nash equilibrium. For any degree of inequity aversion there is the following result for a particulary simple production technology.

Proposition 3.5 *Consider a worker-owned firm of size* N *with* $x(K) = K$ *for all* $K \in \{0, 1, ..., N\}$*, and suppose agents have effort costs* $c < 1$*. If* $N \geq 4$*, there exists a contract* S *that uniquely implements all agents working if and only if* $\beta > (Nc-1)/(Nc)$*.*

Proof: If $\beta \geq (N-1)/N$, all agents working forms the unique Nash equilibrium as $K = N$ maximizes $K - K\,c$ subject to $K \in \{0, 1, ..., N\}$.

If $\beta < (Nc-1)/(Nc)$, there exists no contract that implements all agents working by Proposition 3.3. Thus, there cannot exist a contract that uniquely implements all agents working.

For the remainder of the proof assume $(Nc-1)/(Nc) < \beta < (N-1)/N$. Consider the following contract S. First, S is budget-balancing and equal at the top, $s_i(N) = s_j(N) = 1$ and $s_i(N-1) = s_j(N-1) = (N-1)/N$ for all $i, j \in \mathcal{N}$. Second, $s_1(x) = 1$ and $s_i(x) = 0$ for all $i \neq 1$ if x is even, and $s_2(x) = 1$ and $s_i(x) = 0$ for all $i \neq 2$ if x is uneven. In words, if more than one agent shirks, the entire output is given to agent 1 if the number of working agents is even or zero, and to agent 2 otherwise.

Since the contract is equal at the top all agents working forms a Nash equilibrium according to Proposition 3.3. Below we show that this Nash equilibrium is unique.

First, suppose all but one agent work and output is $x = N - 1$. Given $\beta = (Nc-1)/(Nc)$

all agents working forms a Nash equilibrium where all agents are exactly indifferent between working and shirking if all other agents work. In this case all but one agent working might also form a Nash equilibrium. However, if $\beta > (Nc-1)/(Nc)$ each agent strictly prefers to work if all the other agents work. As we consider the latter case, all but one agent working cannot form a Nash equilibrium.

Next, suppose at least two but not more than $N-2$ agents work so that output is $x \in \{2,..,N-2\}$. In this case there are at least two agents exerting effort, only one agent receiving the entire output, and $N-x$ agents shirking. Call the agent who works but does not receive any monetary reward agent i. Her utility is given by

$$v_i(\langle 1, e_{-i}\rangle, S) = -c - \alpha\left[\frac{1}{N-1}\,x + \frac{N-x}{N-1}\,c\right].$$

If agent i shirks, there are $x-1$ agents left working. Agent i's utility then depends on $S(x-1)$. There are three cases. First, suppose agent i receives the entire output when shirking, $s_i(x-1) = x-1$. Apart form agent i, there are now $N-x$ agents shirking and $x-1$ agents working. None of them receives a monetary payoff. Thus, agent i's utility is given by

$$v_i(\langle 0, e_{-i}\rangle, S) = (x-1) - \beta\left[\frac{N-x}{N-1}(x-1) + \frac{x-1}{N-1}(x-1+c)\right].$$

Second, suppose agent $j \in \{1,2\}$ other than agent i receives the output $x-1$. Suppose agent j works. In addition to agent j, there are now $x-2$ agents working. Agent i, as all the other agents apart form agent j, receives no monetary payoff. Thus, her utility is given by

$$v_i(\langle 0, e_{-i}\rangle, S) = -\alpha\left[\frac{1}{N-1}(x-1-c)\right] - \beta\left[\frac{x-2}{N-1}\,c\right].$$

Thirdly, suppose agent $j \in \{1,2\}$ other than agent i receives the output $x-1$, and agent j shirks. There are now $x-1$ agents working. Agent i, as all the other agents apart form agent j, receives no monetary payoff. Thus, her utility is given by

$$v_i(\langle 0, e_{-i}\rangle, S) = -\alpha\left[\frac{1}{N-1}(x-1)\right] - \beta\left[\frac{x-1}{N-1}\,c\right].$$

Comparison of the above three terms yields that agent i receives the minimum utility when shirking if the now reduced output $x-1$ is given to another agent who herself is shirking. However, even in this worst case agent i's utility when shirking exceeds her utility when working as long as

$$c\left[1 - \beta\frac{x-1}{N-1} + \alpha\frac{N-x}{N-1}\right] + \alpha\frac{1}{N-1} > 0.$$

This is satisfied as $N > x$ and $\beta < 1$. Thus, $x \in \{2, .., N-2\}$ agents working cannot form a Nash equilibrium.

Next, suppose only one agent works so that output is $x = 1$. As $N \geq 4$ there are less than $N - 1$ agents working. Thus, the entire output is given to agent 2. There are two cases. First, suppose agent $i \neq 2$ works. Her utility is then given by

$$v_i = -c - \alpha \left[\frac{N-1}{N-1} c - \frac{1}{N-1} \right] < 0.$$

If agent i shirks, all agents including agent i herself receive a utility of zero. Thus, agent i prefers to shirk. Second, suppose agent 2 works. Consider agent 1, who then receives a utility of

$$v_1 = -\alpha \left[\frac{1}{N-1}(1-c) \right] < 0.$$

Suppose agent 1 works. As $N \geq 4$, there are less than $N - 1$ agents working and agent 1 receives a monetary payoff of 2. Thus, agent 1's utility is given by

$$v_1 = 2 - c - \beta \left[\frac{N-2}{N-1}(2-c) + \frac{1}{N-1} 2 \right],$$

which is larger than zero as $\beta < (N-1)/(N)$. Thus, agent 1 prefers to work. Consequently, one agent working cannot form a Nash equilibrium.
Finally, suppose no agent works. Then agent 2, as all the other agents, receives a utility of $v_2 = 0$. Suppose agent 2 works. As $N \geq 4$, she then receives the entire output, $x = 1$. Thus, her utility when working is given by

$$v_2 = 1 - c - \beta \left[\frac{N-1}{N-1}(1-c) \right] > 0$$

as $\beta < 1$. Thus, no agent working cannot form a Nash equilibrium. Summarizing, all agents working is the unique Nash equilibrium given contract S. *Q.E.D.*

As $\Delta\, x(K) = 1 < c$ for all K it is always efficient if all agents work, and all agents working is the unique maximizer of the joint surplus for any firm size N. If agents are highly inequity averse, $\beta \geq (N-1)/N$, all agents working thus forms the unique Nash equilibrium by Proposition 3.2.

Suppose now that agents are not highly inequity averse, $\beta < (N-1)/N$. Consider the following contract. "In case all or all but one agent work, output is equally distributed. Otherwise, if an even number of agents or nobody works, the entire output is given to, say, agent 1. If an uneven number of agents works then all is given to another agent, say, agent 2." By Proposition 3.3 all agents working forms a Nash equilibrium. As the condition on β is given by the strict inequality this Nash equilibrium is strict and all but one agent working cannot form a Nash equilibrium. Consider now all other candidate equilibria in turn. All agents shirking cannot be an equilibrium. In this case agent 2 has an incentive to work as she then receives the entire output. Only one agent working cannot be an equilibrium either. If in this case agent 2 is working, now agent 1 has an incentive to work as she then receives the entire output. If in this case agent 2 is not working, the working agent incurs effort costs but does not receive a monetary payoff. Hence, she gets a higher utility by shirking and deviates. In all other cases at least one agent works but does not receive a monetary payoff. This agent is better off shirking.

Unfortunately, the above contract may not work for firm sizes smaller than four. If agents are not highly inequity averse, a contract implementing all agents working must give all agents relatively equal monetary payoffs in case all agents or all but one agent work. If firm size is less than four, this imposes restrictions on the vector of monetary payoffs in case two or one agent work. Consider the following example with $N = 2$, $x(K) = K$, and effort costs strictly larger but very close to 1/2. If the contract is equal at the top, both agents receive a monetary payoff of 1/2 if one agent works and the other agent shirks. In this case the working agent receives a rent of $1/2 - c$. As this is smaller than zero no agent working forms a Nash equilibrium. In order to uniquely implement all agents working the contract has to be made 'less equal at the top'. This has two effects. One the one hand, minimum incentives to work are impaired so that at least one agent has reduced incentives to work. On the other hand, abandoning the contractual restriction of equality at the top might make unique implementation possible. For the above example consider the following contract with $s_1(0) = s_2(0) = 0$, $s_1(1) = 2/3$, $s_2(1) = 1/3$ and $s_1(2) = s_2(2) = 1$. It is then possible to show that for effort costs strictly larger but very close to 1/2 there exist levels of inequity aversion $\alpha < (2-3c)/(3c-1)$ and $\beta > (3c-1)/(3c+1)$ with $\alpha > \beta$ so that both agents working forms the unique Nash equilibrium. The condition on β is more restrictive than the condition imposed in Proposition 3.3, but the minimum β is smaller than $(N-1)/N = 1/2$. Thus, it is sometimes possible to uniquely implement all agents working although agents are not sufficiently inequity averse to prompt renegotiation. If inequity aversion is sufficiently high, the additional restriction imposed by unique implementation might be no problem. But the less inequity averse the agents, the more detrimental are the restrictions imposed on the contract at the top. Therefore, for some parameter constellations all agents working can never be uniquely implemented in firms comprising less than four agents although it is implementable.

3.5 Teams and Outside Ownership

Teams are often used in firms with outside ownership. This section analyzes the impact of inequity aversion on a firm's decision to form teams. Including a principal does not change the general implications of inequity aversion on incentive provision. The following proposition summarizes the conditions allowing a principal to implement all agents working in a team of size N. Note that in addition to Condition 3.1 contracts are now required to satisfy Condition 3.2.

Proposition 3.6 *Consider a team of size N.*

1. *If $\beta \geq (N-1)/N$, the principal can implement all agents working if and only if $\Delta\, x(N) \geq c$.*

2. *If $\beta < (N-1)/N$, the principal can implement all agents working if and only if $\Delta\, x(N) \geq (1-\beta)\, N\, c$.*

The costs of implementing all agents working are $y(N) = N\, c$.

Consider an optimal contract implementing all agents working in a team of given size N. First, Proposition 3.1 implies that the contract can be taken to be equal at the top, $s_i(N) = s_j(N)$ and $s_i(N-1) = s_j(N-1)$ for all $i, j \in \mathcal{N}$.

Second, all agents are just compensated for their effort costs in case all agents work. Thus, $s_i(N) = c$ for all $i \in \mathcal{N}$ and $y(N) = N\, c$. As in equilibrium all agents work they just accept the contract. The principal cannot lower these minimum monetary payoffs without forfeiting participation. Moreover, individual incentives are determined by the change $s_i(N) - s_i(N-1)$ in monetary payoffs and not by the absolute magnitude of $s_i(N)$. Thus increasing $s_i(N)$ is not necessary to improve incentives as long as $s_i(N-1)$ may be lowered.

Thirdly, since the contract is equal at the top Propositions 3.2 and 3.3 may be applied. Thus, incentive provision imposes a lower bound on $\Delta\, y(N)$. For given $y(N) = N\, c$, this implies an upper bound on $y(N-1)$. Finally, Condition 3.2 imposes an upper bound on $\Delta\, y(N)$. For given $y(N) = N\, c$, this implies a lower bound on $y(N-1)$. Whenever the last two conditions on $\Delta\, y(N)$ or equivalently $y(N-1)$ are compatible, the principal can give the right incentives for all agents to work while at the same time keeping the contract bribery-proof and thus credible. This implies Proposition 3.6.

Inequity aversion thus simplifies incentive provision for the principal. For a given production technology x, the optimal team - or firm - size is determined by

$$\max_{N \in \Gamma} x(N) - N\, c,$$

where Γ is the set of team sizes N so that the relevant condition on $\Delta\, x(N)$ in Proposition 3.6 is satisfied. The principal can extract all rents from the agents and maximizes joint surplus. As all the employed agents get zero rent an additional agent accepts work if she is just compensated for her effort costs. Thus, the negative impact of inequity aversion on firm size vanishes with the introduction of a principal.

The analysis of the interaction between inequity aversion and team incentives might provide interesting insights into the organization of firms. Splitting the workforce in different divisions might have an impact on the workers' reference groups. This can generate positive or negative incentive effects. Finally, note that different organizational forms can almost always be observed within a single firm. Consultancies and law firms have on the one hand a strong horizontal team orientation but on the other hand relatively strict - informal - vertical hierarchies. One could argue that this might allow the co-existence of teams and individual workers outside of teams.

3.6 Discussion

3.6.1 Agents Comparing only Monetary Payoffs

The theory of inequity aversion by Fehr and Schmidt (1999) explains the seemingly contradictory empirical evidence of a large number of experiments. In these cases agents' reference groups and criteria of comparison are essentially defined by the experimental setting. In more natural environments these points are no longer clear. When applying the theory of inequity aversion to team production it seems plausible that the other agents form the group of reference. We also find it compelling that people take the other team members' effort decisions into account, and this is indeed mandatory to capture the notion of 'shame for cheating'. We thus have assumed throughout the paper that agents compare rents amongst each other.

Notwithstanding suppose that agents compare monetary payoffs only. If the contract is equal in case all agents or all but one agent work, an agent does not suffer from inequity if he alone shirks. Incentives are then not affected by inequity aversion. Could they be improved by a contract that gives one or several agents higher and the other agents lower monetary payoffs in case one agent shirks? This surely enhances incentives for those agents getting a lower payoff, but how about those agents getting more? If these agents are not very inequity averse and their utility is rising in their monetary payoff, their incentives to work are reduced. If agents are homogeneous, minimum incentives over all agents are thus impaired. However, if the agents are very inequity averse and their utility is decreasing in their monetary payoff, there will be renegotiation until all agents share joint production equally. Anticipating this any initial contract boils down to an ex post equal contract, and agents have the same incentives to shirk as if there was no inequity aversion. Thus, inequity aversion cannot improve incentives in teams if agents compare monetary payoffs only.

3.6.2 Agents with Heterogeneous Preferences

In this model all agents are taken to have identical preferences, yet in reality agents probably differ with respect to their concerns for inequity. This would probably have the following consequences. If preferences are observable, the more inequity averse agents can be granted a relatively large fraction of the output in case one agent shirks. As these agents suffer a lot from cheating they do not shirk. The other, less inequity averse agents can then be assigned a low fraction of output in case one agent shirks. This prevents them from shirking even if they do not suffer heavily from social comparisons. Thus, incentive provision may be facilitated with the help of unequal contracts.

If the agents' preferences are unobservable, differentiating between agents with different concerns for inequity is not feasible. This generates a tendency towards more equal contracts. However, it might now be possible and desirable to screen agents according to their preferences. This opens a whole range of new and interesting questions.

3.7 Conclusion

This paper shows how incentive provision in team production is affected if agents are inequity averse in the sense of Fehr and Schmidt (1999). Optimal contracts accounting for inequity aversion involve simple, budget-balancing, and equal sharing rules. Moreover, inequity aversion can provide all agents with sufficient incentives to work in cases where this is unfeasible with agents only caring for their own monetary payoff and effort costs. Our results are driven by the assumption that agents incorporate effort costs in their comparisons. 'Shame from cheating' reduces a shirking agent's utility precisely in those cases in which she actually shirks. As contracts cannot condition on an agent's effort decision, they cannot afflict the above 'intrinsic punishment' with equal precision. Thus, inequity aversion facilitates incentive provision. Furthermore, the conditions under which inequity aversion permits cooperation amongst the agents depend on the size of the team. They usually become less restrictive with decreasing size of the team, which fits the common observation that small teams often work well whereas larger ones suffer from free-riding. Summarizing, the present paper shows that inequity aversion and the associated incentive effects could offer a fruitful new perspective on the internal organization of firms if these can influence its workers' references groups.

Bibliography

Akerlof, G. A. (1982): "Labour Contracts as a Partial Gift Exchange," *Quarterly Journal of Economics*, 97(4), 543–569.

Akerlof, G. A., and J. L. Yellen (1988): "Fairness and Unemployment," *American Economic Review*, 78(2), 44–49.

——— (1990): "The Fair Wage-Effort Hypothesis and Unemployment," *Quarterly Journal of Economics*, 105(2), 255–283.

Alchian, A. A., and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62(5), 777–795.

Bartling, B., and F. A. von Siemens (2004a): "Efficiency in Team Production with Inequity Averse Agents," mimeo, University of Munich.

——— (2004b): "Inequity Aversion and Moral Hazard with Multiple Agents," mimeo, University of Munich.

Bewley, T. F. (1995): "A Depressed Labour Market as Explained by Participants," *American Economic Review*, 85(2), 250–254.

——— (1999): *Why Wages Don't Fall during a Recession*. Harvard University Press.

Blinder, A. S., and D. H. Choi (1990): "A Shred of Evidence on Theories of Wage Stickiness," *Quarterly Journal of Economics*, 105(4), 1003–1015.

Camerer, C., and R. Thaler (1997): "Ultimatums, Dictators, and Manners," *Journal of Economic Perspectives*, 64, 209–219.

Campbell, C. M., and K. S. Kamlani (1997): "The Reasons for Wage Rigidity: Evidence from a Survey of Firms," *Quarterly Journal of Economics*, 112, 759–789.

Carmichael, L., and W. B. MacLeod (2002): "How should a Behavioral Economist do Welfare Economics," mimeo, Queen's Univerity.

——— (2003): "Caring about Sunk Costs: A Behavioural Solution to Hold-Up Problems with Small Stakes," *Journal of Law, Economics and Organization*, 19(1).

Crawford, V. P. (1982): "A Theory of Disagreement in Bargaining," *Econometrica*, 50(3), 607–638.

Davis, D., and C. Holt (1993): *Experimental Economics*. Princeton University Press, Priceton, NJ.

Demougin, D., and C. Fluet (2003): "Group vs. Individual Performance Pay When Workers Are Envious," *CIRPÉE Working Paper 03-18*.

Ellingsen, T., and M. Johannesson (2002a): "Is there a Hold-Up Problem?," mimeo, Stockholm School of Economics.

——— (2002b): "Sunk Costs, Fairness, and Disagreement," mimeo, Stockholm School of Economics.

Englmaier, F., and A. Wambach (2003): "Contracts and Inequity Aversion," mimeo, University of Munich.

Eswaran, M., and A. Kotwal (1984): "The Moral Hazard of Budget-Breaking," *The RAND Journal of Economics*, 15(4), 578–581.

Fehr, E., and S. Gächter (2000): "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90, 980–994.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993): "Does Painess Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics*, 108, 437–460.

Fehr, E., A. Klein, and K. M. Schmidt (2001): "Fairness, Incentives and Contractual Incompleteness," *Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 72*.

Fehr, E., S. Kremhelmer, and K. M. Schmidt (2003): "Fairness and the Optimal Allocation of Ownership Rights," mimeo, University of Munich.

Fehr, E., and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114(3), 817–868.

——— (2000): "Fairness, Incentives and Contractual Choices," *European Economic Review*, 44(4-6), 1057–1068.

——— (2003): "Theories of Fairness and Reciprocity - Evidence and Economic Applications," in *Advances in Economics and Econometrics - 8th World Congress*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, pp. 208–57. Cambridge: Cambridge University Press.

Festinger, L. (1954): "A Theory of Social Comparison Processes," *Human Relations*, 7, 117–140.

Frank, R. H. (1984a): "Are Workers Paid their Marginal Products?," *American Economic Review*, 74(4), 549–571.

——— (1984b): "Interdependent Preferences and the Competitive Wage Structure," *Rand Journal of Economics*, 15(4).

Fudenberg, D., and J. Tirole (1983): "Sequential Bargaining with Incomplete Information," *Review of Economic Studies*, 50(2), 221–247.

——— (1991): *Game Theory*. MIT Press.

Geaneakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, 60–79.

Grossman, S., and O. Hart (1986): "The Costs and Benefits of Ownership: a Theory of Lateral and Vertical Integration," *Journal of Political Economy*, 94, 691–719.

Grout, P. (1982): "Investments and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," *Econometrica*, 52, 449–460.

Gul, F. (2001): "Unobservable Investment and the Hold-Up Problem," *Econometrica*, 69(2), 343–376.

Güth, W., R. Schmittberger, and R. Tietz (1990): "Ultimatum Bargaining Behaviour - A Survey and Comparison of Experimental Results," *Journal of Economic Psychology*, 11, 417–449.

HART, O., AND J. MOORE (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119–1158.

HOLMSTRÖM, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340.

HUCK, S., D. KÜBLER, AND J. WEIBULL (2003): "Social Norms and Economic Incentives in Firms," *The Research Institute of Industrial Economics, Working Paper No. 565.*

ITOH, H. (2003): "Moral Hazard and Others-Regarding Preferences," mimeo, Graduate School of Commerce and Management, Hitotsubashi University.

KAGEL, J., AND A. ROTH (1995): *Handbook of Experimental Economics.* Princeton University Press, Princeton NJ.

KANDEL, E., AND E. P. LAZEAR (1992): "Peer Pressure and Partnerships," *Journal of Political Economy*, 100(4), 801–817.

LAFFONT, J.-J., AND J. TIROLE (1993): *A Theory of Regulation and Procurement.* Cambridge, MIT Press.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory.* Oxford University Press.

MILLER, N. H. (1997): "Efficiency in Partnerships with Joint Monitoring," *Journal of Economic Theory*, 77, 285–299.

REY BIEL, P. (2003): "Inequity Aversion and Team Incentives," mimeo, University College London.

ROTEMBERG, J. J. (2002): "Altruism, Reciprociy and Cooperation in the Workplace," mimeo, Harvard Business School.

SMITH, V. L. (1982): "Microeconomic Systems as an Experimental Science," *American Economic Review*, 72, 923–955.

TIROLE, J. (1999): "Incomplete Contracts: Where Do We Stand?," *Econometrica*, 67(4), 741–781.

WILLIAMS, R. M. (1975): "Relative Deprivation," in *The Ideal of Social Structure: Papers in Honor of Robert K. Merton*, ed. by L. Coser. New York: Harcourt, Brace, Jovanovich.

Lebenslauf

1. Juni 1973	Geboren in Istanbul, Türkei
8. Juli 1992	Abitur am Wilhelmsgymnasium, München
1992-1993	Wehrpflicht
1993-1994	Studium der Volkswirtschaftslehre an der Albert-Ludwigs-Universität Freiburg
1995-1999	Studium der Volkswirtschaftslehre an der Ludwig-Maximilians-Universität München
29. November 1999	Diplom in Volkswirtschaftslehre
2000-2001	Wissenschaftlicher Mitarbeiter am Seminar für Dynamische Modellierung, LMU München
2001-2002	Marie Curie Fellow an der Faculty of Economics and Politics der Universität Cambridge
2002-2004	Wissenschaftlicher Mitarbeiter am Seminar für Dynamische Modellierung, LMU München
21. Juli 2004	Promotion zum Dr. oec. publ.