Essays in Behavioural Contract Theory

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy
by

Harvey Upton

March, 2019

supervised by
Prof. Dominique Demougin and Dr. Christian Bach

The University of Liverpool; Management School

Abstract

This thesis investigates the optimal provision of incentives when employment relationships are characterised by moral hazard and individuals have relative income concerns. In contrast to the existing literature, it is assumed that workers' social comparisons are not limited to others within the firm, but extend to larger groups in society.

The thesis is organised into three chapters. The first chapter provides a comprehensive survey of the literatures which study incentive contracting when parties' preferences are characterised by the related concepts of inequity aversion or loss aversion. The chapter discusses the similarities between the two literatures, highlights some results which are relevant for both preference specifications and establishes the context in which the subsequent analysis should be placed.

The second chapter examines a firm's optimal choice of contract for a worker whose preferences exhibit relative income concerns. This is formalised through use of a stylised model, in which workers have an aversion to falling behind the economy's average income. In this framework, it is shown that the optimal contract takes either a binary or ternary form and that firms benefit from the social comparisons of workers. In addition, there is an interdependence between the contracting of firm-worker pairs which results in an externality effect, so that firms could gain from collective decision making. Moreover, relative income concerns imply a lower economy-wide average wage, as well as a reduced level of inequality as measured by the Gini coefficient.

The third chapter extends the foregoing analysis to an environment featuring a frictional labour market and unemployment; this allows for an investigation into how dismissal can be used by firms to create effort incentives. Dismissing workers for poor performance is shown to act as a substitute for explicit incentive pay, allowing firms to reduce wage costs. Some implications for labour market policies are derived. Increases in the minimum wage are found to aid the creation of incentives, lowering the bonus payment necessary to implement effort. In contrast, increases in unemployment benefits have a negative impact on incentives. These effects are shown to be stronger and more pronounced when workers have relative income concerns.

Overall, the thesis provides several predictions and insights aimed at improving the understanding of incentive contracts, their structure and their effects on individual behaviour.

Contents

Acknowledgements

In preparing this thesis, I received support from many people to whom I am grateful. First and foremost, I wish to thank Dominique Demougin. His guidance throughout the Ph.D. process has been invaluable, both for the composition of this thesis and for my development as an economist. In particular, our joint work together has significantly advanced my understanding of how one should undertake research. The second chapter of this thesis was cowritten with him and contains an earlier version of our paper, "Moral Hazard and Keeping up with the Joneses". I would also like to thank Christian Bach. His support and guidance, especially in the early stages of the Ph.D. process, has been greatly appreciated. Financial assistance has come from the University of Liverpool Management School and the Economic and Social Research Council (ESRC), to whom I am grateful.

I am also thankful to numerous members of the University of Liverpool microeconomic theory group, as well as the wider economics department. This thesis has benefited considerably from their feedback. In particular, Olga Gorelkina, Erin Hengel and Robert Routledge offered valuable comments on the second chapter of this thesis. In addition, I wish to thank Cahal Moran for helpful feedback on an early version of the third chapter. Valued comments have also come from audiences at the NWDTC Economics Conference in Manchester, 2017; the Liverpool Economic Theory & Computation Workshop, 2017; the Conference of the French Association of Law and Economics in Aix-en-Provence, 2017 and the NWDTC Economics Conference in Liverpool, 2018.

Rob Edwards and Sanna Laksá have read many parts of this thesis and provided valuable advice. They have also been a constant source of support and guidance throughout my studies, for which I am especially grateful. I am also thankful to Kaori Narita, Sirui Wu and Yunchou Wu. Finally, I would like to thank my family and friends for their support over the last few years, without which, this thesis would not have been possible.

Introduction

Behavioural contract theory studies how modifications to preferences — specifically, deviations from the classical approach supported by psychological evidence — influence contracting between economic actors.[1] One particular insight of this line of research is that individuals are not purely self-interested, as is commonly assumed, but typically care about how their payoffs relate to those of others around them. For instance, Fehr and Schmidt (1999) document substantial empirical and experimental evidence for individuals having concerns over the fairness of outcomes. In experimental games, some subjects have been shown to consistently reject outcomes characterised by large inequalities, even if they personally lose out as a result of this. Similarly, many individuals have shown a willingness to forego income in order to impose costly punishments on others whose behaviour does not comply with social norms. In recent years, a number of studies have investigated how the optimal provision of incentives within firms is affected by these preferences for fair and equitable outcomes (e.g. Bartling and von Siemens, 2010*b*). A hallmark of this literature is that the inequity concerns of individuals are limited to other workers within the firm.

While there is evidence that such intra-firm wage comparisons can be important (see e.g. Card et al., 2012), there are many studies which suggest that workers may also compare their incomes to others outside of the firm. One particular candidate for income comparisons is "people like me"; i.e. those with similar characteristics such as age, education, location and occupation. For instance, Clark and Oswald (1996) use British data to construct a reference wage for each individual, corresponding to the predicted earnings of a worker with the same characteristics, and show that there is a negative relationship between this reference wage and job satisfaction. Along similar lines, Cappelli and Sherer (1988) find that an individual's pay satisfaction is negatively correlated with the average

[1] Throughout, the word 'classical' is used to refer to models which make traditional assumptions about the nature of individual preferences: namely, that they are self-focused and that utility is determined solely by outcomes themselves, rather than in relation to any beliefs, expectations or reference points. See the summary articles of Rabin (1998) and DellaVigna (2009) for evidence of psychological phenomena that motivates studying deviations from the classical model and Kőszegi (2014) for a recent survey of the behavioural contract theory literature.

wage paid to comparable workers by other firms in the same industry. Taking a geographical approach to reference group formation, Luttmer (2005) provides evidence from US data that, controlling for several characteristics including income, an individual's happiness is negatively correlated with the incomes of others in the local area. Moreover, an increase in the earnings of neighbours and a decrease in one's own income are each associated with a similar sized reduction in happiness. Comparable findings have been reported from studies of households in Canada (Helliwell and Huang, 2010), 18 Latin American countries (Graham and Felton, 2006) and rural China (Knight et al., 2009). The lattermost study is particularly interesting since relative income is shown to have a strong impact on happiness even in regions with high levels of poverty.

In this thesis, we study the optimal provision of incentives when workers' reference groups extend beyond the firm, to larger groups in society. Specifically, we assume that individuals have an aversion to falling behind the average wage in an economy, or a desire to *keep up with the Joneses*.[2] We begin by providing a comprehensive survey of the existing relevant literature. Next, assuming that employment relationships are characterised by moral hazard, we investigate a firm's optimal choice of contract when they have access to a rich performance measure and examine the impact of workers' relative income concerns. We also extend our analysis to allow for unemployment and discuss some implications for labour market policies. The thesis therefore provides a number of predictions and insights which aim to improve our understanding of the structure of incentive contracts and their effects on individual behaviour.

The first chapter of the thesis provides an extensive review of the literatures which study incentive contracting when parties' preferences are characterised by either inequity aversion or loss aversion. The purpose of this survey is threefold. First, we provide an in-depth analysis of the similarities between the two specifications of preferences and their implications for incentive contracting. Intuitively, since in both cases individuals are assumed to make comparisons — either to reference points or reference groups — and have an aversion to falling behind, the utility associated with outcomes which result in relatively low wage payments is reduced in comparison to the standard case. We show that, as a result, there are comparable implications for optimal incentive contracts. Presentation of the two literatures together therefore allows for further insights as well as an analysis of some results which are relevant for both preference specifications.

Second, since our formalisation of relative income concerns also shares similarities with these models of preferences, the literature review establishes the context

[2]The terminology 'keeping up with the Joneses' to describe preferences which include relative income concerns has previously been applied to numerous economic contexts; see for instance Dupor and Liu (2003).

in which the analysis contained in the subsequent chapters of the thesis should be placed. Third, we additionally identify some promising avenues for future research.

This first chapter begins by initially considering the standard model of incentive contracting in the presence of moral hazard and presenting the key results of this literature. Next, we outline the basic models of inequity averse and loss averse preferences, while additionally discussing some issues regarding their application to incentive contracting. We then survey the relevant literature and show that aversion to either inequity or losses typically has a positive impact on incentives, while reducing the expected utility of workers. The key issue in the inequity aversion literature has been the extent to which a worker's wages should depend on the performance of others in the reference group. In contrast, many papers which consider loss aversion have instead focussed on the optimal contractual structure in the presence of a rich performance measure. We also study the impact of these preference models for tournament schemes and team production.

The second chapter of the thesis investigates the optimal design of incentive contracts when workers have relative income concerns and employment relationships are characterised by moral hazard. Specifically, we introduce a stylised model where workers have an aversion to falling behind a reference wage, which, in equilibrium, is assumed to be determined by the average income in a replica economy populated by a continuum of firm-worker pairs.

The first part of the chapter examines the contracting problem of a single pair who take the reference wage as given. Solving for the wage scheme which minimises the firm's costs of implementing a given effort level, the optimal contract is shown to be simple, taking either a binary or ternary form. In addition, we find that the firm benefits from the worker's relative income concerns, since aversion to falling behind the reference wage has a positive impact on incentives.

We next embed this model of contracting into an economy where all firms and workers are assumed to be identical. In order to further the notion of relative income concerns, we let the reference wage be determined endogenously by the average income of workers. In this environment, we show that the optimal contract is ternary and establish some additional comparative static results, before using a series of numerical examples to examine how changes in the parameters of the model affect the economy's equilibrium. We discuss an externality effect, whereby each firm ignores the implications of their contracting decision for the economy's average payment, which affects the employment relationships of others via the reference wage. Finally, we find that higher relative income concerns are associated with a lower average wage, as well as a reduced level of inequality as measured by the Gini Coefficient.

In the third chapter of the thesis, we extend this model to allow for unemployment. One possible interpretation of our result that the optimal contract can take a ternary form is that the wage scheme is binary, with the additional option of dismissal for poor performance. Such a contract is then perceived by the worker as ternary, since there are three distinct possible outcomes: dismissal and a base wage; retention and a base wage; retention, a base wage and a bonus payment. However, it is clear that dismissal can only create additional incentives if workers anticipate a reduction in expected utility following termination of the employment relationship. This is the case if, for instance, workers are unemployed for a long period of time and lose a significant amount of income. With this in mind, we consider a dynamic environment featuring labour market frictions, as formalised using the standard model of job matching (Pissarides, 2000).

Assuming that, as before, workers have relative income concerns and are averse to falling behind the average wage in the economy, we first examine the contracting problem of a single firm-worker pair. We show that when deciding on the probability of dismissal, the firm takes into account two distinct effects. First, by dismissing the worker for poor performance, the firm can create incentives to undertake effort, allowing for a reduction of the bonus payment and lower expected wage costs; this increases current period profits. Second, the possibility of dismissal means there is a positive probability that the employment relationship is terminated at the end of each period. Due to the frictional nature of the labour market, in this case the firm can spend several periods searching for a new match, resulting in a loss of profits. The firm then trades off these two effects when deciding the frequency with which the worker is dismissed.

Clearly, the extent to which the firm can use dismissal as a device to create incentives will depend on the underlying conditions of the labour market. Accordingly, when evaluating labour market policies which impact the payoffs of both employed and unemployed workers, it is important to additionally take into account their effects on employment relationships and the ability of firms to provide effort incentives. To explore this issue, we use a series of numerical examples to examine the impact of changes in the minimum wage and unemployment benefits on the probability of dismissal, the optimal contract and the steady-state equilibrium labour market outcomes. This is in line with Kőszegi's (2014) recommendation that, when studying contract-theoretic models with non-standard preferences, analysis should extend beyond the effects of the behavioural parameters, to variables which are fundamentally of more economic interest.

We show that, in our framework, the minimum wage acts as an efficiency wage which increases an employed worker's utility relative to unemployment, creating larger incentives to undertake costly effort. It follows that dismissal is more effective following a raise in the minimum wage and is therefore used more often. This reduces the size of the necessary bonus payment, although the firm's overall wage costs increase. Additionally, we show that the steady-state unemployment level increases, since existing matches are terminated with higher probability and lower profits push some firms out of the market.

In contrast, an increase in unemployment benefits lowers the relative utility of an employed worker, reducing effort incentives. The threat of dismissal therefore becomes less effective and is used less often. Accordingly, firms must increase the size of the required bonus payment, leading to higher wage costs and lower profits. We also find that a rise in unemployment benefits induces opposing effects on unemployment: while existing matches last for longer on average, the decrease in profits implies that some firms will exit the market.

Moreover, while the foregoing effects exist for the case where workers are self-interested, we show that they are stronger and more pronounced when workers have relative income concerns. Our analysis therefore highlights the importance of considering the implications for incentives in employment relationships when evaluating labour market policies.

In the conclusion to the thesis, we briefly summarise the key results of our analysis and explore how they relate to the existing literature. We discuss some of the assumptions made throughout, in particular with respect to our formalisation of relative income concerns and comment on some possible alternative modelling choices. Finally, we outline some of the ways in which our analysis could be extended and offer some thoughts on the future directions of the literature.

Chapter 1

1 Introduction

The classical approach to economic theory assumes a highly simplified model of human behaviour. Individuals are assumed to be hyperrational: their preferences are self-focused, time-consistent and utility is determined solely by outcomes themselves, rather than in relation to any beliefs, expectations or reference points. While it has been long known that this approach neglects important aspects of human behaviour, its simplicity has allowed for tractable models which have enabled economists to make invaluable predictions, insights and real-world interventions regarding economic phenomena. Nonetheless, in many — if not all — areas of economics, allowing for more accurate descriptions of how individuals behave can lead to significant developments in our comprehension of important economic issues. With this in mind, the behavioural economics literature has both examined the ways in which economic agents consistently violate the assumptions of the classical approach and, following this, developed theoretical models which can capture this behaviour.[1] These models have then been applied to a variety of different topics, resulting in further insights and predictions.

One such topic is the the provision of incentives in the presence of moral hazard. In many environments, incentive contracts will be designed by one party specifically in order to optimally influence the behaviour of another. Accordingly, the structure of these contracts is typically sensitive to variations in the underlying preferences of parties. Due to the ubiquitous nature of contracts which — to at least some extent — are designed in order to provide incentives, it is important to understand both their composition and their impact on the behaviour of the economic actors involved. There is by now a substantial body of work which attempts to explore these issues by assuming that preferences are non-standard. The purpose of this chapter is to provide a survey of a particular subset of this literature. Specifically, we shall discuss incentive contracting when preferences

[1] This literature is vast and as such is not covered in detail here; Rabin (1998) provides a survey while DellaVigna (2009) considers evidence from the field.

deviate from the classical model in two particular ways.

The first deviation relates to individuals who care about the payoffs of others. There is a wealth of evidence that economic agents are not purely self-interested, as is assumed by the standard theory; instead, their preferences have a social aspect and as such their behaviour will vary as the payoffs of others change. In particular, individuals often demonstrate a desire for *fair* or *equitable* outcomes. Such preferences are captured by Fehr and Schmidt's (1999) model of inequity aversion, which also retains a high degree of tractability and is therefore well-suited to economic applications.

The second deviation we consider is reference-dependent preferences and loss aversion. As first documented in the economics literature by Kahneman and Tversky (1979), individuals have a tendency to evaluate outcomes not in isolation, but relative to some subjectively determined reference point. Moreover, deviations below the reference point (*losses*) typically lead to variations in utility which are greater in magnitude than equal sized deviations above the reference point (*gains*). This is the key feature of Kahneman and Tversky's (1979) prospect theory, which has been successfully applied to several economic topics and can help explain many results which are inconsistent with the standard model (see Barberis, 2013). There have also been a number of subsequent formalisations of reference-dependence preferences, which attempt to sharpen the foregoing insights while remaining widely applicable.

This chapter provides a review of how inequity aversion and loss aversion have been applied to the study of incentive contracting. Throughout, we shall follow the literature and predominantly consider the provision of incentives within organisations. Besides providing a natural context in which to discuss the findings of the literature, labour contracts will typically involve parties whose preferences are likely to coincide with the aforementioned deviations from the classical approach.[2]

First, since organisations are social in nature, coworkers represent a natural reference group with whom employees are likely to make wage comparisons.[3] Indeed, there exists evidence both that job satisfaction depends on such relative pay comparisons (Card et al., 2012) and that firms view the need to maintain equality as an important constraint on their internal wage structures (Agell and Lundborg, 1995). Second, a central theme of the literature on contracting in the presence of moral hazard has been the trade-off between incentive provision and the income-risk which this entails. Since a worker's wage is likely to constitute a significant portion of his earnings, it is particularly important to account for

[2]It should be emphasised, however, that many of the literature's findings will extend to other environments which feature moral hazard.

[3]This point is further discussed in Section 3.1.3. Moreover, contracts which condition on the performance of coworkers are likely to further induce such comparisons, by increasing *social proximity* (Bartling, 2012*a*).

risk preferences when studying the optimal form of labour contracts. As noted by Barberis (2013), prospect theory is widely considered to be the most accurate description of decision making under risk currently available, so that its application seems particularly relevant for models of incentive contracting in the workplace.

The chapter is structured as follows. First, Section 2 introduces the standard model of incentive contracting in the presence of moral hazard. We present the basic setup and discuss possible modelling choices, before presenting the key insights of the classical literature.[4] Throughout, we utilise a series of formal examples in order to illustrate these results, which then serve as a benchmark for later comparisons.

Next, in Section 3, we outline the basic models of inequity aversion (3.1) and reference-dependent preferences (3.2), while exploring some details regarding their application to economic environments and in particular to incentive contracting. For the former model, it is important to specify the relevant reference group with whom individuals undertake comparisons as well as the exact nature of these comparisons; that is, how do we define an individual's 'payoff'? For models of reference-dependence, the most important issue when considering applications relates to the correct definition of an individual's reference point. We conclude the section (3.3) by comparing the two theories, discussing some important similarities and arguing that in many environments they can lead to similar predictions.

Section 4 then provides a survey of the relevant literature. We first consider the case in which a firm employs workers who compare their wages with one another and dislike inequalities (4.1). The firm then faces a trade-off, since wage inequality can be an effective way to create incentives, but also reduces the expected utility of workers so that compensation may need to be increased in order to guarantee that the contract is accepted. The key issue in this literature is the extent to which the remuneration of one worker should depend on the performance of others. Through the design of wage schemes, firms are able to either exacerbate or eliminate wage inequality such as to minimise their costs; as we shall see, aversion to inequity then has a significant impact on the composition of the optimal contract. We also explore alternative specifications, such as the case where workers compare their wages to their superiors, rather than coworkers.

Next, we discuss incentive contracting when workers are averse to losses (4.2) and find that such preferences induce a similar trade-off to the aforementioned case of inequity aversion — the firm must then decide the extent to which they wish to expose workers to potential losses. The main focus of this literature has been the structure of the optimal wage scheme when the firm has access to a rich

[4] We do not attempt to provide a comprehensive review of the moral hazard literature when parties have classical preferences. For that purpose, see Gibbons (1998), Prendergast (1999), Bolton and Dewatripont (2005) or Macho-Stadler and Pérez-Castrillo (2018).

performance measure, with reference-dependence and loss aversion implying key differences to the standard case. The remainder of the section then considers the implications of inequity and loss aversion for two common alternative organisational forms: tournament schemes (4.3) and team production (4.4). Finally, Section 5 concludes the chapter by discussing the key results of these literatures and offers some thoughts on future directions.

While the majority of the existing literature is covered, this survey is not intended to be exhaustive and as such some papers have been excluded. Throughout, we limit attention to studies which are primarily theoretical in nature, rather than considering empirical or experimental contributions. Moreover, in both our review of the literature and our formal examples, we typically omit technical discussions regarding issues such as the existence or uniqueness of equilibria, instead focussing on the literature's important results and their associated intuitions.

There exist a number of related works which also survey the literature on incentive contracting when parties have non-standard preferences. Englmaier (2005) reviews the early literature on incentive contracting with social preferences, but does not cover reference-dependence or loss aversion. Kőszegi (2014) surveys the broader behavioural contract theory literature, which addresses additional deviations from the standard model of preferences (e.g. time inconsistency) as well as alternative forms of informational asymmetries (e.g. adverse selection). However, the impressive scope of his paper precludes a detailed analysis of any one specific area, so that much of the material presented here is not covered. Similarly, Macho-Stadler and Pérez-Castrillo (2018) consider the application of non-standard preferences to models of moral hazard, but only briefly discuss a selection of key results.

2 Moral Hazard and Incentive Contracting

2.1 The Basic Framework

The standard model of incentive contracting under moral hazard features a worker (*the agent*) who is hired to exert costly effort $a \in \mathcal{A}$ on behalf of a firm (*the principal*) in return for a wage payment w. The agent's preferences over wage payments and effort can be represented by a separable utility function, $U(w, a) = u(w) - c(a)$. The first term $u(w)$, which captures the agent's preferences over money and risk, is assumed to be an increasing and (weakly) concave function. As we shall see, the nature of the agent's risk preferences has important implications for the outcome of contracting under moral hazard. The second term, $c(a)$, is an increasing and strictly convex function representing the agent's costs of undertaking effort.[5] The agent is under no obligation to participate in the relationship and has access to an outside opportunity which offers utility $\overline{U}$, while we assume that the principal always wishes to participate. Effort creates output for the principal, the value of which is denoted by the increasing concave function $v(a)$. The principal is assumed to be risk neutral, so that her payoff function is given by $\Pi(w,a) = v(a) - w$.

We first consider the full information benchmark, in which a is observable and verifiable for all parties. In order to ensure participation, the agent must be compensated for both his effort costs and outside utility. The principal's problem is therefore:

$$\max_{a,w} \; v(a) - w \tag{1}$$

$$\text{s.t.} \;\; u(w) - c(a) \geq \overline{U} \tag{2}$$

Since the wage w enters negatively into the principal's objective function (1), the *participation constraint* (2) will bind. Accordingly, $u(w) = c(a) + \overline{U}$. Substituting this into (1), the principal then solves for the optimal effort level to be induced:

$$\max_{a} \; v(a) - u^{-1}\left(c(a) + \overline{U}\right) \tag{3}$$

The argument which maximises (3) is often referred to as the *first-best* effort level (i.e. the outcome in the absence of any informational asymmetries) and denoted by a^{FB}. For instance, if the agent is risk neutral so that $u(w) = w$ and $\mathcal{A} = \mathbb{R}_+$, the solution is implicitly defined by $v'(a) = c'(a)$. Since the agent's action is observable, the principal can then implement the desired effort level a^{FB} using a number of wage schemes which condition on a.[6] One simple example is as follows:

$$w^{FB}(a) = \begin{cases} u^{-1}\left(c(a^{FB}) + \overline{U}\right) & a \geq a^{FB} \\ \\ -\infty & a < a^{FB} \end{cases} \tag{4}$$

[5] For technical reasons we further assume that $c(0) = 0$, with $\lim_{a \to 0} c'(a) = 0$.

[6] Note that in our framework, we have assumed that the agent's effort has a deterministic impact on the output accrued to the principal. A common alternative approach is to model effort as influencing the probability distribution over different levels of output, which then acts as a signal of effort provision (e.g. Holmström, 1979). Technically, the principal's output function is then $v(x) = x$, rather than $v(a)$. Nonetheless, since the principal still wishes to design the contract as to minimise her costs of implementing effort, the majority of key results can be derived in either framework. In the full information case, when output is stochastic, the optimal wage scheme must allocate risk efficiently between parties; a risk averse agent is then fully insured by the wage scheme and receives a fixed payment, regardless of the output level.

Moral hazard obtains when effort is either unobservable or unverifiable, in which case effort incentives must be provided by conditioning the agent's wage on some performance measure. Formally, we assume that there exists a number of signals $x \in \mathcal{X}$, which are observable and verifiable for all parties. The agent's effort then influences the probability distribution over these signals, so that varying payment depending on the signal observed will impact the agent's effort choice. The moral hazard literature then aims to analyse the outcome of contracting between the parties, with a particular emphasis on the nature of the resulting wage scheme.

It is standard to assume that the principal has all of the bargaining power in the relationship. The timing of the game is as follows. First, the principal offers a take-it-or-leave-it contract to the agent, who either accepts or rejects. In the case of rejection, each party receives their respective outside utilities. If the contract is accepted, the agent chooses an effort level a to undertake. Next, the uncertainty in the environment is resolved and both parties observe a single signal $x \in \mathcal{X}$. Finally, the principal pays the wage w to the agent as specified by the contract.

The principal's aim is to design the wage scheme which implements the profit maximising effort level at the lowest possible cost, while simultaneously guaranteeing that the agent participates in the contract. Formally, the principal's problem becomes:

$$\max_{a,\, w(x)} \ v(a) - \mathbb{E}\left[w(x)|a\right] \tag{5}$$

$$a \in \arg\max_{\hat{a} \in \mathcal{A}} \ \mathbb{E}\left[u(w(x))|\hat{a}\right] - c(\hat{a}) \tag{6}$$

$$\mathbb{E}\left[u(w(x))|a\right] - c(a) \geq \overline{U} \tag{7}$$

Since effort is unverifiable, the agent is free to choose the value of a which he finds most beneficial. Accordingly, the *incentive compatibility* constraint (6) requires that the principal's desired effort level maximises the agent's expected utility among all other possible effort levels in $\mathcal{A}$. The *participation* constraint (7) requires that the agent's expected utility is sufficiently large to induce him to accept the contract. Finally, some models feature a *limited liability* constraint (8), which imposes a lower bound (often zero) on the available wages which can be paid to the agent.

$$w(x) \geq 0, \ \ \forall x \in \mathcal{X} \tag{8}$$

The main source of variation in models of moral hazard is the structure of the sets $\mathcal{A}$ and $\mathcal{X}$. We briefly consider each of these modelling choices in turn. The set $\mathcal{A}$ is often assumed to take one of two possible forms:

1. *Discrete*; $\mathcal{A} = \{a_1, ..., a_n\}$ for some $n \in \mathbb{N}$.

2. *Continuous*; $\mathcal{A} = [\underline{a}, \overline{a}] \subseteq \mathbb{R}_+$.

The discrete approach typically simplifies analysis since the incentive compatibility constraint reduces to a finite number of inequalities which must be satisfied. In fact, it is common to assume that $\mathcal{A} = \{a_L, a_H\}$ with $a_L < a_H$, so that effort is binary and there is only one inequality to be satisfied. In such cases, the principal's problem becomes designing the wage scheme which implements effort a_H at the lowest cost.[7] In contrast, assuming that $\mathcal{A}$ is continuous involves taking the first-order approach, whereby (6) is replaced by its first-order condition. Letting $\mathcal{A}$ be continuous can be advantageous, since it allows for an analysis of how the principal's optimal effort choice changes with the parameters of the model. However, ensuring concavity of the agent's problem often requires placing restrictive conditions on the performance measure.

Next, we outline three common approaches to modelling the performance measure:

1. *Discrete*; $\mathcal{X} = \{x_1, ..., x_n\}$ for some $n \in \mathbb{N}$. The probability of observing a particular x_i is then given by $p(x_i; a)$, with $\sum_{i=1}^{n} p(x_i; a) = 1, \ \forall a \in \mathcal{A}$.

2. *Continuous*;

   (a) $\mathcal{X} = [\underline{x}, \overline{x}] \subseteq \mathbb{R}$. In this case, the random variable $x \in \mathcal{X}$ is distributed according to the function $F(x; a)$ with associated density $f(x; a)$.

   (b) $\mathcal{X} = (-\infty, +\infty)$, with $x = a + \epsilon$, where the random variable $\epsilon \sim \left(0, \sigma^2\right)$ according to the distribution function $G(x; a)$.

The simplest possible approach is to assume a binary performance measure, so that $\mathcal{X} = \{x_L, x_H\}$. This allows for an examination of the fundamentals of the moral hazard problem, while significantly simplifying analysis. Indeed, many of the important findings of the literature can be illustrated within this basic framework. However, since the number of possible payments a wage scheme can specify is limited by the cardinality of $\mathcal{X}$, the contract offered by the principal in this case will necessarily be binary. This precludes an investigation into the optimal *contractual form* in a given environment. For this purpose, many authors assume that the performance measure is either discrete with an arbitrary (but finite) number of possible realisations, or continuous with outcomes being distributed over an interval. However, while these assumptions allow for a derivation

[7] Note that since a_L minimises the agent's effort costs here, it can be implemented by paying a constant wage.

of many interesting properties of the optimal wage scheme, they tend to require strong assumptions regarding the distribution of outcomes and typically one is still unable to solve for general explicit solutions.

The final approach we consider therefore assumes a continuous measure of performance but imposes further structure by specifying that $x = a + \epsilon$, where ϵ is a random error term which is often assumed to be normally distributed. By making further assumptions, such as constant absolute risk aversion (CARA) utility and restricting attention to linear contracts, authors can then derive closed-form solutions. This is particularly useful for extensions and applications, such as contracting in the presence of multiple agents or multitasking.

In the remainder of this section, we first consider the basic model when the agent is risk neutral and has unlimited liability. We then move onto two key insights of the literature: the *rent vs. efficiency* trade-off when the agent is financially constrained and the *insurance vs. efficiency* trade-off when the agent is averse to risk.

2.2 Risk Neutrality and Unlimited Liability

Consider a model in which effort is continuous, with $\mathcal{A} = \mathbb{R}_+$, while the performance measure is binary so that $\mathcal{X} = \{x_L, x_H\}$. The probability of x_H being realised is denoted by the increasing concave function $p(a)$, with $p(0) = 0$, so that the agent's effort-choice problem is concave in a and the first-order approach is valid. Let w_H and w_L respectively denote the wage to be paid when the signals x_H and x_L are realised. The principal's problem then becomes:

$$\max_{a,\, w_H,\, w_L} \; v(a) - p(a)w_H - [1 - p(a)]\, w_L \tag{9}$$

$$a \in \arg\max_{\hat{a}} \; p(\hat{a})w_H + [1 - p(\hat{a})]\, w_L - c(\hat{a}) \tag{10}$$

$$p(a)w_H + [1 - p(a)]\, w_L - c(a) \geq 0 \tag{11}$$

where we have assumed $\overline{U} = 0$ for simplicity. Setting the derivative of (10) equal to zero and rearranging yields:

$$w_H - w_L = \frac{c'(a)}{p'(a)} \tag{12}$$

Next, note that the principal wishes to set w_L as low as possible while still ensuring the agent's participation; this implies that (11) will bind, so that:

$$w_L = c(a) - p(a)\frac{c'(a)}{p'(a)} \tag{13}$$

Letting $w_L = w_L(a)$, note that our restrictions on $c(\cdot)$ and $p(\cdot)$ imply that $w_L(0) = 0$ and $w_L' < 0$, so that this fixed payment becomes negative for positive effort levels.[8] Inserting (12) and (13) into the principal's objective function (9) then yields the following simplified problem:

$$\max_a \; v(a) - c(a) \tag{14}$$

so that the principal's costs of implementation are equal to the effort costs of the agent, identical to the full information case. Intuitively, the principal is able to provide appropriate effort incentives by making the wage spread sufficiently large. Holding this wage spread constant, since there are no restrictions on the available wage payments, the principal can reduce the lower payment until the expected wage is exactly equal to the agent's effort costs (plus any outside option). A risk neutral agent does not need to be compensated for wage uncertainty and is therefore still willing to accept the contract. Clearly, since the principal's costs of implementing any effort level are equal to those of the full information case, first-best effort a^{FB} will continue to be induced.

Before we continue, we pause to make two remarks regarding the performance measure. First, note that while it seems intuitive that the principal will benefit from access to 'better' signals, our analysis implies that in this environment she can achieve her first-best profit level using *any* binary signal which satisfies our restrictions, regardless of how weakly related it is to the agent's effort provision.[9] Second, in environments where the performance measure has many realisations, our analysis also implies that the principal cannot improve her payoff by offering a contract with more than two possible wages.[10] As we shall see shortly, this is no longer the case when the agent is risk averse.

[8] Taking the derivative of (13) yields:

$$\frac{dw_L}{da} = c'\frac{pp''}{p'p'} - c''\frac{p}{p'} < 0$$

[9] One way to formalise this analysis would be to consider the effort elasticity of $p(\cdot)$, as in Demougin and Fluet (2001). One can show that even as this elasticity tends to zero, the first-best can still be implemented. However, the wage spread required to induce the desired effort becomes unbounded, so that $w_L \to -\infty$.

[10] In fact, in such environments the principal's optimal contract can usually take many different forms. See also Demougin and Fluet (1998).

2.3 Risk Neutrality and Limited Liability

We now consider our previous example, but further assume that the agent is financially constrained so that the principal can only offer non-negative wages. In our environment, the additional constraint (8) then takes the form: $w_H, w_L \geq 0$. Since the incentive compatibility constraint is unchanged, we proceed as before and solve for the necessary wage spread to induce effort a, which again yields (12). The principal again wishes to set w_L as low as possible while ensuring participation. However, we previously found that (13) becomes negative for positive effort levels, which violates the limited liability constraints. Accordingly, the principal sets $w_L = 0$, meaning that the agent's participation constraint will no longer bind.[11] Substituting into (9), the principal's problem then becomes:

$$\max_a \; v(a) - c'(a)\frac{p(a)}{p'(a)} \tag{15}$$

The principal's costs of implementing effort a with a financially constrained agent are therefore represented by the second term in (15), $c'(a)\frac{p(a)}{p'(a)}$. The difference between these costs and the agent's effort costs represents a rent paid to the agent, which we denote by $R(a) = p(a)\frac{c'(a)}{p'(a)} - c(a)$. It is straightforward to show that this rent is both positive for $a > 0$ and increasing in effort.[12] Accordingly, in the presence of limited liability, there is a divergence between the agent's effort costs and the principal's costs of implementing this effort. When deciding on the effort level to be induced, the principal must now take into account not only the marginal effort costs c', but also the marginal rent R' which must be paid to the agent. Typically, this trade-off between efficiency and rent leads to an underprovision of effort relative to the first best level, with the solution to (15) referred to as the *second-best* effort level and denoted by a^{SB}.

2.4 Risk Aversion

The study of incentive provision with risk averse agents has received significant attention in the moral hazard literature. As an introduction, we consider the foregoing environment with continuous effort and a binary performance measure, although we only provide an informal analysis. We revert to our original assumption that the principal faces no constraints on the wage payments that can be offered.

[11] Note that this is not necessarily the case if the agent's outside option is sufficiently large.

[12] To see this, first note that $R(0) = 0$ due to our restrictions on $p(\cdot)$ and $c(\cdot)$. Next, taking the derivative yields:

$$R' = c''\frac{p}{p'} - c'\frac{pp''}{p'p'} > 0$$

Since under moral hazard the performance signal is imperfectly correlated with effort, the provision of incentives necessarily introduces risk into the agent's wage. While we found that this was not an issue when the agent was risk neutral, this is no longer the case with risk aversion. Intuitively, in order to induce participation, the principal must pay an expected wage which covers not only the agent's effort costs and outside option, but also his costs of bearing risk. Typically, since the wage spread between the payments w_H and w_L is increasing in the amount of effort to be induced, this so-called *risk premium* will also be increasing in a. Accordingly, when deciding on the effort to be implemented, the principal must take into account the marginal effects on both the agent's effort costs and risk premium. Similar to the foregoing case with limited liability, this creates a trade-off between efficiency and insurance which will typically lead to the underprovision of effort.

We next turn our attention to a more complex environment, with a continuous performance measure so that $\mathcal{X} = [\underline{x}, \overline{x}]$. It is assumed that x is distributed according to the function $F(x; a)$, with associated density $f(x; a)$. The principal's optimisation problem then becomes:

$$\max_{a,\, w(x)} \; v(a) - \int_{\underline{x}}^{\overline{x}} w(x) f(x; a)\, dx \tag{16}$$

$$a \in \arg\max_{\hat{a}} \; \int_{\underline{x}}^{\overline{x}} u(w(x)) f(x; \hat{a})\, dx - c(\hat{a}) \tag{17}$$

$$\int_{\underline{x}}^{\overline{x}} u(w(x)) f(x; a)\, dx - c(a) \geq \overline{U} \tag{18}$$

As previously discussed, it is standard to replace the agent's optimisation problem (17) with its first-order condition:

$$\int_{\underline{x}}^{\overline{x}} u(w(x)) f_a(x; a)\, dx - c'(a) = 0 \tag{19}$$

However, this method is generally invalid unless we place some restrictions on the distribution function $F(x; a)$. Rogerson (1985) shows that two conditions in particular are sufficient to guarantee validity of the first-order approach. First, the strict Monotone Likelihood Ratio Property (MLRP) requires that:

$$\frac{\partial}{\partial x}\left[\frac{f_a(x; a)}{f(x; a)}\right] > 0, \;\; \forall x \in (\underline{x}, \overline{x}) \tag{20}$$

Second, the strict Convexity of the Distribution Condition (CDFC) states that:

$$F_{aa}(x;a) > 0, \;\; \forall x \in (\underline{x}, \overline{x}) \tag{21}$$

We assume that $F(x;a)$ satisfies both of these properties and, as such, replace (17) with its first-order condition (19).[13] Pointwise optimisation of the Lagrangian associated with the principal's maximisation problem yields the following first-order condition with respect to $w(x)$:

$$\frac{1}{u'(w(x))} = \lambda \frac{f_a(x;a)}{f(x;a)} + \mu \tag{22}$$

where λ and μ are the Lagrangian multipliers associated with the incentive compatibility and participation constraints, respectively. Holmström (1979) shows that both constraints will bind, so that λ, $\mu > 0$. The optimal wage scheme is then characterised by (22) for all $x \in [\underline{x}, \overline{x}]$, along with (18) and (19). Our assumption of MLRP guarantees that the RHS of (22) is strictly increasing in x. This, along with risk aversion of the agent ($u'' < 0$), implies that the optimal wage scheme also is strictly increasing in x, so that outcomes which are indicative of higher effort are rewarded with strictly larger payments. In fact, while we have only considered a single performance measure here, the optimal contract will condition payments on *any* signal which reveals information about the agent's effort. As such, the wage scheme should be contingent on many different variables if they are each able to improve estimates of effort provision. This finding, which is perhaps the most important in the moral hazard literature, is known as the *sufficient statistic result* (Holmström, 1979).

Intuitively, conditioning payment on some relevant piece of information allows for a better estimate of the agent's effort and therefore reduces the risk inherent in the relationship, allowing for a lower risk premium. Accordingly, optimally filtering out risk requires making wages contingent on all such information. In the limit, as the inclusion of a large number of variables allows the principal to infer the agent's effort with certainty, the required risk premium reduces to zero and the first best effort can be implemented at no extra cost (Macho-Stadler and Pérez-Castrillo, 2018).

2.5 Risk Aversion and Multiple Agents

The sufficient statistic result has particularly interesting implications for incentive provision when a principal employs multiple agents. The central question here relates to whether incentive payments to an individual should depend on the

[13] MLRP and CDFC are strong conditions and as a result it is difficult to find distribution functions which satisfy both properties. See LiCalzi and Spaeter (2003) for some examples.

performance of other agents, as well as his own. Moreover, if so, should wages be increasing or decreasing in the performance of others?

To explore these issues, we consider a continuous effort environment ($\mathcal{A} = \mathbb{R}_+$) in which a principal wishes to employ two identical agents, denoted by 1 and 2, to undertake a specific effort level a. The principal has access to a performance measure for each agent $i = 1, 2$, which takes the form $x_i = a_i + \epsilon_i$, where ϵ_i is a normally distributed random variable with zero mean and a variance of σ^2. In addition, the covariance between the two error terms, ϵ_1 and ϵ_2, is denoted by σ_{12}.[14] The utility functions of agents are now assumed to take the constant absolute risk aversion (CARA) form:

$$U(w, a) = -e^{-\rho[w-c(a)]} \tag{23}$$

where $\rho > 0$ is the agent's coefficient of risk aversion. Finally, we restrict our attention to wages which are linear in performance measures for each agent:

$$w_i = \eta + \gamma x_i + \delta x_j \tag{24}$$

Given these assumptions, maximisation of expected utility is equivalent to maximisation of the certainty equivalent, which can be expressed as:

$$CE_i(a_i) = \eta + \gamma a_i + \delta a_j - c(a_i) - \frac{\rho}{2}\left[\gamma^2\sigma^2 + \delta^2\sigma^2 + 2\gamma\delta\sigma_{12}\right] \tag{25}$$

where the final term measures the agent's disutility from exposure to risk.[15] The agent will choose the effort level a_i which maximises (25). Rearranging the first-order condition yields $\gamma = c'(a_i)$; by assuming that $c(\cdot)$ takes a specific functional form we can therefore obtain an explicit solution for γ. Note that compensating the agent based on the performance of others has no impact on incentives. In fact, since participation can be ensured by appropriately adjusting the fixed payment η, the principal will set δ with the sole objective of minimising the agent's risk exposure in order to relax the participation constraint. Accordingly, minimisation of the final term of (25) yields:

$$\delta = -\gamma\frac{\sigma_{12}}{\sigma^2} \tag{26}$$

Since γ, $\sigma^2 > 0$, δ will take the opposite sign to σ_{12}. We first consider the case where $\sigma_{12} > 0$, so that the error terms of Agents 1 and 2 are positively correlated,

[14]Note that our analysis considers an *informational* linkage between the performance of agents, since error terms are correlated. A different rationale for remunerating agents based on the performance of others occurs when there is a *technological* linkage, so that one agent's performance depends upon the effort provision of another. See Holmström (1982) and Mookherjee (1984).

[15]See Bolton and Dewatripont (2005) for a detailed derivation in a similar model.

in which case $\delta < 0$ and an agent's wage is therefore decreasing in the performance of the other agent. This is known in the literature as *relative performance evaluation* (RPE). Intuitively, observing a high measure of performance for Agent 1 suggests that the error term ϵ_1 is positive. Since $\sigma_{12} > 0$, it is then likely that ϵ_2 is also positive. Similarly, a low measure of performance for Agent 1 is suggestive of a negative ϵ_2. The sufficient statistic result therefore implies that the realisation of x_1 should be incorporated into the wage scheme of Agent 2, since it is informative about ϵ_2 and can therefore be used to improve the estimate of Agent 2's effort provision. By penalising Agent 2 as the performance of Agent 1 increases, but rewarding him as it decreases, the principal is able to reduce the variance of Agent 2's wage and therefore reduce the necessary risk premium. In fact, in the limit as $\rho \to 1$ and the two error terms become perfectly correlated, each agent's performance becomes affected by a single, common source of noise; by filtering out this shock the principal is able to eliminate exposure to risk and thus approximate first-best incentives.

The above discussion therefore highlights that the benefit of RPE is *not* to induce competition between agents, but to reduce the risk inherent in the relationship, as argued by Holmström (1982). Indeed, when $\rho = 0$, the optimal δ is also equal to zero and wage schemes become independent. In this case, since there is no correlation between error terms, conditioning an agent's remuneration on the performance of others would actually *increase* risk exposure and therefore provides no benefit. Finally, there may also exist environments in which $\rho < 0$ so that error terms are negatively correlated, in which case $\delta > 0$ and, for similar reasons, a form of *team contracting* in which agents are rewarded for the high performance of coworkers becomes optimal.[16]

3 Behavioural Models

3.1 Other-Regarding Preferences and Inequity Aversion

3.1.1 Introduction

The notion that preferences typically feature a social component is perhaps best illustrated by experimental findings from two simple, two-player games: the ultimatum game and the dictator game. In both games, one player (*the proposer*) decides how a fixed surplus should be shared between the two players. In the former game, the other player (*the responder*) faces a choice between accepting

[16]Our analysis considered the optimal contract required in order to induce a given effort level, a. However, by reducing the risk exposure of agents, basing remuneration on the performance of others can mitigate the insurance vs. incentives problem considered previously and lead the principal to implement higher levels of effort.

the proposed division or destroying the entire surplus so that both players have a zero payoff. In the latter game, the offer can only be accepted. The classical approach predicts that in both games the proposer will end up with (almost) the entire surplus.[17] However, experimental evidence from several different countries, with stakes of various magnitudes, have found the following consistent regularities. First, in the ultimatum game, low offers are often rejected by the responder, demonstrating a willingness to forego individual wealth in order to reduce the payoff of the opponent. Perhaps in anticipation of this, low offers are rare, so that proposers often offer high shares of the surplus. Second — and perhaps even more surprisingly — in the dictator game, some proposers offer positive shares; that is, they are willing to reduce their own income in order to increase the payoff of their opponent.

Beyond these two games, there is a large amount of additional evidence — both experimental and empirical — that individuals care about the payoffs of others.[18] In response, economists have developed a number of theoretical models in order to formalise such preferences. Loosely speaking, they can be separated into two distinct classes of models.[19]

The first class contains models of *intentions-based reciprocity*, whereby individuals care about the *intentions* of others. In this approach, agents who feel that they have been treated kindly by a peer wish to return the favour, while those who feel they have been treated poorly prefer to hurt the peer. Such models therefore account for the intricate nature of social concerns. However, since an agent's utility then depends on their beliefs about the intentions of others, the reciprocity approach must utilise psychological game theory, as developed by Geanakoplos et al. (1989), rather than employing standard game theoretical tools. As such, these models become fairly complex even in simple environments, generally feature multiplicity of equilibria and are therefore often ill-suited for predictive purposes or economic applications (Fehr and Schmidt, 2006).

In the second class of models, those which consider *social preferences*, an individual's utility function is assumed to depend directly on both their own payoff and the payoffs of others within some relevant reference group. Different models of social preferences vary in the way in which the individual's utility depends on

[17] This is clear in the dictator game, since offers cannot be rejected and therefore a purely self-interested individual should choose to claim the entire surplus. For the ultimatum game, depending upon the specifics of the environment, the responder will either i) accept all offers or ii) accept all offers weakly above the lowest possible strictly positive offer, $\epsilon > 0$. The proposer should therefore offer either 0 or ϵ to the responder; see Fehr and Schmidt (1999) for a more detailed discussion.

[18] See for example the survey article of Fehr and Schmidt (2006), who also discuss neuroeconomic evidence.

[19] Fehr and Schmidt (2006) consider a third class of models, in which individuals have social preferences which vary depending on the reference group with which they interact.

the outcomes of others. For instance, models of altruistic agents assume that an individual's utility is always increasing in the payoffs of others, whereas the utility of spiteful agents is always decreasing in these payoffs (see for example Levine, 1998). We will focus on one model of social preferences in particular: *inequity aversion*, as developed by Fehr and Schmidt (1999), in which individuals have a preference for fair and equitable allocations.[20] This implies that an agent's utility, in contrast to the aforementioned models, can either be increasing or decreasing in the payoffs of others, depending on whether these changes move the resulting allocation closer to or further away from the equitable ideal. As shown by Fehr and Schmidt (1999), even a simple formalisation of such preferences allows for a unifying explanation of a wide range of seemingly contradictory phenomena.

In this subsection, we briefly present the Fehr and Schmidt model of inequity aversion before discussing two issues which are particularly important when considering other-regarding preferences within firms.[21] First, what is the relevant social reference group for an individual within an organisation? Second, do comparisons focus only on monetary outcomes, or should they also incorporate the extent to which agents undertake productive effort?

3.1.2 The Fehr and Schmidt Model of Inequity Aversion

Let us initially consider an environment which features two agents, i and j, each of whom have incomes which are denoted by w_i and w_j respectively. The Fehr and Schmidt specification assumes that Agent i's preferences can be represented by the utility function:

$$u_i(w_i, w_j) = w_i - \alpha_i \max\{w_j - w_i, 0\} - \beta_i \max\{w_i - w_j, 0\} \tag{27}$$

The first term measures standard consumption utility, where risk neutrality is assumed for simplicity. The second and third terms measure the disutility from disadvantageous and advantageous inequity, respectively. Agent i's income is compared with that of Agent j. If $w_j > w_i$, then Agent i is behind and suffers from disadvantageous inequity, or feelings of *envy*. The utility loss is equal to the size of the difference in incomes, multiplied by the coefficient of envy, $\alpha_i \geq 0$. Incorporation of this term into the utility function can account for the rejection of positive offers by the responder in the ultimatum game: while accepting the

[20] A similar model is presented by Bolton and Ockenfels (2000), although there are some key differences. Models of this type are sometimes referred to as *inequality aversion*, reflecting the notion that in many environments the most fair or equitable allocation involves equality between the payoffs of individuals. An axiomatic foundation to the Fehr and Schmidt model is provided by Neilson (2006).

[21] Throughout, we slightly abuse terminology and interchangeably use the terms *inequity averse* and *other-regarding* to describe preferences which are characterised by envy, compassion, status-seeking, or some combination of these.

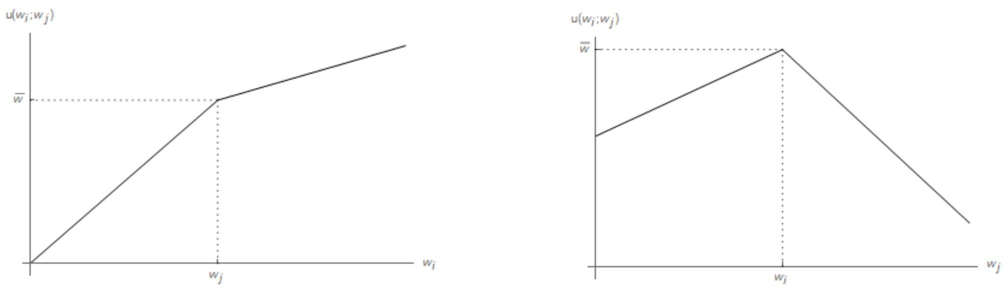

Figure 1: The variation in the function $u_i(w_i, w_j)$, as defined by (27), as w_i and w_j change (LHS and RHS, respectively). When $w_i = w_j = \overline{w}$, there is no loss in utility from inequity and $u(\overline{w}, \overline{w}) = \overline{w}$.

offer would increase the responder's income and associated monetary utility, the resulting allocation may result in feelings of envy which make rejecting the offer preferable.

Alternatively, if $w_i > w_j$, then Agent i is ahead and experiences feelings of advantageous inequity or *compassion*.[22] Similarly, $\beta_i \geq 0$ is the coefficient of compassion, which is multiplied by the income difference to yield the utility loss in this case. The inclusion of compassion into the utility function can explain positive offers by the proposer in the dictator game, since the inequality associated with claiming the entire surplus leads to a utility loss which can outweigh the increase in utility from a higher payoff. Clearly, when incomes are equal and $w_i = w_j = \overline{w}$, then the second and third terms are both zero and $u_i(\overline{w}, \overline{w}) = \overline{w}$. Figure 1 plots (27), showing how utility changes with variations in both w_i and w_j.

The functional form proposed by Fehr and Schmidt is designed to capture inequity concerns in the simplest possible way. In the literature, several authors have either extended or modified (27) in order to provide a more realistic description of individual preferences. It is straightforward to allow for classical risk aversion by replacing the first term in (27) with $m(w_i)$, where $m(\cdot)$ is an increasing concave function as standard. In addition, the assumption of piecewise linearity for inequity concerns can often cause the model to make extreme predictions. For instance, in the dictator game, an individual with $\beta < 0.5$ will choose to take the entire surplus while an individual with $\beta > 0.5$ will choose to equally divide the surplus. Hence all offers are either very fair, or very unfair; the intermediate offers which are often observed in the data are ruled out. As noted by Fehr and Schmidt (1999), this can be remedied by assuming that inequity concerns are captured by a non-linear function, an approach which is furthered by several authors in the

[22] Various terminology has been utilised in the literature to denote advantageous (compassion, guilt, empathy) and disadvantageous (envy, jealousy) inequity concerns. For consistency and simplicity, we use the terms *envy* and *compassion* throughout, but emphasise that this usage is not intended to convey any connotations regarding the underlying source of the emotions associated with a distaste for inequitable outcomes.

literature.[23]

The Fehr and Schmidt model implies that social comparisons enter into the utility function ex post, so that an agent only suffers if any potential inequity actually arises. Some authors deviate from this and model inequity concerns as entering ex ante. Intuitively, this implies that agents suffer from the exposure to the possibility of inequity, even if such outcomes are not actually realised. For instance in the framework of Bartling (2011), such an approach allows for risk aversion over wages but risk neutrality over inequity, which he notes that, in the absence of any evidence to the contrary, is a possible modelling choice.

It makes sense to impose some restrictions on β_i, the coefficient of compassion. Suppose that $w_i > w_j$; an agent with $\beta_i > 1$ would prefer in this situation to dispose of or burn money in order to reduce inequity, since compassion in this case is a stronger motivator for the agent than income. This seems unrealistic. Accordingly, Fehr and Schmidt assume that $\beta_i \leq 1$ throughout, so that in all cases the marginal utility of money will be (at least weakly) positive.[24] They also assume that $\beta_i \leq \alpha_i$, so that an individual's inequity concerns are at least as pronounced when they are behind as when they are ahead. There is strong evidence for the validity of this assumption, which implies that individuals are loss averse in social comparisons (Fehr and Schmidt, 1999). An upper bound for the coefficient of envy α_i is not imposed, however, and they suggest that some individuals may have preferences which are most consistent with $\alpha_i \approx 4$ or higher.

While it is explicitly assumed that $\beta_i \geq 0$, Fehr and Schmidt do note that there will be some situations whereby $\beta_i < 0$, so that individuals gain utility from advantageous inequity and actually prefer to be ahead of others. We follow the literature — in which such cases are often allowed for — and refer to these preferences as *status seeking.*[25] Frank (1985) provides extensive evidence that individuals care deeply about which positions they occupy in income hierarchies, while more generally Fershtman (2008) surveys the role of social status within economics. Fershtman et al. (2012) find evidence that the norms induced by the environment will typically influence whether people like or dislike advantageous inequity.

The above discussion highlights two more general points. First, Fehr and Schmidt emphasise that the extent of inequity aversion will be heterogeneous between individuals and provide some examples of possible distributions (see also

[23] Note that this approach can also eliminate the non-differentiability of the function when $w_i = w_j$, simplifying applications of the model.

[24] Note that in this situation, while an agent with $0.5 < \beta_i < 1$ would not wish to burn money, they would be willing to transfer some of their payoff to the other agent in order to reduce inequality. Accordingly, some authors make the stronger assumption that $\beta_i \leq 0.5$.

[25] Alternative terminology for such preferences found in the literature includes *competitiveness, spitefulness* and *pride.*

Fischbacher and Gachter, 2010 and the references therein). In fact, evidence suggests that there may be systematic differences in aversion to inequity depending on characteristics. Numerous papers have shown that Western Europeans have a stronger preference for more equal income distributions compared to the United States (see for instance Alesina et al., 2004), while Kuhn and Villeval (2015) find evidence consistent with women having stronger dislike of advantageous inequity than men. Second, the strength and nature of an individual's social preferences cannot be thought of as fixed values which are robust to all situations and environments. As we will discuss subsequently, evidence suggests that an individual's aversion to inequity will vary depending on factors such as the social context and the composition of the reference group.

We now consider the more general case of an environment with N agents in the reference group. Let the income vector of all other agents be denoted by w_{-i}; the utility function of Agent i then takes the following form:

$$u_i(w_i, w_{-i}) = w_i - \alpha_i \frac{1}{N-1} \sum_{j \neq i} \max\{w_j - w_i, 0\} - \beta_i \frac{1}{N-1} \sum_{j \neq i} \max\{w_i - w_j, 0\} \quad (28)$$

Agent i makes a pairwise comparison of his income with each other agent in the reference group. When he is ahead of another agent, he experiences compassion; when he is behind, he experiences envy. The overall loss from utility will encompass a mix of these feelings. Notably, the agent's inequity is self-centred in the sense that he does not care about inequitable outcomes between others in the reference group, but solely on how his own payoff compares to those of others. Alternative approaches have been explored in the literature. For example, Bolton and Ockenfels (2000) present a model of inequity aversion in which agents prefer to be as close as possible to the average of all other agents in the reference group. This can lead to conflicting predictions between the two models. Suppose that $N = 3$ and the payoff vector is $(w_1, w_2, w_3) = (0, 100, 50)$. Under Fehr and Schmidt preferences, Agent 3 suffers disutility both from being £50 ahead of Agent 1 and from being £50 behind Agent 2. In contrast, under the Bolton and Ockenfels (2000) specification there is no utility loss, since Agent 3's payoff is exactly equal to the average payoff of Agents 1 and 2. Engelmann and Strobel (2004) design a series of experiments to compare the models and find that the Fehr and Schmidt specification generally outperforms that of Bolton and Ockenfels, while Buckingham and Alicke (2002) find evidence that social comparisons with individuals are stronger than those with averages.

There are, of course, alternatives. For example, agents might limit their comparisons to salient others within the reference group, such as those who have the highest or lowest payoffs; in this case it is the maximum level of inequity between an agent and others in the group which is important.

While there is little evidence of which specification is generally most appropriate, the modelling choice will often have important implications for economic applications. However, for our purpose, the vast majority of papers in the literature study simplified models in which the reference group is made up of two single agents. Accordingly, each individual compares their payoff to that of one single other, so that all the previously mentioned specifications become qualitatively similar.

Finally, note that inequity concerns in the Fehr and Schmidt model are weighted by the term $\frac{1}{N-1}$, so that the overall coefficient of disutility from inequity remains constant as the reference group expands or contracts. There is very little evidence regarding the accuracy of this assumption, with some authors taking an alternative approach (e.g. Goel and Thakor, 2006).

3.1.3 What is the correct reference group?

In many situations, it is not exactly clear how the relevant reference group should be defined. Fehr and Schmidt (1999) argue that it will depend upon factors such as the social context, the saliency of particular individuals and social proximity. The institutional environment in which interaction takes place is also likely to play a key role. In the context of a labour relationship, there are numerous possible candidates. Within the firm, a worker might compare their wage with others in the same team, others who have a similar role or with all workers at a comparable level. They may even make vertical pay comparisons with either management or subordinates. Beyond the firm, the reference group might include those in similar roles at different firms, friends, family or neighbours. Workers may even be interested in their wage relative to industry-wide or national averages.

Several authors have noted that there is scarce scientific evidence on this issue. Experimental evidence provides little guidance here, since in the laboratory reference groups will be largely determined by the experimental design. However, the question is important in our context since there are subtle implications for principal-agent models. For instance, in the standard approach to incentive contracting, it is common to model an agent's outside utility as being exogenously given and fixed. When agents have other-regarding preferences, however, this assumption essentially implies that the agent no longer makes social comparisons when he is not employed by the firm.[26]

[26]This is particularly important when undertaking a comparative static analysis. To see this, note that for a given contract, an increase in the strength of an agent's other-regarding preferences will often *ceteris paribus* lead to a loss in utility. In cases where the outside utility is exogenously

The prevailing approach in the literature has been to continue to make this assumption and therefore assume that the reference group is limited to coworkers.[27] There are some arguments to support this approach. Festinger (1954) develops a theory of social comparison in which the tendency of individuals to compare themselves to others becomes stronger as the relevant referents become more similar. Accordingly, given a range of possible referents, individuals will tend to choose those who are considered to be equals. A similar argument has been made specifically with respect to envy; Elster (1991) quotes Aristotle (Rhetoric, 1388a), who argues that "we envy those who are near us in time, place, age, or reputation." Both Adams (1963) and Akerlof and Yellen (1988) have noted that coworkers will often satisfy this criteria better than others. The issue has also received significant attention in the organisational psychology literature, where the related matter of information availability has also been explored (Kulik and Ambrose, 1992). Intuitively, the extent of comparisons with others will necessarily depend on how much of the relevant information is available to an individual.

Some interesting empirical support comes from cases of mergers and acquisitions, where expansions in the boundary of the firm can lead to changes in the relevant reference group. Both Williamson (1985) and Kole and Lehn (2000) provide case studies of mergers which failed in part due to the need to ensure internal equity between newly integrated workforces. Similarly, Demougin et al. (2006) document initial difficulties in the merger between Daimler Benz and Chrysler Corporation due to pay disparity between senior executives, necessitating significant pay increases in order to reduce inequalities. While most of these intra-firm social comparisons are assumed to be horizontal (i.e. with comparable coworkers such as members of the same team), there is also some evidence for vertical comparisons, whereby workers show concern for the pay of managers. For example, Englmaier and Wambach (2010) describe the unrest at American Airlines in 2003 when large wage cuts were imposed upon workers in order to avoid bankruptcy, but the pay of executives was left unchanged.

Despite the prevalence of the assumption that the reference group is limited to those within the firm, there does exist some evidence to suggest that this may not always be the case. Clark and Oswald (1996) provide empirical evidence that happiness at work is significantly negatively correlated with a comparison income, calculated by predicting the typical income of someone with the individual's observable characteristics. Similarly, Luttmer (2005) finds that self-reported happiness is decreasing in the earnings of neighbours, defined as those who live in the same

given and the participation constraint is initially binding, this utility loss implies that the agent now prefers to reject the contract. However, this is not necessarily the case if the agent's outside utility also varies with the parameter(s). Accordingly, the impact on the principal's costs can depend upon which modelling specification is chosen.

[27] A notable exception is Goel and Thakor (2006).

locality, each of which has a population of 150,000 on average. Surprisingly, the findings suggest that similar sized increases in own earnings and reductions in the earnings of neighbours each lead to a fall in happiness of approximately the same magnitude.[28] Babcock et al. (1996) provide evidence that during the course of wage bargaining between teacher's unions and school boards, parties make use of reference wages from schools in comparable districts.

A common justification for assuming that the reference group is limited to comparisons within the firm comes from the survey evidence of Bewley (1998, 2004), which suggests that employees know little about pay rates at other companies. However, lacking knowledge of the wages of others does not preclude comparisons based on estimations. Moreover, while information about wages at a particular firm may indeed be sparse, individuals may have reasonable knowledge of industry averages, especially if labour unions promote awareness of pay rates in order to stimulate the interest of members (Bewley, 2004). Additionally, in recent years there has been a movement towards wage transparency (Mas, 2017).

Finally, even if we are able to pin down exactly which others will constitute the reference group in a given environment, its composition may have a significant impact on concerns for inequity. Loewenstein et al. (1989) provide evidence that the degree of inequity aversion shown by individuals varies both with the relationship to the relevant other (positive, neutral or negative) and the social context of the interaction (business or personal). Casciaro and Lobo (2008) document psychological evidence on the reciprocity of perceived liking; similarly, it is likely that the extent to which individuals exhibit other-regarding preferences (and compassion in particular) will depend in part upon how they feel this would be reciprocated following alternative, counterfactual outcomes. Different corporate cultures and hiring policies will thus also contribute to variations in the prevalence of inequity concerns between organisations (Englmaier and Wambach, 2010).

3.1.4 What is the correct comparison?

In the Fehr and Schmidt model, agents make unidimensional comparisons of monetary payoffs with one another. However, when studying incentive contracting, there is the possibility that agents incorporate effort provision when considering equitable outcomes. The approaches in the literature we review are varied. Some authors model agents as simply comparing wages with one another, while others assume that the comparison is between wages net of effort costs, so called *net wages*.

Evidence suggests that the latter specification may be more realistic in at least

[28] See also the studies discussed in the introduction to this thesis; Clark et al. (2008) offer a survey of this literature.

some environments. Equity Theory is a social psychological theory of fairness which has its roots in the work of Aristotle and proposes that the fair or equitable ratio of outcomes is proportional to the ratio of inputs (Konow, 2003). Adams (1965) argues that feelings of inequity arise when an individual perceives their ratio of outcomes and inputs to be unbalanced relative to others, resulting in unpleasant emotional states such as dissatisfaction, anger or guilt. In our context, this theory suggests that individuals will take both wages (outcomes) and effort (inputs) into consideration when making social comparisons.

Experimentally, in dictator games in which the surplus to be divided is dependent on production in a prior stage, researchers have found that proposers offer more to those who were more productive and as such contributed more to the surplus (see for instance Frohlich et al., 2004). Further evidence comes from Abeler et al. (2010), who conduct a gift-exchange experiment in which a principal observes the effort of two agents and decides on appropriate wage payments. They find that effort provision is substantially lower when the principal must pay equal wages, compared to the case where wage differentiation is allowed. In the case of imposed wage equality, agents who initially undertake higher effort than their coworker get discouraged and subsequently reduce their effort, consistent with the notion that workers perceive equal pay for unequal work as being unfair.

However, the nature of comparisons may be context dependent. Researchers in social psychology have found evidence which suggests that the way in which individuals compare tends to be biased in a self-serving manner. For instance, workers who exert higher effort than others will compare net wages and also view wages proportional to effort as being fair. In contrast, those who exert lower effort than others will compare only wages and believe that a fair wage pays workers equally (see Neilson and Stowe, 2010 and the references therein). In addition, when comparisons are multidimensional — such as wages and effort, or wages and job perks — agents may focus on the dimension in which they are worse off (Goel and Thakor, 2006).

As a final remark, one might also question the extent to which workers are able to make any form of wage comparisons with colleagues, since many firms have wage secrecy policies. However, in these cases there is evidence to suggest that individuals continue to make comparisons based on estimations, with a tendency to overestimate the wages of others resulting in heightened perceptions of inequity (see for instance Mahoney and Weitzel, 1978). With respect to effort considerations, clearly comparisons will be easier (and therefore likely be more salient) when individuals work closely with one another and can readily observe each other's behaviour. It follows that the majority of papers which consider team production model agents as incorporating both wages and effort into their comparisons.

3.2 Prospect Theory, Reference-Dependent Preferences and Loss Aversion

3.2.1 Introduction

To this day, the most commonly applied model of decision making under risk in economics is expected utility theory, despite there existing a wealth of laboratory evidence that people's behaviour systematically violates its predictions. In response to some of these inconsistencies, Kahneman and Tversky in their seminal 1979 paper developed a new model of decision making named *prospect theory*, which was able to explain much of this evidence and remains widely seen as the best available description of risk preferences (Barberis, 2013).[29] Prospect theory has four key features which distinguish it from the standard model:

i) *Reference-dependence.* Individuals are assumed to evaluate outcomes relative to some reference point r, so that we have $u(w;r)$.

ii) *Loss aversion.* Individuals are more sensitive to losses (outcomes below the reference point) compared to gains (outcomes above the reference point) of the same magnitude.

iii) *Diminishing sensitivity.* Individuals are much more sensitive to variations in outcomes when they are close to the reference point. Mathematically, this implies that marginal utilities are decreasing as we move further away from the reference point. Accordingly, the utility function is convex in the loss region and concave in the gain region. Individuals are then risk seeking for losses, but risk averse for gains.

iv) *Probability weighting.* Individuals do not weight outcomes by their objective probabilities, but rather by transformed subjective probabilities or decision weights. In particular, they tend to overweight low probability events, while underweighting high probability events.

The first three of these features represent a departure from the utility function of standard economic theory, which is usually considered to be reference independent and everywhere (weakly) concave.[30] The fourth feature implies a movement away from the linear probabilities of expected utility theory. Together, these properties

[29] For a more detailed description, see the original 1979 article or the updated version of the model, also known as *Cumulative Prospect Theory* (Tversky and Kahneman, 1992). Barberis (2013) provides a simple introduction, before reviewing some of the ways the ideas of prospect theory have been applied to economic settings.

[30] It is noteworthy that Kahneman and Tversky (1979) introduced prospect theory as having a *value function*, often denoted $v(\cdot)$, rather than a utility function denoted by $u(\cdot)$. This convention is largely followed in much of the following literature. For the sake of clarity, however, we continue to use the terminology *utility function* and as such use the notation $u(\cdot)$.

can account for a number of results which are inconsistent with the classical model. For instance, Rabin (2000) shows that the standard theory cannot explain observed attitudes to risk over small-stake gambles. Suppose an individual turns down a 50-50 bet between losing £100 and gaining £110. Under the standard model, aversion to risk arises solely due to diminishing marginal utility, leading to concavity of the utility function. Since the aforementioned gamble has a positive expected value, rejection implies that there is a significant degree of local curvature around the current level of wealth. Now further suppose that the individual would continue to turn down this gamble for all initial wealth levels. In this case, Rabin (2000) shows that the classical theory implies 50-50 bets between losing £1000 and gaining *any* sum of money would also be turned down. Intuitively, rejection of the gamble at all wealth levels implies that the aforementioned significant curvature is present everywhere, resulting in massive curvature when we begin to consider large-stake gambles.

A similar argument comes from a famous observation by Samuelson (1963), who offered a colleague a 50-50 bet between losing \$100 and gaining \$200. The colleague rejected the bet, but claimed that if it was offered 100 times then he would accept. Samuelson (1963) proved that such behaviour is inconsistent with expected utility theory: an individual who rejects a given gamble would also reject an aggregate gamble in which the bet is independently played multiple times. As noted by Rabin (2000), however, the preferences of Samuelson's colleague seem very plausible. Many would reject the single bet. Yet the aggregate bet has an expected value of \$5000 with negligible risk of losing any money at all (1/700), so that most individuals would surely accept it. A direct explanation for the modest scale risk aversion described here comes from loss aversion, which can reconcile significant levels of aversion to risk over small-stakes while maintaining plausible degrees of risk aversion over larger stakes.

Beyond explaining attitudes toward risk, reference-dependent preferences have been used to analyse topics such as the endowment effect (whereby there is a disparity between willingness-to-pay and willingness-to-accept for a good; Kahneman et al., 1991), the famous paradoxes of Allais (1953) (see Kahneman and Tversky, 1979) and negative wage-elasticities for labour supply decisions (Camerer et al., 1997).[31]

In this subsection, we first explore the ways in which economic applications have modelled the ideas of prospect theory, before considering a question of key importance: in a given environment, how is an individual's reference point determined? Finally, we present the Kőszegi and Rabin (2006, 2007) model of

[31] For surveys of applications of prospect theory and reference-dependent preferences to various economic phenomena, see Camerer (2000), Barberis (2013) and O'Donoghue and Sprenger (Forthcoming).

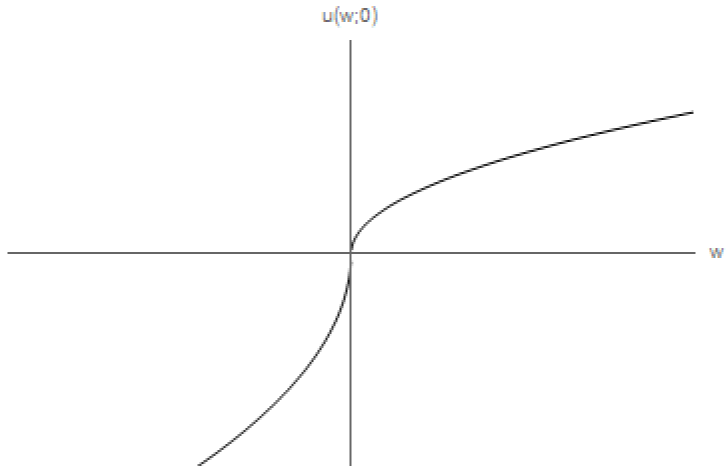

Figure 2: The function $u(w;0)$ as defined by (29), with $\alpha = 0.5$ and $\lambda = 2.25$.

expectations-based reference-dependent preferences, a recent development which has received significant attention in the literature and inspired several important economic applications.

3.2.2 Models of Reference-Dependent Preferences

Tversky and Kahneman (1992) use experimental evidence to provide an example of a possible utility function for preferences under prospect theory. They first assume that $r = 0$, which is consistent with the reference point being determined by the status quo: all outcomes which leave the individual with higher wealth are gains while all those which leave him with lower wealth are losses. Next, they impose the functional form:

$$u(w;0) = \begin{cases} -\lambda(-w)^{\alpha} & w < 0 \\ \\ w^{\alpha} & w \geq 0 \end{cases} \tag{29}$$

and estimate $\alpha = 0.88$ and $\lambda = 2.25$. Figure 2 plots this function, with $\alpha = 0.5$ chosen in order to accentuate the curvature below and above the reference point.

The parameter λ captures loss aversion and it is clear from the figure that the function is steeper immediately below zero compared to immediately above. Diminishing sensitivity is captured by $\alpha < 1$; as we move further away from the reference point in either direction, it can be seen that marginal utility is diminishing, so that the function is convex below the reference point but concave above. Kahneman (2003) argues that preferences can generally be approximated fairly well by a function which takes the form of (29).

Most other studies which attempt to estimate the utility function under prospect theory use either experimental evidence, or evidence from field experiments such as game shows as in Post et al. (2008). Estimations can take either a parametric or non-parametric approach; see the discussion in Booij et al. (2010) for a comparison of the advantages and disadvantages of each, as well as further references to the literature. As noted by Abdellaoui et al. (2007), estimations in particular for the coefficient of loss aversion λ are further complicated by the lack of a commonly agreed definition in the literature. Most, however, relate to the ratio of marginal utility for losses over the marginal utility for gains for one or many outcomes and usually suggest that the utility function is between 1 and 3 times steeper for outcomes below the reference point (see Table 1 in Abdellaoui et al., 2007), with $\lambda \approx 2.25$ often cited as being a reasonable approximation (Wakker, 2010). However, loss aversion is not homogeneous among individuals and some studies suggest that those with certain characteristics will tend to be more averse to losses. For instance, Booij and Van de Kuilen (2009) find that women and those with lower levels of education are significantly more loss averse. Moreover, as noted by Wakker (2010), loss aversion is volatile and depends heavily on framing. Accordingly, $\lambda = 2.25$ should not be viewed as a universal constant which will prevail in any given situation. In fact, there is recent evidence to suggest that loss aversion is by no means a universal trait: Goette et al. (2018) classify only 36% of their subjects as loss averse, with the remainder being either loss neutral or loss loving.

While there is by now a large amount of evidence for each of the features of prospect theory, there have been surprisingly few applications to economics. Barberis (2013) argues that this is because it is difficult to know exactly *how* to apply the theory, since it is not ready-made for such applications. Firstly, in many situations it is often not clear how exactly the reference point should be determined. We shall return to this issue shortly. Secondly, application of the theory introduces unwelcome additional mathematical complexity to economic models, for instance due to the presence of both convexity and concavity, or due to the lack of differentiability at the reference point. Accordingly, many of these applications limit attention to specific features of the theory. For example, the incentive contracting literature which we review subsequently entirely ignores the issue of probability weighting. Instead, these applications focus solely on reference-dependence and loss aversion, with some additionally considering diminishing sensitivity.

The most basic possible form assumes a utility function which is piecewise

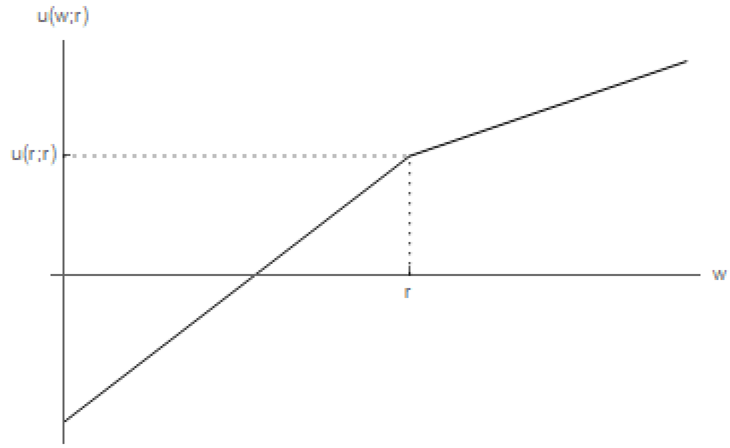

Figure 3: The function $u(w;r)$ as defined by (30), with $\lambda = 2.25$.

linear around the reference point:

$$u(w;r) = \begin{cases} \lambda w - (\lambda - 1)\, r & w < r \\ \\ w & w \geq r \end{cases} \tag{30}$$

as is plotted by Figure 3. While this specification still suffers from the lack of differentiability at r, it is everywhere (weakly) concave and the piecewise linear structure significantly reduces mathematical complexity. This form can be extended by further assuming classical risk aversion (see for instance de Meza and Webb, 2007), so that the function is everywhere strictly concave but still kinked at the reference point.

The fact that this approach ignores diminishing sensitivity should not necessarily be considered a limitation. Since convexity over the loss region is at odds with the concavity implied by the standard economic argument of diminishing marginal utility, such a function is not expected to describe preferences for losses that are large relative to total assets (Kahneman, 2003). Wakker et al. (2007) argue that a compromise in which utility for losses is mildly convex and closer to linear than for gains may be more appropriate. For sufficiently large gambles, however, the diminishing marginal utility effect will dominate, resulting in a concave utility function once more. It is unclear which functional form is more appropriate for the incentive contracting applications which we consider. While real-world incentive payments can indeed be significant, especially in the case of CEO compensation, such contracts do not typically feature outcomes for which the individual is left in ruin. Unfortunately, there is little empirical evidence to guide us here.[32]

[32] More generally, testing for risk preferences over high monetary stakes has proven difficult.

An alternative approach to modelling reference-dependent preferences involves separation of the utility function into two components. For example, Kőszegi and Rabin (2006, 2007) present a model in which the utility function has the form:

$$u(w;r) = m(w) + \eta\mu\left(m(w) - m(r)\right) \tag{31}$$

which is additively separable in two terms.[33] The first, $m(\cdot)$, represents the standard economic notion of intrinsic consumption utility and is therefore assumed to be everywhere increasing and weakly concave. The second, $\mu(\cdot)$, captures both reference-dependence and loss aversion. One possible simple formalisation is as follows:

$$\mu(x) = \begin{cases} \lambda x & x \leq 0 \\ \\ x & x > 0 \end{cases} \tag{32}$$

so that $\mu(\cdot)$ is piecewise linear.[34] The parameter η then measures the relative strength of gain-loss utility in the agent's preferences, with the standard model being captured as a special case when $\eta = 0$.

Their theory also allows for stochastic reference points; let R be a gamble with k outcomes and let each of these outcomes and their associated probabilities be denoted by r_i and q_i respectively for $i = \{1, ..., k\}$. The utility associated with the monetary outcome w is then:

$$u(w;R) = m(w) + \eta\sum_{i=1}^{k} q_i \cdot \mu\left(m(w) - m(r_i)\right) \tag{33}$$

Intuitively, if an individual's reference point is a gamble between £0 and £100, then actually receiving £50 feels like a gain relative to £0, but a loss relative to £100. The overall sensation is then a mix between these two feelings, weighted

Since most evidence is experimental in nature — and since researchers are limited in the money available to them — investigation of this issue is challenging. One approach is to perform experiments in developing countries so that the monetary stakes are large in real terms (see Kachelmeier et al., 1992), while an alternative is to use data from field experiments such as game shows (see Post et al., 2008). One noteworthy study comes from Pope and Schweitzer (2011), who use data from professional golf and find evidence suggesting that loss aversion persists with high stakes, intense competition and experienced agents.

[33]While Kőszegi and Rabin (2006) formalise a model of consumption across multiple dimensions, the simplified version presented here follows Kőszegi and Rabin (2007) by assuming that individuals have one-dimensional utility over monetary outcomes.

[34]Since the argument of $\mu(\cdot)$ depends on the function $m(\cdot)$, note that this formalisation entails a strong link between standard consumption utility and gain-loss utility. While we have imposed that $\mu(\cdot)$ is piecewise linear, Kőszegi and Rabin allow for more general forms, including those which feature diminishing sensitivity so that $\mu(\cdot)$ satisfies the same conditions as prospect theory's value function. In this case, whether $u(\cdot)$ is convex or concave over losses will depend on the trade-off between the functions $m(\cdot)$ and $\mu(\cdot)$. Most applications, however, assume this piecewise linear form.

by the appropriate probabilities. Larsen et al. (2004) provide evidence suggesting the existence of such mixed feelings in the context of expectations acting as a reference point.

Finally, an agent's ex ante expected utility for a risky gamble is then calculated as usual, by summing over ex post utilities weighted by their respective probabilities. Individuals are assumed to process these probabilities objectively, so that the Kőszegi and Rabin model does not feature the probability weighting component of prospect theory.

3.2.3 What Determines the Reference Point?

As noted by Barberis (2013), one of the main obstacles to applications of prospect theory relates to the determination of the reference point. In any given environment, it is often unclear what exactly constitutes the reference point and therefore whether a particular outcome should be defined as a loss or a gain. Kahneman and Tversky offered relatively little guidance here. In their original 1979 article, they state (p. 286) that:

> "So far in this paper ... the reference point was taken to be the status quo, or one's current assets. Although this is probably true for most choice problems, there are situations in which gains and losses are coded relative to an expectation or aspiration level that differs from the status quo."

Similarly, in Tversky and Kahneman (1991, pp. 1046-47), they write:

> "Although the reference state usually corresponds to the decision maker's current position, it can also be influenced by aspirations, expectations, norms, and social comparisons."

The question of what determines the reference point has been a major topic of discussion in the literature, especially since the existing evidence is scarce. While an individual's current wealth level (the status quo) seems appropriate for the simple experimental gambles which are often studied in the context of loss aversion, the determinants of reference points in more complex environments are much less obvious. Evidence from the psychology literature indicates that, as suggested by the above quotations, reference points may be influenced by various phenomena other than the status quo. While not discussed in detail here, goals (Heath et al., 1999), aspirations (Payne et al., 1980), social comparisons (Schwerter, 2016) and foregone alternatives (Mellers et al., 1997) have all been found to play a role.

In the specific context of wages, there is some evidence that an individual's existing wage acts as a reference point. For instance, several authors have documented changes in behaviour following pay cuts, which may be indicative of loss aversion with respect to the status quo wage. Lord and Hohenfeld (1979) analyse falls in the performance of professional baseball players who, due to a change in regulations, experience a substantial reduction of their salary, while Greenberg (1993) describes a field experiment in which employee theft significantly increased following a wage cut. In addition, reference points being determined by an individual's existing wage is also related to the habit formation literature (see for instance Bowman et al., 1999), one feature of which is an aversion to reductions in current consumption levels.

In the case of variable pay, it seems likely that in many cases the *base wage* becomes salient. Brink and Rankin (2013) assert that the base salary tends to be interpreted by employees as a guaranteed amount, while Luft (1994) goes further, suggesting that it may be viewed by workers almost as an *entitled* wage. However, de Meza and Webb (2007) argue that the reference point may be impacted by the wage schedule, so that it becomes determined endogenously by the contract. Suppose for instance that a manager's contract offers a base salary, with a 90% chance of receiving an additional bonus for good performance. In such a case, it seems likely that the manager is disappointed when receiving only the base salary, so that this particular outcome is perceived as a loss. This would occur if, for example, the manager's reference point was determined by the *mean* or the *certainty equivalent* of the income distribution implied by the contract. de Meza and Webb (2007) pursue this approach, but also note that neither formalisation seems applicable to all situations. Let us now suppose that the manager receives the bonus with a probability of only 1%, so that 99% of the time the base salary is paid. The latter outcome is unlikely to cause the manager too much disappointment, yet would be seen as a loss under either formulation. With this in mind, de Meza and Webb (2007) additionally consider the *median* as a reference point.

There is even evidence to suggest that in some environments individuals compare outcomes to multiple points of reference (Copeland and Cuccia, 2002; Koop and Johnson, 2012). For example, a worker's satisfaction with a particular wage payment may simultaneously depend on comparisons to last year's income, to other possible unrealised wage payments, to the worker's own goals and the wages of salient others such as coworkers. As is noted by O'Donoghue and Sprenger (Forthcoming), the development of models which incorporate multiple reference points has the potential to help explain various real-world phenomena and as such represents an important research agenda for the future.

The foregoing discussion emphasises that in any given environment, the reference point can often be defined in several different plausible ways. This has lead to a pervasive criticism of applications of reference-dependent preferences: without a consistent approach to reference point determination, models benefit from an extra degree of freedom which can potentially accommodate a range of different behaviours across various environments (Pesendorfer, 2006).

In recent years, some authors have therefore turned their attention to developing disciplined models with a consistent approach to defining the reference point. In one prominent example, Kőszegi and Rabin (2006, 2007) present a model of reference-dependent preferences in which the reference point is determined by an individual's expectations, whereby agents are assumed to be fully rational in the sense that they are able to perfectly predict both their environment and their behaviour in this environment.[35]

This approach has several advantages. First, expectations acting as a reference point has an intuitive appeal. For example, an individual who is expecting to receive a payment of £100 will likely feel dissatisfaction if they actually receive only £50. Yet, since current wealth has increased, the lesser payment would still be interpreted as a gain were the reference point to be defined by the status quo.

Second, Kőszegi and Rabin (2006) note that virtually all of the existing evidence which equates the reference point with the status quo comes from contexts where people plausibly expect to maintain the status quo, so that this evidence is also consistent with expectations determining the reference point.

Third, Kőszegi and Rabin (2006) go on to argue that there are several economic environments in which expectations differ from the status quo and, in addition, reference points determined by expectations generally make better predictions in such environments. Consider for instance a consumer who plans to purchase a ticket for a particular concert, but finds that tickets are sold out. While it is likely that the consumer experiences negative feelings as a result of this, his current endowment of concert tickets is identical to that of someone who did not plan to purchase a ticket, so that equating the reference point with the status quo predicts identical gain-loss utility of zero in these situations.

Finally, since gain-loss utility and the reference point can both be derived from consumption utility and the economic environment, their formalisation represents a step closer to a universally applicable, zero degrees of freedom theory.

The Kőszegi and Rabin model has received substantial attention in the recent literature and has been used to analyse various topics of economic importance, such as labour supply (Crawford and Meng, 2011), job search (DellaVigna et al., 2017), optimal provision of incentives (Herweg et al., 2010) and monopoly pricing (Heidhues and Kőszegi, 2014).[36] The model itself will be presented in detail shortly. However, we first pause to consider some of the evidence for expectations

[35]Earlier models of expectations-based reference-dependent preferences include Bell (1985), Loomes and Sugden (1986) and Gul (1991). See O'Donoghue and Sprenger (Forthcoming) for an overview.

[36]See O'Donoghue and Sprenger (Forthcoming) for a review.

acting as a reference point as outlined by their model, which is currently an ongoing discussion within the literature.

Empirical support has come from a variety of contexts: American Football games and domestic violence (Card and Dahl, 2011); casino slot machines (Lien and Zheng, 2015); the labour supply of taxi drivers (Crawford and Meng, 2011 and the literature therein); professional golfers (Pope and Schweitzer, 2011); police performance (Mas, 2006) and labour supply decisions in developing countries (Spears, 2012). There has also been experimental evidence, in contexts such as consumption decisions (Karle et al., 2015); the endowment effect (Knetsch and Wong, 2009); auctions (Banerji and Gupta, 2014) and risk preferences (Sprenger, 2015).[37] However, several other studies find evidence which is inconsistent with reference points being determined by expectations as outlined by the Kőszegi and Rabin model, such as Heffetz and List (2014), Smith (2012), März (2016), Goette et al. (Forthcoming) and Senn (2015).

There is some particularly relevant experimental evidence in the context of effort provision. Abeler et al. (2011) conduct a real effort experiment in which subjects are paid a piecerate for undertaking a tedious and repetitive task. After each repetition of the task, they decide whether to stop working or carry on. When they finally stop, their remuneration is decided in the following way: with equal probability they receive either their piecerate earnings, or a predetermined fixed payment instead. Varying the fixed payment between treatments allows for manipulation of participants' expectations. Classical economic theory suggests that effort provision should be independent of the fixed payment: agents should decide by trading off the marginal costs and benefits of an additional unit of effort. Similarly, theories which assume the status quo as a reference point do not predict differences in effort provision, since the status quo is not influenced by the magnitude of the fixed payment. In contrast, when expectations act as a reference point, agents are predicted to match their piecerate earnings with the fixed payment in order to minimise the scale of potential losses, so that effort provision increases as this payment is raised. The authors find evidence consistent with both of these predictions.

A similar experiment comes from Gill and Prowse (2012), who study competition between participants in a two-player sequential tournament featuring a *leader* and a *responder*. The game is designed such that the marginal impact of the responder's effort on their probability of winning does not depend on the effort provision of the leader; accordingly, the classical theory predicts that the responder's effort should be invariant to how hard the leader worked. However, the authors find evidence of a *discouragement* effect, whereby the responder shies

[37] See also Ericson and Fuster (2011), Fehr and Goette (2007) and Song (2016).

away from working when the leader's effort provision is high, but works hard when it is low. They argue that this is consistent with reference points determined by expected payoffs which adjust instantaneously to effort choices.

There has also been conflicting evidence. Gneezy et al. (2017) are able to replicate the findings of Abeler et al. (2011), discussed above, by showing that increasing the fixed payment results in higher effort provision. However, they also find that decreasing the fixed payment to zero also causes subjects to work harder, so that effort provision is non-monotonic in payments; this is at contrast with the predictions of the theory. They also find similar violations when manipulating probabilities, suggesting that fixed payments and expectations influence reference points in more complex ways than the theory predicts.

As such, at the present moment, overall evidence for expectations acting as reference points as predicted by the Kőszegi and Rabin model is mixed, with some authors turning their attention to possible explanations for this. For instance, Goette et al. (2018) propose that heterogeneity in loss aversion may be responsible. They gather experimental evidence which at the aggregate level shows no support for the predictions of models of expectation-based loss aversion, but does so once the authors control for variations in attitudes towards loss among subjects. Alternatively, Heffetz (2018) offers a possible explanation relating to the nature of how expectations adjust. Future research will continue to shed light on this issue, allowing for improvements in the understanding of how best to apply expectations-based models of reference-dependent preferences.

Finally, to conclude our discussion of how reference points may be determined, we briefly turn our attention to a related question. Even in situations where we can identify a single plausible reference point, if the environment is dynamic we must also consider if and how it may change over time. While evidence here is also scarce, there is a small experimental literature which attempts to explore this issue in the context of financial decision making. Gneezy (2005) finds evidence that in an environment where the price of an asset changes over time, individuals tend to focus on the peak price, which then acts as a reference point. This is in contrast to Baucells et al. (2011), whose findings suggest that it is the *initial* and *most recent* prices which are most salient, with intermediate prices being less influential. In a similar setting, Arkes et al. (2008) find evidence to suggest that the reference point adjusts asymmetrically: the upward movement after a gain is larger in magnitude than the downward movement after an equal sized loss. The authors suggest that this may be due to hedonic considerations, whereby individuals adjust their reference points in such a way as to improve their overall feelings about a situation. A subsequent study finds similar results, but also evidence to suggest that there are cross-cultural differences in the adjustment process (Arkes et al., 2010).

In the expectations-based approach, the relevant question relates to the speed of adjustment of the recently held beliefs which determine the reference point. Studies by Gill and Prowse (2012), Smith (2012) and Song (2016) suggest that there is a quick or even instantaneous adjustment, while conflicting evidence comes from Post et al. (2008) and Card and Dahl (2011). Heffetz (2018, p.5) argues that it is not time, but accustomisation, which is important:

> "Under our hypothesis, what moves the reference point is not the passage of time per se, but some sense of internalization of, or getting used to, the new expectations—which we refer to as sink-in. It is not inconceivable that sink-in takes as little as a few minutes in a low-stakes, no-prior-expectations lab experiment, but takes much longer in a higher-stakes, long-held-expectations setting."

Finally, there is evidence that individuals tend to underestimate the extent to which their preferences will change over time (Loewenstein et al., 2003). This so-called *projection bias* would then suggest that individuals are unable to perfectly anticipate adjustments in the reference point, leading to suboptimal decision making (Kőszegi and Rabin, 2006).

3.2.4 The Kőszegi and Rabin Model of Expectations-Based Reference-Dependent Preferences

The Kőszegi and Rabin (2006, 2007) model of reference-dependent preferences assumes that individuals evaluate outcomes relative to their expectations. More specifically, the reference point is defined as the agent's probabilistic beliefs held in the *recent* past about outcomes. Intuitively, consider an individual who learns that they will not receive a long-expected payment five minutes before it was due to be paid. While their expectations will adjust immediately, it is likely that five minutes later they will still assess receiving zero money as a loss; accordingly, preferences are dependent on *lagged* rather than current expectations. Since in many environments expectations are likely to feature uncertainty, the reference point will often be stochastic, with the individual's utility function taking the form outlined by (33).

The model also requires a description of how expectations themselves are formed. Kőszegi and Rabin assume that individuals are extremely rational in this regard: they perfectly anticipate both the environment that they face and their own behaviour in this environment, with beliefs then reflecting the true probability distribution of outcomes. They argue that this captures, albeit in an extreme way, the plausible notion that individuals have some ability to predict their own future behaviour. For the purpose of applications, the model offers two distinct solution concepts depending on the temporal proximity between decisions being made and outcomes being realised.

First, when decisions are made shortly before outcomes are realised, the reference point will be fixed by past expectations and as such cannot be influenced by the individual's choice. Accordingly, the individual maximises utility taking the reference point as given, with the assumption of perfect rationality requiring that agents can only expect a certain decision if they are willing to follow through with it, given a reference point generated by the expectation to do so. This is known as an *unacclimating personal equilibrium* (UPE), in which the utility maximising choice given certain expectations actually induces these same expectations. Typically, in any given environment, multiple UPEs may exist. As an equilibrium selection device, Kőszegi and Rabin argue that since an individual can make any plan that is rational, in the sense that they know they will follow it through, agents will choose their preferred UPE, or the one which yields highest ex ante expected utility.[38] It is important to note however, that a UPE will often not maximise ex ante expected utility among choices available to the individual, as expectations are taken as given and not internalised when undertaking the decision.

Second, when outcomes are realised a long time after all decisions have been made, the expectations relative to which outcomes are evaluated will be formed after these decisions and will therefore incorporate their implications. For these cases, the appropriate solution concept is a *choice-acclimating personal equilibrium* (CPE): a decision which maximises expected utility given that it determines both the outcome and the expectations which dictate the reference point. In contrast to a UPE, a CPE will always maximise ex ante expected utility, since the effects of decisions on expectations are internalised. CPE seems appropriate for the incentive contracting applications we later consider, where the outcome of the performance measure is often realised some time after effort has been undertaken.

In order to aid our discussion of the Kőszegi and Rabin model, we consider the following simple example. Suppose an individual faces a gamble between receiving a high payment w_H with probability p and a low payment $w_L < w_H$ with probability $1 - p$. This gamble, which we denote by $\overline{W}$, then acts as a stochastic reference point for the individual. Suppose for simplicity that the agent

[38] This is because there may be multiple self-fulfilling expectations. For instance, consider a consumer who formulates a purchasing plan in the presence of price uncertainty and ex ante expects to purchase a good when faced with a price £x. In this case, when they are actually faced with £x, they may indeed prefer to follow through on their plan and make the purchase. However, it is also possible that in the alternative case, where they had initially expected not to buy when faced with the price £x, that similarly they would again prefer to follow through on these expectations and decline to purchase. Accordingly, both scenarios are equilibria in this environment. PPE then predicts that the actual decision will be the UPE which maximises ex ante expected utility.

is risk neutral so that $m(w) = w$, while $\mu(\cdot)$ is given by (32). Then ex post utility when the actual outcomes are w_H and w_L is respectively given by the following equations:

$$u(w_H;\overline{W}) = w_H + \eta p \underbrace{(w_H - w_H)}_{=0} + \eta (1-p)(w_H - w_L) \tag{34}$$

$$u(w_L;\overline{W}) = w_L + \eta\lambda p (w_L - w_H) + \eta (1-p) \underbrace{(w_L - w_L)}_{=0} \tag{35}$$

The first term in (34) represents consumption utility, while the remaining terms capture gain-loss utility: the actual outcome w_H is compared with the expected outcomes w_H and w_L, weighted by their respective ex ante expected probabilities p and $(1-p)$. Since the actual outcome coincides with the expected outcome in the former case, this particular comparison induces gain-loss utility of zero so that the second term of (34) vanishes. The utility associated with the outcome w_L, (35), can be explained in a similar way. However, the key difference is that the comparison of the actual outcome w_L with the expected outcome w_H corresponds to a loss for the agent and is weighted by the parameter λ accordingly.

Calculating ex ante expected utility then yields:

$$U(\overline{W};\overline{W}) = p \cdot u(w_H;\overline{W}) + (1-p) \cdot u(w_L;\overline{W}) \tag{36}$$

$$= p\, w_H + (1-p)\, w_L - \eta (\lambda - 1)\, p\, (1-p)\, (w_H - w_L) \tag{37}$$

The first two terms in (37) represent expected utility from consumption as usual; in this case, the expected value of the gamble. The third term captures gain-loss utility. Let us pause to make some remarks. First, note that since loss aversion implies $\lambda > 1$, the addition of gain-loss utility entails a reduction of expected utility when outcomes are uncertain, even in the case of risk neutral standard consumption utility. Second, fixing outcomes at w_H and w_L, the utility loss from uncertainty is weighted by the term $p(1-p)$, which is inverse U-shaped in p and attains a maximum when $p = 0.5$. Third, as $\eta(\lambda - 1)$ becomes large, the individual's main concern becomes avoiding potential losses. As a result, they may obtain higher utility from a gamble where p is low compared to a gamble where $p \approx 0.5$. That is, under this preference specification individuals may prefer a stochastically dominated gamble, since raising expectations of a gain makes an outcome of no gain feel more painful. Under CPE, since the individual's choice also influences expectations, this implies that in some cases they would in fact *choose*, when offered two gambles, the one which is stochastically dominated. Kőszegi and Rabin (2007) argue that this is not necessarily a weakness of the model and may be consistent with real-world preferences in some environments, citing Frederick and Loewenstein's (1999) discussion of a prisoner who is made worse off by a negligible chance of early release. In contrast, this type of behaviour cannot occur under UPE, since the expectations which determine the reference point are fixed.

In order to now consider the behaviour of individuals, we extend the environment by assuming that the probability of receiving the high payment is now dependent on a costly action $a \in \mathbb{R}_+$, so that $p = p(a)$ with $p' > 0$, $p'' < 0$ and $\lim_{a\to\infty} p(a) = 1$. Moreover, let the cost of the agent's action be denoted by the increasing convex function $c(a)$, with $c(0) = c'(0) = 0$.[39] One interpretation of this environment is an agent's choice of effort when faced with a simple incentive contract, similar to those discussed in Section 2.

We first consider the case where the individual's choice of a is made shortly before the uncertainty is resolved, so that the reference point is fixed by past expectations. These expectations are defined uniquely by the expected action, which we denote by $\tilde{a}$. The individual's problem of maximising utility then becomes:

$$\max_a \; p(a)\left[w_H + \eta(1 - p(\tilde{a}))\,(w_H - w_L)\right] + (1 - p(a))\left[w_L + \eta\lambda p(\tilde{a})\,(w_L - w_H)\right] - c(a) \quad (38)$$

The terms in square brackets represent the ex post utility associated with receiving w_H and w_L respectively, and comparing the relevant outcome with the expectations induced by $\tilde{a}$; to see this, compare these terms with (34) and (35). Taking the derivative and rearranging, the optimal effort level is implicitly defined by the following equation:

$$p'(a)\left[w_H - w_L + \eta\left[1 - p(\tilde{a}) + \lambda p(\tilde{a})\right](w_H - w_L)\right] = c'(a) \quad (39)$$

By also requiring that $\tilde{a} = a$, a UPE in this environment is then characterised by the following condition:

$$p'(a)\left[w_H - w_L + \eta\left[1 - p(a) + \lambda p(a)\right](w_H - w_L)\right] = c'(a) \quad (40)$$

Next, consider the case where the individual commits to a certain action long before the uncertainty is resolved. Since the choice of action then also determines

[39] Strictly speaking, the introduction of a costly action undertaken by the agent should induce additional gain-loss utility, depending on whether effort costs are higher or lower than expected. For simplicity, however, we restrict attention here to reference-dependence over monetary outcomes only and ignore this additional dimension.

the reference point, the relevant utility maximisation problem becomes:

$$\max_{a} \; p(a)\left[w_H + \eta\left[1 - p(a)\right](w_H - w_L)\right] \\ + \left[1 - p(a)\right]\left[w_L + \eta\lambda p(a)(w_L - w_H)\right] - c(a) \quad (41)$$

Taking the derivative and rearranging, a CPE is defined by:

$$p'(a)\left[w_H - w_L + \eta(1 - \lambda)\left[1 - 2p(a)\right](w_H - w_L)\right] = c'(a) \quad (42)$$

Note that decision making will typically vary under the different solution concepts of UPE and CPE; this can be seen in our example by comparing (40) and (42).

By defining the reference point as an individual's expectations, combined with the restriction that beliefs must be formed rationally, the Kőszegi and Rabin model provides a consistent and disciplined approach to modelling reference-dependent preferences. This goes some way to addressing the aforementioned criticism relating to the extra degree of freedom which such models often benefit from, since, given a full specification of $\mu(\cdot)$, both the reference point and gain-loss utility follow directly from consumption utility and the economic environment. However, Kőszegi and Rabin (2006) caution that some judgement is still required when applying the model. Indeed, O'Donoghue and Sprenger (Forthcoming) note that the existence of multiple solution concepts provides some freedom which can be exploited, with CPE being used more often in applications due to its tractability even in situations where it may not be appropriate.

Finally, Kőszegi and Rabin (2009) develop a similar model of reference-dependent preferences which allows for *anticipatory utility*, whereby individuals derive pleasure from anticipating future consumption. An individual's utility function is assumed to be made up of three components: *i)* standard consumption utility, *ii) contemporaneous* gain-loss utility, derived from differences between current consumption and recent prior expectations of current consumption and *iii) prospective* gain-loss utility, resulting from changes in current beliefs over future consumption. Two properties of prospective gain-loss utility are particularly noteworthy. First, individuals are loss averse in comparisons so that bad news about future consumption is more painful than good news is pleasant. Second, individuals are more sensitive to changes in beliefs about imminent outcomes compared to distant outcomes. As before, beliefs are required to be formed rationally based on credible plans for future behaviour. The model seems particularly relevant for applications to multi-period incentive contracting, a direction which is pursued by Macera (2018*a*).

3.3 Discussion

There is by now a vast body of evidence, from both the laboratory and the field, that the preferences of individuals often exhibit both inequity aversion and reference-dependence. Additionally, incorporating these concerns into economic applications has improved the accuracy of predictions and lead to many valuable insights.

So far, inequity aversion and reference-dependence have been presented as distinct concepts. However, they share certain similarities: in both cases outcomes are compared with referents (relating to either a reference *group* or a reference *point*) and in both cases outcomes below the referent lead to relatively larger impacts on utility. These common features can lead to similarities in the utility functions which are used to represent such preferences and, as we shall see in the next section, also to comparable implications for economic applications. This is especially true when a loss averse individual's reference point is influenced by social considerations, which is natural in many environments.[40]

In fact, the two concepts are very closely related. Studies of the behaviour of capuchin monkeys have provided evidence for both inequity aversion (Brosnan and De Waal, 2003) and loss aversion (Chen et al., 2006), which suggests that such preferences have a biological component and are (at least to some degree) innate to humans. Chen and Santos (2006) propose that the evolutionary origins of inequity aversion, in both humans and other species, may lie in tendencies to make comparisons with social reference points. An organism which observes the payoffs of other individuals living in the same environment can gather valuable information, beyond that which is amassed from simply focussing on its own experiences. That is, even in the absence of any social interaction, attending to the payoffs of others may confer selective advantages. For example, in foraging animals, the

[40] Schwerter (2016) provides experimental evidence for social reference points when individuals undertake risky decisions. Each lab session consists of two subjects who are randomly assigned to one of two roles: one acts as a peer, while the other is the decision maker. Peers received a fixed payment, while the decision maker chooses their preferred binary lottery from a menu of choices. Each lottery has a downside of zero; decision makers then face a trade-off between the size of a lottery's upside and the likelihood with which this upside is received. The fixed payment of peers is varied between treatments. The main result is that — as predicted by a theory of loss aversion around social reference points — subjects choose riskier lotteries when they observe a high fixed payment for the peer in an attempt to avoid an outcome in which they earn less. Further evidence then suggests that this result cannot be explained by non-social reference points such as expectations. In a similar setup, Linde and Sonnemans (2012) find no evidence for the presence of diminishing sensitivity with a social reference point (see also Vendrik and Woltjer, 2007). These studies are part of a growing literature which analyses risk preferences in social contexts; see for instance Gamba et al. (2017) and Müller and Rau (2017) for recent contributions. Real-world evidence for socially determined reference points comes from Kuhn et al. (2011), who study the consumption behaviour of the neighbours of Dutch Postcode Lottery winners and find that living next to a winner significantly increases the probability of households purchasing a new car over the next six months.

relative availability of food will be reflected in the payoffs of others, so that peer observations are informative regarding the returns to foraging effort in the current environment. As a cognitive mechanism, an animal which experiences envy will then be further driven to increase foraging effort when observing the high payoffs of peers; i.e. in exactly those situations where increases in this effort are likely to yield high returns. Clearly, this provides an evolutionary advantage by increasing foraging efficiency. Similarly, Rayo and Becker (2007) employ approaches from economic theory to argue that social comparisons may have been evolutionarily advantageous. In a principal-agent framework, they imagine a principal (representing the process of natural selection) who must design an agent's emotional responses to outcomes in such a way as to maximise fitness and show that peer comparisons are a feature of the optimal mechanism.[41]

However, on a psychological level there is a clear distinction between the two concepts. While inequity aversion is often discussed in the context of fairness and having an actual concern for the payoffs of others, an individual who exhibits reference-dependent preferences with respect to a social reference point does not care about the payoff of others per se; instead, they merely use this information to evaluate their own payoff.

Despite this distinction, both of these concepts can lead to similar behaviour, so that it is unclear which mechanism is at work in any given situation. For instance, Raihani and McAuliffe (2012) provide a discussion of the cognitive process behind an individual's decision to punish cheats in social interactions. Observations of cheating often inspire negative emotions such as anger or disgust, with neurological evidence suggesting that it is these emotions which underpin the decision to punish rather than cognitive processes; the administration of punishment may negate these negative feelings by activating brain reward centres (see Raihani and McAuliffe, 2012 for further discussion and references). However, the source of the negative emotions is unclear. While they could arise from fairness concerns relating to inequity aversion, another alternative is that the cheating behaviour leads an individual's payoff to fall below some reference point, resulting in dissatisfaction.

Finally, as mentioned previously, economic models which incorporate these concepts into the preferences of individuals often share several similarities. For instance, compare the Fehr and Schmidt model of preferences in a two agent environment (27) with the simple model of loss aversion (30). Both functions are piecewise linear around a kink point; by setting the reference point r in (30) equal to agent j's payoff x_j and appropriately adjusting the remaining parameters, the

[41] Along similar lines, Falk and Knell (2004) present a theoretical model, along with supporting evidence, in which individuals have relative income concerns and *choose* their relevant social reference group in such a way as to promote self-improvement and self-enhancement.

two functions become equal. Similarly, slightly adjusting the Kőszegi and Rabin model of reference-dependent preferences with a stochastic reference point (33) by equating each r_i with another agent's payoff x_j and setting all weights q_i equal to $\frac{1}{N-1}$ yields — given risk neutral consumption utility and piecewise linear gain-loss utility — the general Fehr and Schmidt model for N agents, (28). Accordingly, it is not surprising that in many economic applications, inequity aversion and reference-dependence can lead to similar or identical outcomes.

While the incorporation of these features of individual preferences into economic analysis has already enhanced our understanding of a wide range of phenomena, there are still many open questions relating to both inequity aversion and reference-dependent preferences which require future research. Some of these — such as the composition of the reference group or the nature of reference points — we have touched upon in this section.[42] In particular, an understudied area relates to the interplay between the two concepts. Recent developments, such as the aforementioned investigations into risk preferences in social contexts, will help clarify this and in turn improve economic applications of these models.

4 Incentive Contracting with Inequity Aversion and Loss Aversion

4.1 Inequity Aversion

4.1.1 Independent Contracting

To begin our exploration of how incentive contracting is impacted by inequity averse preferences, we initially consider a simple environment with continuous efforts and binary performance measures, so that $\mathcal{A} = \mathbb{R}_+$ and $\mathcal{X} = \{x_L, x_H\}$. The model presented here is based primarily on those of Itoh (2004) and Demougin and Fluet (2006).[43] A principal employs two homogeneous agents, denoted 1 and 2, to undertake identical tasks. The agents are inequity averse à la Fehr and Schmidt and compare their wages with one another, but not the principal, so that their preferences can be represented by the utility function (27).[44] We normalise

[42] Some further possible future directions are discussed in the conclusion to this chapter.

[43] While these papers share a similar theme — both examine incentive contracting with independent and team contracts when agents are other-regarding and performance measures are binary — there are a number of differences. The model of Demougin and Fluet (2006) features continuous effort, while attention is restricted to envious agents who compare net wages; that is, individuals include effort costs when making comparisons. In contrast, Itoh (2004) models effort as binary, while agents compare wages only. Envious, compassionate and status-seeking preferences are all permitted in his model.

[44] Alternative specifications include agents who compare net wages (i.e. wages minus effort costs) with one another, or agents who compare their wage to the principal's income. We discuss both of these cases later in this subsection.

the reservation utility of each agent to zero.[45] It is assumed that the principal has access to two performance measures — one for each agent — and that these signals are independent of one another. That is, the probability of observing a signal of high performance for agent i does not depend on the effort provision, nor the performance signal, of agent j. Moreover, we will initially assume that the principal offers *independent* contracts, such that wage payments to agent i do not depend on the signal of agent j's performance and that the same contract is offered to each agent. Accordingly, we analyse the principal's problem of designing the cost minimising contract to implement a given effort level a for Agent 1, given that in equilibrium an identical contract will be offered to Agent 2.

As in our previous examples, we let $p(\cdot)$ denote the probability of observing a signal of high performance conditional on effort, so that the principal's problem is:

$$\min_{w_H, w_L} \; p(a)w_H + [1 - p(a)]\, w_L \tag{43}$$

subject to incentive compatibility and participation constraints as usual. Denoting the effort levels of Agent 1 and Agent 2 by a_1 and a_2 respectively, the expected utility of Agent 1 is given by:

$$p(a_1)w_H + [1 - p(a_1)]\, w_L - p(a_1)\,[1 - p(a_2)]\,\beta\,(w_H - w_L) \\ - [1 - p(a_1)]\, p(a_2)\alpha\,(w_H - w_L) - c(a_1) \tag{44}$$

The first two terms of (44) represent the agent's expected consumption utility, while the final term captures the agent's disutility from undertaking effort as usual. The expected utility impact of inequity is then captured by terms three and four. Intuitively, wage inequality occurs when the principal observes different signal realisations for each agent, which results in a wage spread equal to $w_H - w_L$. Term three captures inequity concerns when Agent 1 receives a higher wage, which occurs with probability $p(a_1)\,[1 - p(a_2)]$; term four captures the reverse case where Agent 2 receives a higher wage, the probability of which is $[1 - p(a_1)]\, p(a_2)$.

Incentive compatibility requires that a is the argument that maximises Agent 1's expected utility. We take the first-order approach and differentiate (44) with respect to a_1, setting the result equal to zero. This yields:

$$p'(a_1)\,(w_H - w_L) - p'(a_1)\,(w_H - w_L)\,[[1 - p(a_2)]\,\beta - p(a_2)\alpha] \\ - c'(a_1) = 0 \tag{45}$$

[45] As previously discussed, this assumption is not without loss of generality since it implies agents no longer undertake social comparisons when not employed by the firm. See footnote 26.

In equilibrium, since wage schemes are incentive compatible and identical, both agents will undertake effort a. Substituting and rearranging yields the wage spread required to induce effort a:

$$w_H - w_L = \frac{c'(a)}{p'(a)} \cdot \frac{1}{1 + p(a)\alpha - [1 - p(a)]\,\beta} \tag{46}$$

This allows us to make some observations. First, note that the RHS of (46) is decreasing in the coefficient of envy, α. Intuitively, the principal creates effort incentives by setting wages as to generate differences in (expected) utility between high and low realisations of the performance measure.[46] Envious individuals experience disutility from receiving lower wages than others. In our model, this is only possible when an agent's performance is low; accordingly, envy reduces the expected utility associated with a low signal of performance, which provides additional effort incentives to the agent. This allows for lower explicit incentives and a reduction in the necessary wage spread.

Second, the RHS of (46) is increasing in the coefficient of compassion, β. The intuition here is similar to the foregoing case. Since compassionate individuals suffer from receiving a higher wage than others, the expected utility associated with a high performance signal is reduced, which weakens effort incentives and therefore a higher wage spread is required. Alternatively, as status-seeking agents enjoy receiving higher wages than others, this expected utility is increased and the necessary wage spread is lower. These results can be summarised by the following remark.

Remark. Envious ($\alpha > 0$) and status seeking ($\beta < 0$) preferences create a positive *incentive effect*, leading to a lower wage spread required in order to implement effort a. For compassionate ($\beta > 0$) preferences, the incentive effect is negative.

When agents are both envious and compassionate, the overall sign of this incentive effect will depend on the magnitude of both parameters as well as the relative probabilities of high and low performance being observed. Finally, the absence of any inequity concerns ($\alpha = \beta = 0$) yields the wage spread of the standard self-interested case, (12).

We now move onto the agent's participation constraint, which can be written, using (44) along with $a_1 = a_2 = a$, as:

$$p(a)w_H + [1 - p(a)]\,w_L - p(a)\,[1 - p(a)]\,(w_H - w_L)\,(\alpha + \beta) - c(a) \geq 0 \tag{47}$$

[46] Note that, for a given contract, the utility associated with a specific realisation of agent i's performance measure is stochastic when agents are other-regarding, since it also depends on the performance signal for agent j. Accordingly, we discuss the *expected* utility associated with high and low realisations of the performance measure for agent i.

where the third term measures the utility impact of inequity concerns; otherwise, the constraint is identical to the self-interested case, as shown by (11). First, we consider the case where there are no restrictions on wage payments, so that the participation constraint is binding. Using (46) and rearranging (47) then yields:

$$w_L = c(a) - p(a)\frac{c'(a)}{p'(a)} \cdot \frac{1 - [1-p(a)](\alpha+\beta)}{1 + p(a)\alpha - [1-p(a)]\beta} \tag{48}$$

so that the principal's costs become:

$$C^P(a) = c(a) + [1-p(a)]\,p(a)\frac{c'(a)}{p'(a)} \cdot \frac{(\alpha+\beta)}{1 + p(a)\alpha - [1-p(a)]\beta} \tag{49}$$

Since the principal faces no restrictions on wage payments, w_L will be set such that the agent extracts no rent from the relationship. As standard, the agent must be compensated for both his effort costs $c(a)$ and his outside utility, in this case zero. However, in order to accept the contract, an other-regarding agent must also receive compensation for exposure to potential inequity: we refer to the second term of (49) as an *inequity premium*. Since both envy and compassion result in disutility for agents, in the case of a binding participation constraint the principal's costs are therefore increasing in both α and β.[47] However, since a status-seeking agent experiences higher utility, the principal's costs will be reduced relative to the self-interested case. Summarising:

Remark. Envious ($\alpha > 0$) and compassionate ($\beta > 0$) preferences create a negative *participation effect*, which increases the principal's overall costs of implementing effort when the participation constraint is binding. Status seeking ($\beta < 0$) preferences, on the other hand, create a positive participation effect.

Second, we consider the case where wages are restricted to be non-negative. In this case, if the participation constraint does not bind then the principal will set $w_L = 0$ so that her costs can be expressed as:

$$C^P(a) = p(a)\frac{c'(a)}{p'(a)} \cdot \frac{1}{1 + p(a)\alpha - [1-p(a)]\beta} \tag{50}$$

which is decreasing in α but increasing in β. Intuitively, since the limited liability constraints imply that the agent will extract a positive rent, the participation constraint plays no role. As such, there is no participation effect of other-regarding preferences in this case and the impact on the principal's costs is determined solely by the incentive effect. Overall, the following remark summarises the effects of

[47] This is clearly true for β. For α, the result follows from:

$$\frac{\partial}{\partial\alpha}\left[\frac{(\alpha+\beta)}{1 + p(a)\alpha - (1-p(a))\beta}\right] = \frac{1-\beta}{[1 + p(a)\alpha - (1-p(a))\beta]^2} > 0$$

other-regarding preferences on the principal's costs of implementing effort:

Remark. Other-regarding preferences on the part of the agent have the following impact on the principal's costs:

- Compassionate ($\beta > 0$) preferences are always detrimental to the principal. They induce a negative incentive effect, which requires a higher wage spread in order to induce a given effort level a. In the case of limited liability, where $w_L = 0$, this implies that a higher w_H is necessary. Moreover, compassionate preferences induce a negative participation effect since the agent's expected utility is reduced. When wages are not restricted and the participation constraint binds, w_L must then be increased in order to guarantee that the agent accepts the contract. In either case, the principal's costs increase.

- Status-seeking ($\beta < 0$) preferences always benefit the principal. Both the incentive effect and the participation effect are positive, so that the principal's costs decrease regardless of whether the participation constraint binds.

- The effect of envious ($\alpha > 0$) preferences depends on the specifics of the economic environment. Since there is a positive incentive effect, the required wage spread to induce effort is reduced. When limited liability restricts the wages available to the principal so that $w_L = 0$, the high payment w_H can be lowered, leading to a decrease in the principal's costs. However, when the participation constraint is binding, the principal must compensate the agent for the disutility associated with envy, leading to a negative participation effect and an increase in the principal's costs.

The majority of the literature on incentive contracting with inequity averse agents centres around the incentive and participation effects of other-regarding preferences described above. In particular, many studies examine how contracts can be designed in order to eliminate — or generate — unequal wages between agents, to the principal's advantage. These will be examined shortly. It should also be noted that several papers limit attention to agents who are purely envious. There are two key reasons for this. First, as discussed in Section 3.1, it is typical to assume that $|\alpha| \geq |\beta|$ so that disadvantageous inequity concerns are more important to agents and as such envy is the driving force of preferences even when allowing for compassion or status-seeking. Second, as outlined in the above remark, envy has particularly interesting implications for contracting since the incentive and participation effects have opposing signs.

Before moving onto more complex contractual structures, we briefly consider some analyses of independent incentive contracting in the presence of other-regarding agents. Neilson and Stowe (2010) consider an environment where performance measures take the form $x_i = a_i + \epsilon_i$ and restrict attention to linear contracts. They first analyse the incentive effect and derive conditions for the existence of a symmetric equilibrium in which other-regarding agents undertake higher effort for a given piecerate, relative to the self-interested case. Next, they consider the participation effect and show that the required inequity premium is increasing in the piecerate, since this magnifies expected inequality between agents. Accordingly, when deciding on the optimal piecerate (and therefore the effort to be induced), the principal faces a trade-off between providing incentives and minimising the necessary inequity premium. Note that this is similar to the classical trade-off between incentives and insurance in the case of risk aversion. As a result, the optimal piecerate can be lower than the self-interested case; the authors suggest that this may provide an explanation for observed wage compression in firms.

Bartling and von Siemens (2010*b*) restrict attention to envy, but allow for risk aversion and an arbitrarily informative performance measure to consider how the effects of other-regarding preferences depend on the specifics of the economic environment. They consider the principal's cost minimisation problem for implementing a given effort level and show that this depends on envy in two particular ways. First, for a given interval of possible wages, envy influences the set of utilities which the principal can impose on the agent. More specifically, envy decreases the lower bound of this set and therefore increases the principal's ability to punish the agent for low performance. However, note that this positive incentive effect only plays a role in the presence of limited liability constraints; if the principal's choice of wage scheme is unrestricted, then it is possible to impose any utility level on the agent. Accordingly, envy can only reduce the principal's costs in environments where the set of feasible wages is restricted. Second, the principal's costs of providing an agent with a certain utility level are increasing in envy, since the agent dislikes the possibility of receiving lower wages than his coworkers. Bartling and von Siemens (2010*b*) are therefore able to replicate the foregoing results regarding the positive incentive and negative participation effects of envy in a much more general setting.[48]

4.1.2 Team Contracting

We have seen that when we limit our attention to independent incentive contracts, the introduction of other-regarding preferences on the part of agents can lead to an increase in the principal's costs. A natural question then arises as to whether the

[48]While Bartling and von Siemens (2010*b*) do not explicitly model compassionate preferences, they argue that the effects on the principal's costs will be unambiguously negative. Compassion will reduce the upper bound of the set of utilities which can be provided for a given set of wages, as well as increase the principal's costs of imposing a given effort level. Neither of these effects can be beneficial.

principal may prefer to use some sort of team contract, whereby the remuneration of an agent is dependent on the performance measure of others.

In order to consider this possibility, we extend the foregoing model to the case where the principal offers an *extreme team contract*, in which the wage w_H is paid if and only if a high signal of performance is observed for both agents. In all other cases, the wage w_L is paid.[49] As before, we analyse the principal's problem of designing the cost minimising contract to implement a given effort level a for Agent 1, again assuming that Agent 2 will be offered the same contract.

The extreme team contract implies that Agent 1 will only be paid the high wage w_H if the signal x_H is observed for both himself and Agent 2. This occurs with probability $p(a_1)p(a_2)$, so that the principal's problem becomes:

$$\min_{w_H, w_L} \; p(a)p(a)w_H + [1 - p(a)p(a)]\, w_L \tag{51}$$

subject to incentive compatibility and participation constraints. Agent 1's expected utility, conditional on the effort levels a_1 and a_2, can now be expressed as:

$$p(a_1)p(a_2)w_H + [1 - p(a_1)p(a_2)]\, w_L - c(a_1) \tag{52}$$

The most striking feature of (52) is that inequity concerns play no role. This is because the extreme team contract entirely shields agents from all wage inequality; both agents receive w_L for all possible realisations of their performance measures, except when x_H is observed for both agents, in which case they each receive the payment w_H. Accordingly, there is neither an incentive nor a participation effect when the extreme team contract is used. The incentive compatibility constraint requires that:

$$p'(a_1)p(a_2)w_H - p'(a_1)p(a_2)w_L - c'(a_1) = 0 \tag{53}$$

Since in equilibrium both agents will undertake effort a, the necessary wage spread is therefore:

$$w_H - w_L = \frac{c'(a)}{p'(a)p(a)} \tag{54}$$

Note that this is larger than the case of independent contracting in the absence of inequity concerns, (12). Intuitively, the probability of Agent 1 receiving the high wage payment is now less responsive to his effort input (since it also depends on the signal for Agent 2) and therefore the wage spread required to induce effort a

[49] Loosely speaking, any wage scheme in which one agent's payment is increasing in the performance of others can be thought of as a team contract. However, we will restrict our formal analysis to the case of the extreme team contract.

becomes higher. Inserting (54) into the participation constraint then yields:

$$w_L + p(a)\frac{c'(a)}{p'(a)} - c(a) \geq 0 \tag{55}$$

As before, we consider the cases of a binding and non-binding constraint in turn. First, in the absence of limited liability constraints, (55) holds with equality which yields:

$$w_L = c(a) - p(a)\frac{c'(a)}{p'(a)} \tag{56}$$

This implies that the principal's costs become once again equal to the agent's effort costs:

$$C^P(a) = c(a) \tag{57}$$

The participation effect of other-regarding preferences can therefore be nullified entirely when the principal is able to use a team contract which does not allow for wage inequalities. When either envy or compassion are the driving force of an agent's inequity preferences (i.e. when status-seeking plays a relatively small role), this implies that the principal's costs in the absence of limited liability constraints are lower when using the extreme team contract (57) than when using an independent contract (49). This represents a violation of the sufficient statistic result introduced in Section 2.4. Even though the performance measures of other agents are not informative about an agent's effort choice, they should nonetheless be incorporated into the optimal wage scheme, since they can be used to reduce (or eliminate) the necessary premium required for exposure to inequality.[50]

Remark. Inequity averse preferences provide a rationale for team contracting in the absence of informational or technological linkage between agents, violating the sufficient statistic result.

Next, in the case where limited liability constraints enable the agent to extract a rent, the principal sets $w_L = 0$ and her resulting costs can be expressed as:

$$C^P(a) = p(a)\frac{c'(a)}{p'(a)} \tag{58}$$

Whether the principal prefers to use an independent or team contract in this case

[50]The sufficient statistic result applies to the case of risk averse agents, rather than those who are risk neutral as we assumed in our simple example. When agents are risk averse, contracting on the performance of others induces a trade-off between exposure to risk and inequality. As noted in particular by Bartling and von Siemens (2010*b*), while the optimal contract for a risk and inequity averse agent will typically include information about the performance of others, this may not be the case if the resulting risk exposure is sufficiently large.

depends on which of (50) and (58) are larger. When the agent's preferences are predominantly envious, she will tend to prefer an independent contract due to the positive incentive effect of envy in this case, leading to lower costs.

The aforementioned analyses of Itoh (2004) and Demougin and Fluet (2006) both consider team contracting when performance measures are binary and discuss the conditions under which such contracts become optimal. Goel and Thakor (2006) present a very general model featuring n simultaneously risk averse and inequity averse agents, while both effort and the performance measure for each agent are assumed to be continuous. They then solve for the principal's cost-minimising contract in order to implement an effort level a. Unlimited liability is assumed throughout, so that participation constraints of agents always bind. As before, independence between signals of performance is assumed so that the classical theory predicts independent contracts. They show that the optimal payment scheme for an agent features a wage which is increasing in both his own performance and the performance of others, extending the finding that inequity concerns can lead to violations of the sufficient statistic result to a more general setting.

Since in their model agents are risk averse and signals are independent, varying wage payments with the performance of others increases the risk imposed upon an agent and therefore the necessary risk premium in order to induce participation. However, it also decreases the required inequity premium; the optimal wage scheme then trades off these two effects in such a way as to minimise the principal's costs, although payments will always be more sensitive to increases in an agent's own performance so that the provision of incentives remains the primary determinant of the wage schedule. Moreover, restricting attention to linear contracts, Goel and Thakor (2006) show that piecerates will decrease as the strength of other-regarding preferences increases, resulting in lower powered incentives, similar to the case of independent contracts discussed previously.

4.1.3 Relative Performance Evaluation

So far we have considered environments in which the performance measures of agents are independent of one another. However, we have previously seen in Section 2.5 that when these signals are positively correlated, the classical theory predicts that relative performance evaluation (RPE) — in which an agent's wage is negatively related to the performance of others — becomes optimal. This creates an interesting tension, since RPE will typically increase wage inequalities between agents.

A formal analysis of this issue is provided by Bartling (2011), who presents a two-agent model similar to that of Section 2.5. As before, there exists a performance measure $x_i = a_i + \epsilon_i$ for each agent, where ϵ_i is a normally distributed noise term which is positively correlated between agents, while the principal is restricted to offering identical linear contracts of the form $w_i = \eta + \gamma x_i + \delta x_j$. Each agent's preferences are now assumed to be represented by the CARA utility function:

$$U(w, a) = -e^{-\rho[w - c(a) - L(w,a)]} \tag{59}$$

where $L(w, a)$ denotes the agent's ex ante expected loss from inequality, given both the contract and the effort he undertakes. Inequity concerns therefore enter the utility function as a fixed loss, equivalent to a wealth effect; this approach implies that agents are risk averse over wages, but risk neutral over losses from inequitable outcomes. When choosing the variables γ and δ, the principal must take into account their impact on the term $L(w; a)$ and the resulting inequity premium required to ensure participation.

Bartling (2011) shows that the principal will choose the sign of δ and thus the type of contract (team, independent or RPE) by trading off three different forces associated with inequity aversion. First, as we have seen previously, other-regarding preferences create an incentive effect. If envy is the most prominent aspect of an agent's social preferences, then this incentive effect is positive and will be maximised when the contract features RPE ($\delta < 0$). Intuitively, an increase in performance will then raise the agent's own wage while reducing the wage of his coworker, significantly reducing the likelihood of the agent receiving a lower payment and experiencing envy. In contrast, with a team contract, increases in performance raise the wages of both agents, so that this incentive effect is much weaker. In the extreme case where $\gamma = \delta$ (a form of *extreme team contract*), there is no inequality since wages are identical and depend only on the sum of performance measures $(x_1 + x_2)$ so that the incentive effect is eliminated entirely.

Second, as in the standard model, the principal's choice of δ will influence the risk imposed upon the agent. As shown in Section 2.5, in the case of positive correlation between error terms, RPE is optimal for shielding the agent from risk.

Third, other-regarding preferences create disutility from inequity concerns. Note that the difference in wages between agents can be expressed as:

$$w_1 - w_2 = (\gamma - \delta)(x_1 - x_2) \tag{60}$$

For given levels of performance, the magnitude of this difference is maximised when $\delta < 0$, or when we have RPE. In this case, even small differences in performance lead to relatively large wage disparities and therefore a high loss from inequity $L(w; a)$. In contrast, by setting $\delta > 0$, differences in wages for given performance levels are reduced and can even be eliminated with an extreme team contract ($\gamma = \delta$).

While incentive provision and risk shielding therefore provide a motivation for RPE, maximal reduction of inequality requires a team contract.[51] The sign of the optimal δ will be determined by this three-way trade-off. If agents are sufficiently averse to inequitable outcomes, a team contract can become optimal even when error terms are positively correlated so that the standard theory predicts that RPE is optimal.[52] Bartling (2011) argues that this may provide an explanation for the lack of empirical evidence for real-world use of relative performance evaluation. Moreover, the magnitude of γ is likely to be reduced relative to the standard case, since a high piecerate will not only increase exposure to risk but also expected inequality.

In a related paper, Bartling (2012*a*) argues that firms may decide against using RPE not only because it increases wage inequality, but also because it can *generate* social comparisons by lowering the 'social proximity' between agents and making such comparisons more salient. In this case, when there is relatively little correlation between performance measures, the benefits of RPE are limited and as such the principal will use an independent contract in order to avoid the costs associated with agents comparing wages with one another. RPE only becomes beneficial for the principal when performance measures are strongly correlated, in which case risk exposure can be significantly reduced, while the difference in wages — and therefore the required inequity premium — will typically be small.

Krapp and Sandner (2016) show that movements away from RPE and toward team contracting can also be motivated by a principal's desire to adhere to an equal pay norm, even when agents are self-interested. This is modelled by assuming that ex post inequality between the wages of agents enters the principal's payoff function directly as a cost. The optimal contract will then result from a trade-off between the desire for equal pay and the need to shield agents from risk. It is not surprising that the impact of an equal pay norm is similar to that of inequity averse agents. In the latter case, the principal must *indirectly* bear the costs of wage inequality through the increased inequity premium which must be paid to agents in order to induce participation, while in the former case these costs enter *directly*. In both instances, the principal has an incentive to reduce the inequality exposure of the optimal contract.

Other papers have explored how relative performance evaluation may become

[51] Note that in the case of negatively correlated performance measures, a team contract is optimal not only for reducing inequality, but also for reducing exposure to risk.

[52] DeMarzo and Kaniel (2017) find similar results when agents care about how their wage compares to the *average* of the reference group. In the limit, as agents' relative income concerns become the driving force of preferences, an extreme team contract where wages are paid on the basis of aggregate performance becomes optimal. See also Fershtman et al. (2003), Miglietta (2008) and Goukasian and Wan (2010).

optimal with other-regarding agents even in the absence of correlated performance signals. For example, as discussed previously, status-seeking preferences create positive incentive and participation effects. Itoh (2004) then shows that if this is the driving force of agents' preferences, the principal can take advantage of these positive effects by increasing the probability of inequitable outcomes through use of RPE. Alternatively, Rey-Biel (2008) analyses how relative performance evaluation may be optimal for the principal in the absence of moral hazard, where the effort choices of agents are perfectly contractable. In the standard case, inducing an agent to undertake high effort then simply requires compensating him for his additional effort costs. However, with other-regarding preferences, the principal may be able to induce high effort by threatening an inequitable outcome in the case where an agent shirks; this then reduces the compensation required for the agent to undertake this high effort and leads to lower costs for the principal. Note, however, that this can only be the case when the agent extracts a rent (for instance, due to minimum wage constraints), otherwise the agent must be compensated fully for the additional effort undertaken in order to participate.

As a final remark, note that throughout our discussion we have so far considered agents who compare wages, rather than accounting for any differences in effort costs. At first glance this is of minor importance, since most papers in the literature consider homogeneous agents who are offered identical contracts, so that there is no difference in equilibrium effort levels or effort costs. However, the exact nature of the comparison has subtle implications for the impact of inequity aversion on incentives. To illustrate this, recall our finding that envious preferences have a positive incentive effect since, by increasing effort provision, an agent can reduce the probability of an outcome which entails disadvantageous inequality. When individuals care about disparities in net wages, there is an additional effect: should such an outcome be realised, the agent's higher effort provision means that the inequality is exacerbated; clearly, this will negatively impact incentives. Analogous results hold for the incentive effects of compassion and status-seeking.

Moreover, as emphasised by Bartling (2011), these additional effects have implications for the optimal contractual form. For instance, the incentive effect of envy becomes negative with an extreme team contract because — since wages are always equal — an agent who undertakes higher effort will decrease his net wage. The principal must then take such effects into account when designing the wage scheme.

4.1.4 Comparisons with the Principal

While most papers in the literature consider multiple agents who make social comparisons with one another, in some environments it may be plausible that an agent compares his income with that of the principal. Some authors have explored this possibility, typically in a model featuring one agent and one principal.

Itoh (2004) considers a simple framework, in which a risk neutral but inequity averse agent works on a binary-outcome project (*success* or *failure*) on behalf of the principal, whereby effort increases the probability of success. The relevant question in this environment regards how the benefits from a successful project should be split between parties. Assuming that the principal has a strictly higher income than the agent following a successful project, introducing envious preferences requires stronger incentives due to the disutility associated with the inequitable outcome, increasing the principal's costs.[53] This is true regardless of whether the participation constraint binds. In case it does, an inequity premium must also be paid, just as in the environments considered previously.

A more general analysis is presented by Englmaier and Wambach (2010). They study the classical moral hazard framework of Holmström (1979), in which a risk averse agent's effort influences the probability distribution over different levels of output $x \in [\underline{x}, \overline{x}]$, which then acts as a performance measure for the principal (see footnote 6). In their environment, the optimal wage scheme $w(x)$ therefore specifies how output should be shared between parties and is assumed to have a slope $w'(x) \in [0, 1]$ for all x.[54] When the agent is self-interested, there is the standard trade-off between providing maximal effort incentives (which calls for $w'(x) = 1$) and insurance against risk (which calls for a fixed wage and therefore $w'(x) = 0$).

Englmaier and Wambach (2010) introduce inequity aversion, but importantly, deviate from the Fehr and Schmidt specification by modelling the inequity loss as a strictly convex, U-shaped function $G(\cdot)$ around a minimum of zero, so that the agent displays both envy and compassion. An example of $-G(\cdot)$ is shown by the left-hand panel of Figure 4. Crucially, this formalisation implies that the agent dislikes lotteries over different levels of inequity, in contrast to the model of Fehr and Schmidt. As a result, minimising the necessary inequity premium requires that an additional unit of output should be shared equally between parties so that $w'(x) = \frac{1}{2}$ and the income differential remains constant over realisations of x. This additional force must then be taken into account when determining the optimal contract (see the right-hand panel of Figure 4). In the limit, as

[53]Following a similar logic, in the case where the *agent* is ahead following a successful project, the principal's costs are increasing in compassion, but decreasing in status-seeking.

[54]This is a common assumption in the moral hazard literature. One possible justification is that parties may be able to boost or destroy output; see Innes (1990).

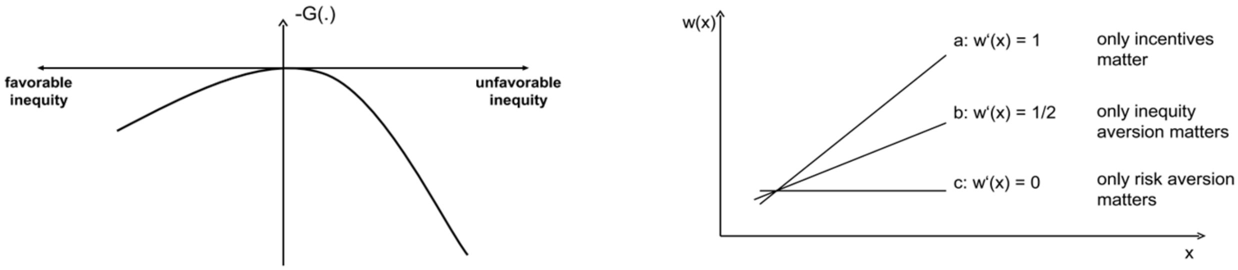

Figure 4: The function $-G(\cdot)$ (LHS) and the forces which determine the optimal wage scheme (RHS); reproduced from Englmaier and Wambach (2010).

the agent's dislike of inequity becomes the driving force of his preferences, the principal maximises her profit by utilising an equal sharing rule in which $w'(x) = \frac{1}{2}$ for all x. Moreover, any additional information about the principal's income — even if it is uninformative with respect to the agent's effort choice — should be included in the contract, so that once again the sufficient statistic result is violated.[55]

Assuming a similar specification of inequity aversion, Dur and Glazer (2007) show that an agent's dislike of lotteries over levels of inequality can lead to variable wage payments even in the absence of moral hazard, so that the standard theory would predict a fixed wage since the agent is risk averse. While there is no need to provide incentives since effort is contractable, a variable wage can be used to insure the agent against variations in the income differential between parties and therefore reduce the necessary inequity premium. The authors argue that this may provide an explanation for firms offering stock-options to low level workers, a phenomenon which cannot be explained by incentive provision since their effort will have a negligible impact on the stock price.

Finally, Banerjee and Sarkar (2017) show that introducing other-regarding preferences on the part of the principal (with respect to the agent) yields similar results to the self-interested case, since the principal will still design the contract in order to maximise profits. This holds unless the principal is strongly compassionate, in which case output will be split such as to ensure income equality between parties (see also Itoh, 2004).

4.1.5 Further Applications

To conclude this subsection, we consider how the concept of inequity aversion has been applied to a broad range of topics associated with incentive contracting. Demougin et al. (2006) allow for technological dependence between the tasks of

[55] See Cato and Ebina (2014) for an extension of the Englmaier and Wambach (2010) model to the multi-period case.

asymmetric agents in order to study the impact of inequity aversion on output. Increasing the effort provision of one agent requires higher incentive pay and impacts expected wage inequality and, therefore, the required inequity premia for coworkers. This introduces a further interdependence between the effort of agents. They show that increases in inequity aversion will typically lead to a reduction in total output, especially when tasks are complementary. Intuitively, in this case, implementing lower effort from one agent in order to alleviate inequalities has the additional effect of reducing the marginal return to effort of other agents.

Teyssier (2007) considers a competitive market in which principals compete for agents with heterogeneous social preferences. Self-interested agents prefer a competitive tournament-like scheme which generates high output, whereas inequity averse agents dislike the wage inequality inherent in such a scheme and prefer a contract in which output is shared equally. It can be shown that a separating equilibrium exists in which each type of agent prefers to self-select into their preferred wage scheme.

So far our analysis has assumed the existence of a suitable contractable performance measure. However, when this is not the case, effort incentives can still be provided using long-term *implicit* or *relational* contracts, which make use of non-verifiable signals of performance and must be self-enforcing, since courts cannot punish parties for deviating from the contractual terms. Kragl and Schmid (2009) consider a multi-agent model in which, at the end of each period, the principal can choose to renege on agreements by refusing to pay the agreed upon bonus payments. This yields an immediate benefit in the form of lower wage costs, but damages the principal's reputation and prevents contracting in subsequent periods, leading to zero future profits. Requiring contracts to be self-enforcing thus imposes the constraint that bonus payments must be smaller than the value of the continuation game.

Introducing envious preferences on the part of the agents has two key effects on this *credibility constraint.* First, due to the incentive effect of envy, the bonus payment required to implement a given effort level is smaller so that the constraint becomes easier to satisfy. Second, due to the participation effect of envy, profits in any given period are lower which means that honouring the contracts becomes less attractive; this makes satisfying the constraint more difficult. The overall impact of envy will depend on which effect dominates. Kragl and Schmid (2009) derive conditions under which envious preferences relax the credibility constraint and expand the set of self-enforcable contracts, allowing higher profits for the principal so that envy can be beneficial even when the participation constraint is binding and inequity premia must be paid.

Two further papers then explore alternative contractual forms when agreements must be self-enforcing. Kragl (2015) considers team contracts, which as we have seen can eliminate inequity concerns but often require larger bonus payments relative to individual wage schemes, while Kragl (2016) analyses tournament schemes which can avoid commitment problems but necessarily result in unequal wages between agents.

An important topic in the classical moral hazard literature is multitasking, whereby an agent's effort provision has numerous dimensions, each corresponding to a different task he has been assigned. The key finding of this literature, known as the *equal compensation principle*, is that the contract must be designed such that the agent's marginal return to each task is equalised. If not, then tasks which yield low returns to the agent will be ignored completely (Holmström and Milgrom, 1991). Bartling (2012*b*) studies an environment in which agents are required to undertake two tasks. The first, individual production, potentially entails wage inequalities between agents since the relationship between effort and output is stochastic. The second consists of contributing toward team production, in which all agents are rewarded based on their aggregate output, and therefore the wages associated with this task are necessarily equal. As we have previously seen, inequity aversion can cause the principal to lower incentive intensity when there is the potential for wage inequalities — in this case for individual production — in order to reduce the necessary inequity premium. The equal compensation principle then requires that incentives for team production must also be reduced, despite the fact that this task cannot contribute toward differences in wages.

Finally, Grund and Przemeck (2012) investigate the impact of inequity aversion on performance appraisals within organisations. There is significant empirical evidence to suggest that firms have a systematic tendency to *i)* overrate the performance of their employees (a *leniency bias*) and *ii)* differentiate only slightly between difference performance levels between employees (a *centrality bias*), so that appraisals of workers therefore tend to be grouped together toward the higher end of the performance scale. Grund and Przemeck (2012) show that this can be explained using a simple model with an altruistic supervisor who appraises the performance of inequity averse agents, whose wages are dependent on their appraisal. Note that since agents are inequity averse, the wage inequality associated with widely contrasting reports will induce disutility. Accordingly, since the altruistic supervisor cares about the utility of the agents, she will tend to award similar reports, resulting in the centrality bias. Similarly, overstating the performance of agents leads to increases in their wages and utility, from which the leniency bias results.

4.2 Reference-Dependent Preferences and Loss Aversion

4.2.1 A Simple Model

We now move onto studying the implications of reference-dependent preferences and loss aversion for incentive contracting. To begin our analysis, we once again consider a simple example with continuous effort and a binary performance measure. We assume that the agent's monetary preferences are represented by the piecewise linear function (30), where $\lambda > 1$ and $r > 0$. Moreover, we restrict attention to cases whereby $w_L < r < w_H$.[56] Given these assumptions, the agent's expected utility can be written as:

$$p(a)w_H + [1 - p(a)]\,[\lambda w_L - (\lambda - 1)\,r] - c(a) \qquad (61)$$

As before, the principal wishes to minimise the expected wage payment, so that her problem continues to be given by (43). To ensure incentive compatibility, we again take the first-order approach and set the derivative of (61) with respect to a equal to zero. Rearranging for the bonus payment w_H then yields:

$$w_H = \lambda w_L - (\lambda - 1)\,r + \frac{c'(a)}{p'(a)} \qquad (62)$$

It is clear to see that the RHS of (62) is decreasing in both r and λ. This implies that, for a fixed w_L, the necessary bonus payment (and therefore wage spread) is decreasing in both the reference point and the agent's aversion to losses. It follows that loss aversion induces a positive incentive effect. Intuitively, a higher λ increases the extent to which the loss associated with receiving the wage w_L is felt by the agent, while a higher reference point r increases the relative size of this loss. Both effects reduce the utility associated with w_L, while leaving the utility associated with w_H unchanged, resulting in higher effort incentives.

We next consider the effect on the agent's participation decision. Similar to our previous analysis, we assume that the agent's expected utility must be weakly greater than his outside utility level, which we set equal to zero.[57] Substituting

[56] If this was not the case, both wage payments would be on the same portion of the utility function (30), which would then be linear over the relevant region.

[57] This assumption implies that the agent's outside utility is invariant to both his level of loss aversion and the reference point. The literature has typically followed this approach, especially in contexts where the reference point is determined endogenously by the contract. However, one could imagine scenarios — such as when r is determined by last year's salary — in which the agent will continue to make comparisons with the reference point when rejecting the contract. Moreover, the exact assumption made will determine the impact of changes in λ and r on the participation constraint and therefore has important implications; see footnote 26 for a similar argument in the context of inequity aversion.

(62) into (61), the participation constraint then becomes:

$$\lambda w_L - (\lambda - 1)\, r + p(a)\frac{c'(a)}{p'(a)} - c(a) \geq 0 \tag{63}$$

When wage payments are unrestricted, so that this constraint binds, the low wage w_L can be expressed as:

$$w_L = \frac{1}{\lambda}\left[c(a) + (\lambda - 1)\, r - p(a)\frac{c'(a)}{p'(a)}\right] \tag{64}$$

This implies that the principal's costs of implementing effort are:

$$C^P(a) = c(a)\left[p(a) + [1 - p(a)]\,\frac{1}{\lambda}\right] + \frac{\lambda - 1}{\lambda}\,[1 - p(a)]\left[r + p(a)\frac{c'(a)}{p'(a)}\right] \tag{65}$$

Studying (65) allows us to make some observations. First, note that by setting $\lambda = 1$, the utility function (30) becomes $u(w) = w$, so that we are in the standard risk neutral case; the principal's costs of implementing effort then reduce to $C^P(a) = c(a)$. Second, for $\lambda > 1$, the RHS of (65) is increasing in both r and λ, so that loss aversion induces a negative participation effect, raising the principal's costs.[58] We have seen that increases in both r and λ reduce the utility associated with the wage payment w_L, aiding the creation of incentives. However, for the same reason these increases also reduce the agent's expected utility, so that a *loss premium* must be paid by the principal.

Alternatively, if limited liability restricts the principal's choice of wages such that the agent extracts a rent, she will set $w_L = 0$, which implies the following costs:

$$C^P(a) = p(a)\left[\frac{c'(a)}{p'(a)} - (\lambda - 1)\, r\right] \tag{66}$$

which are clearly decreasing in both r and λ, since only the incentive effect of loss aversion applies. Our findings are summarised by the following remark.

Remark. An increase in the reference point r or the agent's aversion to losses λ leads to a positive incentive effect, reducing the wage spread required to implement effort a. When limited liability restricts the wages available to the principal, so that $w_L = 0$, the high payment w_H can then be lowered, leading to a decrease in the principal's costs. However, when the participation constraint binds, the principal must compensate the agent for the disutility associated with exposure

[58] This is clearly true for r. Differentiating the RHS of (65) with respect to λ yields:

$$\frac{\partial C^P(a)}{\partial \lambda} = \frac{1}{\lambda^2}\,[1 - p(a)]\left[r + p(a)\frac{c'(a)}{p'(a)} - c(a)\right]$$

which can be shown to be positive; see footnote 12.

to losses, increasing her costs.

Note that there is a direct parallel between the impact of loss aversion and the impact of envy, studied in Section 4.1: both lead to a positive incentive effect, but a negative participation effect, with the overall consequences for the principal's costs dependent on the economic environment. Moreover, similar to envy, loss aversion will typically lead the principal to implement a reduced effort level relative to the classical case.

4.2.2 The Optimal Contractual Form

The literature on incentive provision in the presence of loss aversion has typically focussed on studying the form of the optimal contract when the principal has access to a rich performance measure. An early contribution by de Meza and Webb (2007) considers a binary effort environment ($\mathcal{A} = \{a_L, a_H\}$) with a continuous signal of performance ($\mathcal{X} = [\underline{x}, \overline{x}]$), in which the agent's utility function is additively separable in terms capturing *i)* intrinsic consumption utility, *ii)* reference-dependence and loss aversion and *iii)* effort costs, so that the agent's preferences over monetary outcomes are similar to (31). Consumption utility is assumed to be strictly concave, so that the standard theory would predict an everywhere strictly increasing wage scheme. de Meza and Webb (2007) then solve for the principal's cost minimising contract in order to implement the effort level a_H, making various assumptions about both the reference point and the nature of the agent's loss aversion.[59]

They first consider the case of linear loss aversion with respect to an exogenously given reference point; one possible interpretation of this might be that the agent compares his wage to the previous year's earnings. In this case, they show that the optimal wage scheme typically features a region over which pay is insensitive to performance, with wage payments here equal to the reference point.[60]

As illustrated by Figure 5, this region can occur at the start, middle or end of the wage schedule, depending on the value of the reference point.[61] de Meza and Webb (2007) argue that this pay insensitivity results from the principal's need to balance creating effort incentives with ensuring the participation of the agent.

[59]Note that solving for the optimal contract typically entails differentiation of the agent's utility function with respect to the wage scheme. This becomes difficult when the utility function is kinked, due to the non-differentiability at r. Many papers in the literature therefore require the use of subdifferential calculus to solve the principal's problem.

[60]Kanbur et al. (2008) study the issue of income taxation under moral hazard when agents have reference-dependent preferences and similarly find that full insurance around the reference point is optimal.

[61]There do exist cases in which the agent's wage is everywhere strictly increasing. However, this occurs only when the reference point is either extremely high or extremely low, so that the whole wage schedule is entirely in either the loss or gain space.

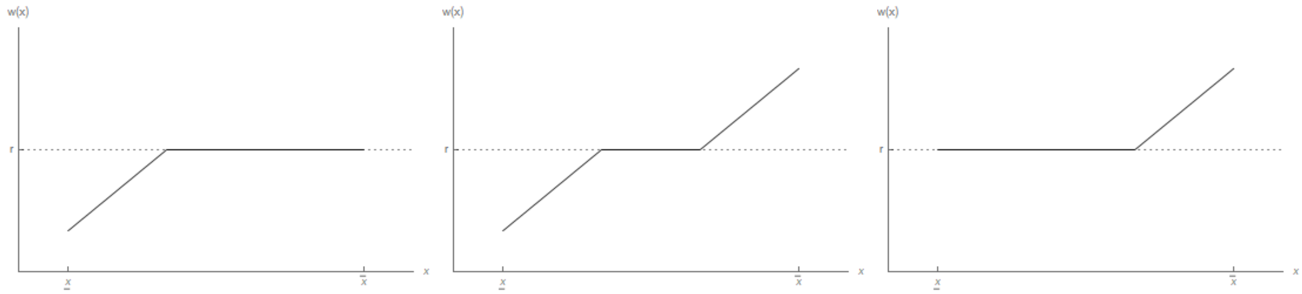

Figure 5: Three possibilities for the shape of the optimal wage scheme; adapted from de Meza and Webb (2007).

Intuitively, receiving a wage payment which is perceived as a loss reduces the agent's utility. If the participation constraint is binding, then in order to meet the agent's outside option, the principal is forced to pay a loss premium which compensates for the possibility of such an event. The optimal contract takes this into account; while a large degree of wage differentiation is an effective means of inducing effort, it also raises the probability that the agent experiences a loss, hence increasing the required loss premium. This generates a trade-off for the principal. Exposure to losses for poor performance is a powerful way to create incentives, but necessitates paying higher wages elsewhere to ensure participation. Accordingly, the principal finds it advantageous to shield the agent from losses for some realisations of the performance measure so that the wage scheme becomes partially unresponsive around the reference point. As we shall see, this has been the key finding in the literature on incentive contracting with loss averse agents and is recurrent in several studies.

Next, de Meza and Webb (2007) introduce diminishing sensitivity into the agent's utility function, which implies convexity below the reference point. Imposing a lower bound on the set of possible wage payments, they show that the principal will never choose to pay a wage in between this lower bound and the reference point. Intuitively, since the agent is now risk loving in the loss region, the utility level associated with a particular wage could be provided to the agent at a lower cost by paying a lottery. However, they further show that lotteries are not a feature of the optimal contract; in order to best create incentives, any wages which are below the reference point will be bunched at the lower bound. The wage scheme then potentially has two flat regions: one at the lower bound, the other at the reference point. In some cases, this will account for the entire wage schedule, so that there are only two possible wage payments and the contract becomes binary.[62]

[62]Dittmann et al. (2010) and Iantchev (2009) also consider agents with reference-dependent preferences, loss aversion and diminishing sensitivity, so that utility is convex below the reference point. Dittmann et al. (2010) assume that preferences over wages are given by (29), as suggested

Finally, they once again consider linear loss aversion but allow for endogenous reference points, which are determined by the design of the contract. In many cases this will be more plausible than an exogenous formulation and is closely related to the concept of expectations acting as a reference point. They consider two alternative formalisations.

First, it is assumed that the reference point is determined by the certainty equivalent of the distribution implied by the wage scheme. The principal's problem is simplified since the certainty equivalent is tied down by the participation constraint; accordingly, the optimal contractual form is identical to the case of an exogenous reference point. However, an important difference is that the region over which pay is insensitive is now always strictly on the interior of the wage scheme.

Second, they consider the median of this distribution, finding that the wage schedule again features wage insensitivities, but also discontinuities.

de Meza and Webb (2007) argue that their results are consistent with real world incentive schemes, which typically feature pay insensitivity and shielding from losses. For example, CEOs are often remunerated via a base salary plus stock options so that they are rewarded when performance is high, but not punished when it is low. This protection from losses cannot be explained by classical models of incentive provision.

Jofre et al. (2015) extend the framework of de Meza and Webb (2007) to a (finite-horizon) dynamic environment, assuming that the previous period's wage acts as a reference point for the agent. That is, the reference point varies over time and in period t is determined by the wage payment received in period $t-1$. They assume that the agent's utility function is everywhere strictly concave, but kinked, and show that under these assumptions the optimal contract exhibits two key differences relative to the classical case.

First, in each period the wage schedule features an interval over which pay is insensitive to performance and $w(x) = r$, replicating the result of de Meza and Webb (2007). This also implies the possibility of wage persistence, since there is always a positive probability of the period t payment being equal to that of period $t-1$.

Second, the principal understands that today's wage payment will determine the agent's reference point tomorrow and takes this into account when designing the contract. Since a high wage increases the probability of the agent incurring

by Tversky and Kahneman (1992), while Iantchev (2009) imposes a piecewise quadratic form. Both papers also find that the optimal wage scheme features a discontinuous drop below the reference point to the lowest possible wage payment, a feature which is interpreted as dismissal from employment. Wages are found to be continuously increasing elsewhere. The optimal contracts are then compared with real-world data, with both studies finding that loss aversion can explain observed wage schemes better than risk aversion alone.

a loss in the next period, the principal prefers to pay lower wages in order to reduce the prospect of a loss and lower the necessary loss premium, so that wages become compressed relative to the classical model. In particular, they do not rule out the case where wages are fixed in each period except the last — so that it is not possible for the agent to experience a loss — with incentives being provided solely through payments in the final period.[63]

The optimal contractual form has also been studied using the Kőszegi and Rabin (2006, 2007) model of expectation-based reference-dependent preferences outlined in Section 3.2.4. Herweg et al. (2010) consider a continuous effort environment with a discrete performance measure so that $\mathcal{X} = \{x_1, ..., x_n\}$. The agent's preferences over money are described by (33), where the (stochastic) reference point is determined endogenously by the expectations induced by both the wage scheme and the agent's action, in a choice-acclimating personal equilibrium (CPE). The gain-loss function capturing reference-dependence is assumed to be piecewise linear, while effort costs enter into the utility function as normal.

When the agent's intrinsic consumption utility is strictly concave, in the absence of loss aversion the optimal contract is fully contingent and specifies a different wage for each signal x_i. However, Herweg et al. (2010) show that the optimal contract for a loss averse agent can take a binary form, specifying only two distinct wage payments. They argue that this may provide an explanation for the real-world prevalence of bonus contracts, which are commonly used to motivate workers despite suffering from well-known drawbacks.[64]

Intuitively, recall that under the Kőszegi and Rabin (2006, 2007) model of preferences, an agent ex post compares his actual wage with each other possible wage, weighted by the ex ante probability with which the alternative outcome occurs. Due to loss aversion — and as shown previously in our simple example by (37) — these comparisons lead to an ex ante reduction in the agent's expected utility, requiring the principal to pay a loss premium. Herweg et al. (2010) show that this loss premium is increasing in the contract's degree of wage differentiation: a scheme which specifies many different wage payments leads to a large number of comparisons and causes the agent to ex ante consider deviations from the (stochastic) reference point likely. In order to reduce the necessary loss pre-

[63]Hori and Osano (2014) introduce loss aversion into a continuous-time agency model to explain the dynamics of CEO compensation, dividends and capital structure. Since this modification entails various departures from the standard principal-agent framework, we do not provide a detailed review of their analysis. However, consistent with the literature, they also find that the optimal contract features a range of outcomes over which compensation is invariant to outcomes.

[64]For instance, a worker who has already met a sales target has no further incentive to undertake effort if they already know they will be awarded the bonus. Moreover, they are motivated to delay further sales until subsequent periods. Both of these issues can be avoided by using alternative forms of remuneration, such as piecerates. See also the discussion in Herweg et al. (2010).

mium, the principal therefore has an incentive to pay the same wage for several different signals of performance. While the agent's risk aversion provides the usual motivation for a strictly increasing wage schedule, if this plays a relatively minor role compared to loss aversion, then the optimal contract minimises wage differentiation and pays only two distinct wages.

Further properties of the optimal wage scheme can also be derived. Recall from (37) that the agent's reduction in expected utility from loss aversion is maximised when outcomes are equally likely — in that example when $p = 0.5$. Accordingly, in order to minimise the loss premium, the optimal contract will pay the bonus either very often or very rarely, since this minimises the weight which the agent puts ex ante on experiencing a loss when the bonus is not paid. Finally, while a loss averse agent may allow the principal to use a lower powered incentive scheme relative to the loss neutral case, it can be shown that the principal's costs of implementing effort are strictly increasing in loss aversion, similar to our earlier results.[65]

In order to explain the real world prevalence of deferred incentives, Macera (2018*a*) considers a two-period dynamic environment in which the agent has preferences as described by Kőszegi and Rabin (2009), so that utility is affected by changes in the agent's rational expectations over both current and future payments. In equilibrium, the possibility of changes in expectations always results in an ex ante utility loss, since the prospect of being disappointed outweighs the prospect of being pleasantly surprised. As the agent is more sensitive to changes in expectations over current-period payments relative to future payments, this utility loss can be reduced by paying a fixed wage in the first period and deferring

[65] Herweg et al. (2010) also consider higher degrees of loss aversion and show that in such cases, the principal may be able to reduce implementation costs by committing to stochastically ignore the outcome of the performance measure, or 'turn a blind eye'. This allows for payment of the bonus even when the principal observes a signal which is suggestive of low effort. Daido and Murooka (2016) find similar results for even moderate levels of loss aversion.

Intuitively, in some environments, an agent may increase their ex ante expected loss by working harder, an effect which dampens effort incentives. This is especially true, for instance, when the performance measure is binary and the probability of a signal of high performance is low (c.f. equation 37). By committing to stochastically ignore the performance measure with a certain probability, the principal can influence both the agent's expected loss and the marginal impact of effort on this expected loss. Accordingly, such stochastic ignorance may be beneficial for the creation of effort incentives.

Daido and Murooka (2016) argue that since the principal may find it difficult to commit to such a contract, the same effect can be generated through the use of *team incentives,* since this allows for a low performing agent to still be awarded a bonus, so long as coworkers have performed well.

Along similar lines, Marchegiani et al. (2016) argue that by having a systematic tendency to overrate the performance of their employees (a *leniency bias*), firms are able to increase wage expectations; this aids the creation of incentives when agents are loss averse since the pain associated with receiving a low wage in the case of poor performance is increased. They then provide support for this argument by conducting a real-effort laboratory experiment, showing that a contract which occasionally fails to reward hard-working subjects induces significantly less effort than a contract which occassionally rewards those who shirk.

all incentives to the second period. This result is in contrast to the classical risk averse case, where the optimal contract uses a combination of both present and future incentives. In an accompanying paper, Macera (2018*b*) provides support from an experiment in which subjects were required to undertake a real-effort task over the course of two sessions and were allowed to choose between two incentive schemes. The majority of subjects preferred a scheme which defers incentives into the future, contrary to the predictions of the classical model.

The optimal contractual form under reference-dependence and loss aversion has therefore been studied in various environments, both static and dynamic, and with several different specifications of the agent's preferences. As we have seen, a recurrent finding has been that the optimal wage scheme is partially unresponsive to the performance measure, violating the sufficient statistic result. This is the key result of the literature and is summarised by the following remark.

Remark. When the agent's preferences are characterised by reference-dependence and loss aversion, the optimal contract typically features some degree of insensitivity, whereby wages do not increase in response to a higher signal of performance. This constitutes a violation of the sufficient statistic result, since contracts do not make use of all relevant information. Intuitively, wage insensitivities occur since the principal wishes to shield the agent from losses, in order to reduce the necessary premium required to induce participation.

4.2.3 Further Applications

Reference-dependent preferences have also been applied to various different aspects of incentive contracting. One particularly interesting application has explored how effort provision is impacted by the *framing* of incentives. For instance, a contract which specifies two different wage levels contingent on performance can be presented in two different ways: as a *bonus* contract, which offers a low base wage along with an additional payment for high performance, or as a *malus* contract, which combines a high base wage with a punishment or fine for low performance. While classical economic theory predicts that the way in which a contract is framed should be inconsequential for the provision of effort, this is not necessarily the case if the principal's choice of frame is able to influence the agent's preferences.

Armantier and Boly (2015) study a model in which a contract consists of a schedule of wages, one of which is specified by the principal as the base wage. This payment then acts as the agent's reference point. Due to loss aversion, contracts which implement effort using the threat of punishment are then more effective and should induce agents to work harder.[66] However, Armantier and Boly (2015) note that diminishing sensitivity also plays a role: since the agent's utility function is convex below the reference point, for large losses the agent becomes less sensitive to changes in income. Accordingly, when the base wage is unrealistically high, the marginal utility of income is low for most payments, reducing effort incentives. The authors find experimental evidence for these results and suggest that incentive contracts will be most effective when they combine the use of both penalties and rewards.

Just and Wu (2005) and Hilken et al. (2013) consider similar frameworks but also analyse the agent's participation decision. In this case, when choosing the contract frame, the principal must take into account not only the incentive effect but also the impact on the agent's expected utility. Consistent with the broader literature, they then find that since the disutility associated with exposure to losses requires a higher average payment to induce participation and therefore increases the principal's costs, the optimal contract will feature a low base wage with the possibility of bonus payments, so that punishments do not play a role.[67]

A related issue concerns how principals may be able to create effort incentives by using performance targets, even if they are not associated with monetary rewards. Corgnet et al. (2018) note that while economists have traditionally ignored the role of intrinsic motivation, psychologists have long argued that the motivation of individuals will typically depend on a variety of factors, including non-monetary incentives. For instance, a performance target may act as a yardstick by which workers can assess their ability, with the desire to appear competent creating incentives to meet this target, increasing effort provision.

Abstracting from explicit monetary incentives, Rablen (2010) studies a framework in which the agent's utility depends on how his performance compares to a target set by the principal, in a manner consistent with the value function of prospect theory. Since this target determines the agent's reference point, the principal can manipulate effort incentives through her choice of performance goal. Due to diminishing sensitivity, the marginal utility of performance is highest around the reference point. This implies that effort incentives will be inverse U-shaped in the performance target, with the optimal goal being some intermediate level which is challenging, but attainable. Rablen (2010) also notes that there is an additional argument against setting the performance goal too high if the agent

[66]See also Pierre (2016, 2018) for a similar argument in the static and dynamic cases, respectively.

[67]Note that this only holds when the agent's outside option is invariant to the principal's choice of frame. This may not always be the case. For example, suppose a worker is negotiating a contract with his current employer but also has a job offer elsewhere; it seems plausible that by modifying the terms of the contract, the employer can influence the worker's perception of this job offer. If the principal can reduce the agent's outside utility by increasing the base wage (and therefore the reference point), then the optimal contract will typically feature punishments.

has some discretion over the degree of risk in the production process. A high goal means that performance will often fall below the target, with the associated convexity below the reference point implying that the agent becomes prone to risky behaviour, even when this has no effect on the expected level of output.

Corgnet et al. (2018) consider a similar model in which the principal can set a wage-irrelevant performance target in order to create incentives, but also allow for explicit incentive pay in order to analyse the interaction between the two. They show that these non-monetary incentives act as a substitute for performance pay, allowing the principal to reduce the incentive payments required to implement effort. However, the agent must be ex ante compensated with a higher fixed wage in order to guarantee participation due to the possibility of a utility loss from failing to meet the target. In their framework, the first effect dominates so that principals benefit from setting wage-irrelevant performance goals, with the optimal contract typically making use of both monetary and non-monetary incentives. They complement their theoretical findings with experimental evidence and suggest that non-monetary incentives may help explain the real world prevalence of low powered wage schemes.[68]

The efficient assignment of tasks within firms when agents have expectation-based reference-dependent preferences has also been studied. Daido et al. (2013) consider an environment in which there is uncertainty over which agent will be more productive for a certain task. Assuming that the principal can write a state-specific contract, with classical preferences the task will be assigned in each state of the world to the agent who is most productive. However, this may no longer be the case. Intuitively, a state-contingent assignment creates uncertainty over both wages and effort costs, which, with expectation-based loss aversion, leads to a reduction in the expected utility of agents. Since this increases the compensation required to ensure participation, a state-independent task assignment may be optimal if loss aversion is sufficiently strong. Balmaceda (2018) similarly explores how such preferences might influence a principal's choice between assigning a number of tasks all to one agent (multitasking), or to several agents, each of whom undertakes a single task (specialisation).

Finally, Daido and Itoh (2007) analyse two effects in relation to self-fulfilling prophecies within a principal-agent framework. As we have previously seen, in

[68] This strand of the literature highlights a more general point: changes to the agent's reference point — whether controlled by the principal or not — will typically influence the outcomes of contracting between parties. For instance, while classical economic theory predicts that non-binding regulatory caps on executive pay are inconsequential, Städter (2018) shows that this may no longer be true if reducing the set of feasible contracts impacts the agent's reference point. Alternatively, Keefer (2016) argues that the organisational culture of a firm can lead to changes in reference points by, for example, emphasising the importance of either work-related effort or of job-related perks. In such cases, the induced change to the agent's preferences will have implications for both wages and effort provision.

the Kőszegi and Rabin (2006, 2007) model of preferences, differences in initial expectations can lead to different behaviours in any given environment. One example of this concerns an agent's beliefs about his own future performance: if an agent expects high performance, then he will be more likely to work hard in order to meet this goal. This notion — that a worker's enhanced expectations can lead to improved performance — is known as the *Galatea Effect*. Moreover, the principal may be able to influence what kind of expectations are formed, for instance, if the agent cares about how his performance compares to the principal's own beliefs. This is then known as the *Pygmalion Effect*, whereby a manager's high expectations induce workers to act in ways as to fulfil these expectations. Daido and Itoh (2007) show that both effects can be captured by reference-dependent preferences and can be used by the principal to lower the costs of implementing effort.

4.3 Tournaments

An alternative method by which a principal can create effort incentives for several agents is through the use of a tournament scheme, whereby the remuneration of an agent depends on how his performance is ranked relative to others. The use of tournaments has numerous advantages over independent incentive schemes: only ordinal information about the performance of agents is needed, which can lower measurement costs; a principal can credibly commit to the scheme even when performance signals are unverifiable; the principal's wage expenditure is fixed and known in advance; results will be unaffected by common shocks to performance and, finally, tournaments often arise naturally in real-world situations, such as when many agents compete for promotion within a firm. However, there are also some disadvantages: tournaments may induce undesirable activities such as sabotage or collusion; incentives can also be drastically reduced if agents have access to interim information.

As a starting point for our analysis, we briefly consider a simple model of rank-order tournaments, based on the work of Lazear and Rosen (1981). A principal employs two identical agents, again denoted by 1 and 2, to undertake costly effort on her behalf and has access to a performance measure for each, which takes the form $x_i = a_i + \epsilon_i$. The error terms ϵ_i are assumed to be normally distributed and independent, with zero mean and a variance of σ^2. After effort has been undertaken and the values of x_1 and x_2 have been realised, the agents' performances are ranked. A *winner's prize*, or a high wage, denoted by w_H is awarded to the agent who achieves the highest performance level; the other agent receives the low wage or *loser's prize* w_L.[69] The effort incentives created by the

[69] A complete scheme must also specify the outcome should the tournament end in a draw.

tournament will depend upon the prize structure and in particular the prize spread between w_H and w_L.

Formally, the principal's problem is to implement the desired effort for each agent while minimising total wage expenditure $w_H + w_L$. The probability of Agent 1 winning the tournament is:

$$\Pr\{x_1 > x_2\} = \Pr\{\epsilon_1 - \epsilon_2 > a_2 - a_1\} \tag{67}$$

Let $G(\cdot)$ denote the cumulative distribution function of the random variable $\xi = \epsilon_1 - \epsilon_2$, with associated density $g(\cdot)$. It follows that $\Pr\{x_1 > x_2\} = 1 - G(a_2 - a_1)$. Agent 1's expected utility is then:

$$[1 - G(a_2 - a_1)]\, w_H + G(a_2 - a_1) w_L - c(a_1) \tag{68}$$

The prize structure must be such that agents are induced to both participate in the tournament and undertake the desired effort level. As in the case of independent contracting, the agent's incentive compatibility constraint is replaced by the first-order condition of (68) with respect to a_1, which is set equal to zero. Rearranging then yields:

$$w_H - w_L = \frac{c'(a_1)}{g(a_2 - a_1)} \tag{69}$$

In a symmetric Nash equilibrium, both agents undertake identical effort $a = a_1 = a_2$, which implies:

$$w_H - w_L = \frac{c'(a)}{g(0)} \tag{70}$$

As before, the participation constraint requires that each agent's expected utility is weakly greater than his outside option, which we again set equal to zero. From substituting the foregoing into (68) — along with $a_1 = a_2$ and $G(0) = \frac{1}{2}$ — we then have:

$$w_L + \frac{1}{2}\frac{c'(a)}{g(0)} - c(a) \geq 0 \tag{71}$$

If limited liability does not restrict the prizes which the principal can award, this constraint will bind, implying the following loser's prize:

$$w_L = c(a) - \frac{1}{2}\frac{c'(a)}{g(0)} \tag{72}$$

For instance, a winner could be selected at random, or each agent could receive an equal wage of $\frac{1}{2}[w_H + w_L]$. However, the exact rule is inconsequential for our analysis since performances are equal with probability zero in the current framework.

Note that the expected cost per agent of implementing the desired effort level is $\frac{1}{2}[w_L + w_H]$. Substituting (70) and (72) then yields:

$$C^P(a) = c(a) \tag{73}$$

It follows that since the principal's costs of implementation are equal to the agent's effort costs, a rank-order tournament with risk neutral participants and unlimited liability leads to the first-best outcome, similar to an independent scheme (Lazear and Rosen, 1981). Moreover, both Lazear and Rosen (1981) and Green and Stokey (1983) argue that tournaments can outperform independent contracts if error terms are sufficiently correlated between contestants, similar to the case of relative performance evaluation. In the presence of limited liability constraints, however, the principal's choice of prizes may be restricted such that agents earn a rent under the optimal scheme. As in the case of independent contracting, this will typically lead to the implementation of suboptimal effort levels.

The study of tournament schemes when agents are inequity averse is particularly important, since — unlike the other incentive schemes we have previously considered — tournaments necessarily result in unequal wages being paid to participants. Moreover, the fact that contestants are competing with one another for rewards is likely to make wage comparisons especially salient. Grund and Sliwka (2005) study a two-player tournament when agents are both compassionate and envious. They show that many of the results we have previously derived extend to the case of tournaments. First, for a fixed wage structure, an agent's effort choice under a tournament scheme will be increasing in envy, but decreasing in compassion. The intuition behind this incentive effect is identical to the case of independent contracting with inequity averse agents: envy reduces the utility associated with low performance (relative to the other agent), whereas compassion reduces the utility associated with high performance.

Next, since inequity aversion decreases expected utility, there is again a negative participation effect due to the inequity premium. When agents do not extract a rent so that the participation constraint binds, this increases the principal's costs and, as a result, the optimal tournament fails to implement first-best effort, in contrast to the self-interested case. In a similar model, Demougin and Fluet (2003) introduce limited liability constraints and show that when agents do extract a rent, so that only the incentive effect remains, the principal's costs will be reduced if envy is the driving force of preferences. Compassion, on the other hand, can never be beneficial for the principal.[70]

Through these effects, there is a direct parallel between the impact of inequity

[70] See also Kräkel (2000), who studies a model of tournaments in which agents wish to minimise the relative income difference between themselves and all others who are better off.

aversion on tournament schemes and the case of independent contracting studied in Section 4.1. There may, however, be a difference in the *strength* of the effects. Demougin and Fluet (2003) establish that when agents do not extract a rent, individual incentive pay will typically entail lower costs for the principal than a tournament. The intuition here is straightforward: tournaments lead to unequal wages between agents with certainty, whereas this is not necessarily the case under an individual scheme. This then implies that inequity premia will be larger under a tournament scheme, so that the principal's costs are also increased.

The case where agents extract a rent is less clear. Demougin and Fluet (2003) argue that this same guarantee of inequality can lead to a stronger incentive effect under a tournament scheme, meaning that the required wage spread is reduced — but this will depend on the exact specifications of the model. Similarly, in a slightly different framework, Dubey et al. (2013) show that tournaments can lead to stronger effort incentives when agents are envious and status-seeking, since under a tournament scheme increased effort leads to not only a higher expected wage for the agent, but also a reduction in the expected wages of competitors.

Ederer and Patacconi (2010) also consider envious and status-seeking agents, in a framework where reference points are determined endogenously by the prize structure of an n-player tournament. This captures the notions of interpersonal comparisons and preferences for high status, both of which seem natural in a tournament environment. The output of agents is ranked, with the top $k < n$ participants being awarded a high wage (winner's prize), while the remainder receive a low wage (loser's prize). The point of reference for agents is assumed to be the *modal wage* in the tournament; whether this is the high or low prize will therefore depend on the principal's choice of k. Ederer and Patacconi (2010) show that the principal benefits from keeping the number of winners k sufficiently low such that the reference wage is equal to the loser's prize. In this case, since all participants receive a wage weakly greater than the reference point, the principal is able to eliminate the disutility which agents associate with low status. In fact, since wages above this reference point yield extra utility to status-seeking agents, the principal is able to reduce wage costs relative to the self-interested case.

Schöttner (2005) investigates the use of *J-type tournaments*, which, rather than fixing the prize structure, fix a bonus pool which is then split between agents based on their relative performance. These incentive schemes are particularly valuable when agents are inequity averse, since they are able to retain some of the advantages of standard tournaments (for instance, fixed wage expenditure) without exposing participants to the same level of expected inequality.

Moving away from inequity aversion, Gill and Stone (2010) study two-player tournaments in a framework where agents have perceived entitlements which are sensitive to how hard they worked relative to a rival, or alternatively, care about receiving their *just deserts*. Formally, each agent has a reference point which is defined by his expected wage, conditional on the effort provision of both himself and his competitor. They assume that agents dislike wages which fall below this reference point, while wages above the reference point can result in either a positive or negative utility impact. Their specification therefore captures elements of both inequity aversion and expectation-based reference-dependent preferences.[71] In particular, they argue that these 'desert' preferences capture a more sophisticated notion of fairness than inequity aversion, since agents additionally account for the extent to which they feel any inequality is deserved.

Gill and Stone (2010) show that the expected loss in utility from desert in tournaments is maximised when agents undertake identical efforts. Intuitively, in this case, both agents feel that they deserve an equal share of the winner's prize. Yet, since one will win and one will lose, they will both receive a payoff far away from this common reference point. In contrast, when there is a significant difference between effort provision, the expected loss from desert is small: the hard working agent's expected payoff (and therefore reference point) is close to the winner's prize, while the shirking agent's expected payoff is close to the loser's prize. Moreover, since the hard working agent often wins the tournament, both players typically receive wages which are close to their reference points. If desert concerns are sufficiently strong, symmetric equilibria cannot exist, since one player would always prefer to 'resign' by lowering effort provision, thereby reducing the expected desert loss. This result is particularly important, since many findings of the tournament theory literature typically rest upon the existence of a symmetric equilibrium (Dato et al., Forthcoming).

Dato et al. (Forthcoming) emphasise the relationship between Gill and Stone's (2010) notion of just deserts and the Kőszegi and Rabin (2006, 2007) model of preferences, showing that the foregoing result is consistent with a choice-acclimating equilibrium (CPE) in the latter framework.[72] However, it is feasible that in some

[71] Since agents care about how their actual payoff compares to their expected payoff, there is a particularly strong resemblance to some of the early models of reference-dependent preferences in which the reference point is determined by expectations. However, there are some key differences. For instance, Gill and Stone (2010) later consider 'unfair' tournaments, where one player has an advantage which is felt to be undeserved. In this environment, when agents undertake identical efforts, the player with the advantage faces a higher probability of winning the tournament and therefore a higher expected payoff. However, since the advantage is unfair, expected payoffs no longer coincide with perceived entitlements. In such cases, Gill and Stone (2010) define the reference point as being the *counterfactual* expected payoff had the tournament been fair, so that there is a clear departure from models which assume that the reference point is defined by actual expectations.

[72] Similarly, Bergerhoff and Vosen (2015) consider dynamic tournaments, in which interim information is available to participants, when agents have expectation-based reference-dependent preferences and the solution concept is a CPE. The foregoing logic also applies in this framework, leading to asymmetric effort provision at the interim stage, since this reduces each agent's

tournaments — where the time between decision making and emergence of results is small — expectations may not have time to acclimate and as such remain fixed. Dato et al. (Forthcoming) therefore consider the alternative solution concept of choice-unacclimating equilibrium (UPE). Since an agent's expectations are then fixed with respect to effort provision, there is no longer any benefit from resigning, as this would serve only to reduce the probability of winning and increase the expected loss. Instead, there will typically exist several equilibria around (and including) the standard symmetric equilibrium; this multiplicity occurs since agents become 'attached' to their initial plans and therefore unwilling to deviate. However, similar to Gill and Stone (2010), as agents become increasingly sensitive to how their wage compares to the reference point, only outcomes in which agents undertake asymmetric effort levels can be a preferred personal equilibrium for both agents.

4.4 Teams

Another important topic in the classical literature on incentive contracting has been the study of *team production* or *partnerships*, whereby several agents undertake effort which contributes to the joint production of output, with the resulting profits distributed between parties according to some *sharing rule*. The key assumption is then that neither effort provision nor individual contributions to output are contractable, so that sharing rules can condition only on the total level of joint output. As noted by Bartling and von Siemens (2010*a*), partnerships are the prevalent organisational form in many industries, making their study an important economic topic.

We consider the following simple model based on the seminal analysis of Holmström (1982), in which n identical agents each choose an effort level $a_i \in \mathcal{A} = \mathbb{R}_+$. The cost of this effort for each agent is denoted by $c(a_i)$ and is assumed to satisfy the usual restrictions. The efforts of agents jointly determine output according to the deterministic production function $\Pi : \mathbb{R}^n_+ \to \mathbb{R}_+$, which is assumed to be strictly increasing, concave and differentiable. For simplicity, we initially restrict attention to a sharing rule in which the output produced is divided equally between agents, so that each receives $\frac{1}{n}\Pi(a_1, ..., a_n)$.

In this environment, the total surplus from production is:

$$\Pi(a_1, ..., a_n) - \sum_{i=1}^{n} c(a_i) \tag{74}$$

expected loss. For tournaments which are sufficiently 'tight' at the interim stage, there exist equilibria in which the player who is initially behind undertakes strictly higher effort and becomes the favourite to win the tournament. They argue that such 'turnarounds' cannot be explained by the standard theory.

which is simply the joint output produced minus the sum of effort costs. It follows that the Pareto efficient effort level for each agent is defined by the following condition:

$$\frac{\partial \Pi(a_1, ..., a_n)}{\partial a_i} = c'(a_i) \tag{75}$$

so that each agent equates his marginal contribution to total output with his marginal effort costs. However, since agents only receive a fraction of total output, it is straightforward to show that utility maximisation implies effort provision defined by:

$$\frac{1}{n}\frac{\partial \Pi(a_1, ..., a_n)}{\partial a_i} = c'(a_i) \tag{76}$$

so that agents undertake too little effort relative to the Pareto optimal quantity.

Intuitively, a misalignment of incentives occurs since each agent bears the full cost of their effort, while receiving only a fraction of the benefit, leading to an equilibrium underprovision which is reminiscent in nature of the classical economic topic of public goods.[73] In fact, while we have limited our analysis to an equal sharing rule, Holmström (1982) shows that this *free-rider problem* will persist for all sharing rules which are *budget-balancing* (i.e. those which exhaustively distribute the entirety of output between agents). Since individual effort provision is not observable and budget-balancing precludes sharing rules which punish all parties for low output, at least one agent will always have an incentive to shirk.

If we allow for non budget-balancing sharing rules, Holmström (1982) notes that typically many mechanisms can lead to an efficient outcome, since group penalties can then be imposed; one simple example is a mechanism which stipulates a zero payment to each agent unless output is at its optimal level, in which case it will be split equally. However, non budget-balancing sharing rules may be infeasible since in the event of suboptimal output, agents will have an ex post preference to renegotiate the contract rather than 'burn' output, so that potential penalties may not actually be enforced. In this case, the free-riding problem reappears, unless there exists a principal who can act as the residual claimant of suboptimal output and therefore enforce the original contract.

Much of the subsequent literature on team production has studied the ways in which the free-rider problem can be mitigated under equal sharing rules. One such contribution comes from Kandel and Lazear (1992), who study an environment in which agents can subject one another to *peer pressure.* Formally, let us consider the foregoing environment but now additionally assume that there exists a peer

[73] Intuitively, in both cases, underprovision results since the costly inputs of individuals (either effort or monetary contributions) confer positive externalities on other agents. Public good provision has been studied with non-standard preferences (see e.g. Fehr and Schmidt, 1999), with findings which are directly relevant for the analysis of team production. However, this literature lies outside the scope of the current discussion.

pressure function $P_i(a_1, ..., a_n)$ which acts as a cost for each agent, so that utility takes the form:

$$\frac{1}{n}\Pi(a_1, ..., a_n) - c(a_i) - P_i(a_1, ..., a_n) \tag{77}$$

The effort provision of a utility maximising agent is then given by:

$$\frac{1}{n}\frac{\partial \Pi(a_1, ..., a_n)}{\partial a_i} = c'(a_i) + \frac{\partial P_i(a_1, ..., a_n)}{\partial a_i} \tag{78}$$

When $\frac{\partial P_i}{\partial a_i} < 0$, the effort level that solves (78) is strictly greater than the level that solves (76), as shown by Kandel and Lazear (1992). The essence of this assumption is that the negative impact of peer pressure on an agent is reduced by undertaking higher effort; this may be the case if, for example, agents feel guilt from shirking or have a desire to conform to a social norm of high effort provision. Some authors have then extended this analysis to show that inequity averse preferences can lead to similar effects. Intuitively, when agents compare wages net of effort costs, an equal sharing rule implies that differences in payoffs can only result from differences in effort provision. Compassionate agents then dislike shirking (since this leaves them ahead), while envious agents dislike the shirking of others (since this leaves them behind).[74]

Mohnen et al. (2008) consider a model in which two agents contribute to a group task over two periods and are subsequently paid a wage on the basis of total output. Agents are assumed to dislike any differences in aggregate effort provision across the two periods. In such an environment, inequity aversion can partially alleviate the free-rider problem, similar to peer pressure, when agents are able to observe the effort of their colleague after the first period. Intuitively, since agents dislike effort disparity, there is an incentive to choose second period effort such that these differences are reduced. For instance, if agent i undertakes strictly higher effort in period 1, agent j will then work harder in order to reduce the gap in aggregate effort contributions. It then follows that an agent's marginal return to first-period effort is increased, since high contributions have the additional effect of inducing the coworker to increase his second period effort. The result is that equilibria exist in which both agents undertake strictly higher total effort relative to the self-interested case, with the paper's findings being supported by the outcome of a real effort experiment. In a related model, Huck and Rey-Biel (2006) show that output will be maximised when agents move sequentially, rather than simultaneously, since the first-mover is similarly able to induce his coworker to undertake higher effort. As a result, leadership can emerge endogenously as an

[74] Note that compassionate preferences therefore have a *positive* incentive effect in the framework of team production, contrary to our earlier findings for the case of independent incentive contracts.

equilibrium if agents are able to choose the timing of their efforts.

Masclet (2002) shows that the free-rider problem can be similarly alleviated if agents are able to punish one another after observing effort provision. In the self-interested case, no punishments will occur if inflicting them is costly, since agents do not care about the payoffs of others. However, this is no longer true when agents sufficiently dislike inequality, in which case they are willing to incur the costs associated with punishment in order to reduce the payoffs of those who shirk. It follows that punishment becomes a credible threat, which can then induce agents to undertake high effort in equilibrium.[75]

Kölle et al. (2016) explore the case of heterogeneity in productivity between agents, where payoff inequalities will naturally occur since high ability agents undertake relatively more effort, but receive the same wage. As we have seen, inequity aversion then creates incentives to adjust effort provision such that payoff disparities are reduced, so that a high ability agent undertakes less effort relative to the self-interested case. However, this can be avoided if the ex ante wealths of agents are unequal. Specifically, if high ability agents initially have higher incomes than their low ability counterparts, the desire for ex post payoff equality increases their willingness to exert effort, which increases both total output and the welfare of all agents.

The literature we have covered so far has assumed the use of an equal sharing rule to divide output between parties. Bartling and von Siemens (2010*a*) note that such mechanisms are commonly used in practice and argue that inequity aversion may provide a rationale for this observation. The intuition here is simple: with an equal sharing rule, there exists no payoff inequality when all agents undertake the efficient level of effort, while the benefits for any agent who chooses to unilaterally shirk are reduced by compassion. Accordingly, incentives to shirk are low. In contrast, unequal sharing rules imply painful inequality even in the case of efficient effort provision by all agents. Moreover, if envy is stronger than compassion, as is typically assumed, the agent who receives the smallest share of output according to the sharing rule may prefer to shirk — thereby increasing his net wage since effort costs are reduced, and possibly moving ahead of others — rather than undertake high effort and suffer from envy.

While the analysis of Bartling and von Siemens (2010*a*) highlights the fact that compassion can help mitigate the free-rider problem and maintain high effort provision, it may also lead to inefficiently small teams. To see this, suppose there exists a team which employs an equal sharing rule when dividing output among members. Now further suppose that another agent wishes to join the team and that this is efficient (since the resulting increase in output is greater than the

[75] Similar results hold in the context of public good provision when inequity averse agents are able to punish one another; see for instance Fehr and Schmidt (1999).

newcomer's effort costs) but leads to a reduction in the average output per team member. With self-interested agents, there may be a compromise in which all parties benefit: for instance, the newcomer's effort costs could be compensated and the remaining additional output divided equally between existing members. However, such a solution implies payoff differences among team members and therefore may not be feasible when agents are compassionate.

Alternatively, Li (2009) shows that there exists a stochastic, budget-balancing sharing rule which can elicit efficient effort when agents are inequity averse, in contrast to the result of Holmström (1982). The mechanism works as follows. If total output is equal to (or above) the optimal level, it is split equally between agents. Otherwise, if output is below this level, then at least one agent has shirked; in this case, a subset of agents is randomly chosen, who are then subject to the largest fine possible given limited liability constraints. The revenue from these fines, along with output, is then divided equally between the remaining partners. Accordingly, each agent is either 'lucky', in which case they receive a higher income but suffer from negative feelings of compassion, or 'unlucky', which results in both lower income and disutility from envy. For agents who are sufficiently other-regarding, the mechanism therefore enforces efficient effort by imposing massive inequality — and therefore a reduction in expected utility — as a result of shirking. Li (2009) further argues that this sharing rule is most effective when team sizes are small, or when the subset of agents who are chosen for punishment is large.

Beyond inequity aversion, there have been limited applications of reference-dependent preferences to team contracting. Gill and Stone (2015) apply the concept of *just deserts*, whereby agents care about their payoff relative to their perceived entitlement.[76] Specifically, this will depend on how an agent's effort provision compares to that of his teammates: he believes he deserves a higher wage if he works harder, but a lower wage if he works less. They note that desert guilt — whereby agents dislike receiving more than they feel they deserve — is a realistic assumption in the context of team production. In this case, desert preferences lead to endogenous complementarities since agents have an incentive to match the efforts of others. To see this, suppose that all agents initially undertake identical efforts. Each agent then believes they deserve an equal share of the team output and, due to the equal sharing rule, each agent actually receives this share. Now consider a single agent's incentive to deviate. Increasing effort provision results in a reference point above the equal share, while decreasing effort provision leads to a reference point below this amount; in both cases however, the agent continues to receive an equal share so that there is a disparity between his actual payoff

[76]'Desert' preferences were previously discussed in Section 4.3.

and his perceived entitlement, resulting in a utility loss. This disinclination to deviate from a common effort level means that a range of symmetric equilibria are sustainable, both below and above the equilibrium self-interested effort level. Desert preferences of this type therefore generate behaviour consistent with both positive and negative reciprocity and can both exacerbate and alleviate the free-riding problem in teams.[77]

5 Conclusion

As we have seen, the introduction of inequity averse and loss averse preferences into studies of incentive contracting has a significant impact on the outcomes predicted by economic models. This impact can typically be delineated into two distinct effects. First, there is a positive incentive effect which allows for lower-powered explicit incentives in order to implement a given effort level. Second, there is a negative participation effect, whereby reductions in an agent's expected utility must be compensated by a higher expected wage if the contract is to be accepted. The latter effect then implies that the effort level chosen to be implemented by the principal is typically reduced relative to the standard case. Accordingly, the findings of the literature are consistent with empirical evidence, both that wages tend to be compressed within organisations (Prendergast, 1999) and that incentive pay is typically less sensitive to performance than predicted by the standard theory (Jensen and Murphy, 1990).

The aforementioned effects also have an impact on the structure of the optimal contract. The literature which considers inequity aversion has predominantly focused on the extent to which a worker's wage should be dependent on the performance of others, and, if so, whether there is a positive or negative relationship between the two. In particular, firms may find it beneficial to eliminate wage inequalities using team contracts, or exacerbate them through relative performance evaluation or tournament schemes. We have also seen how inequity averse preferences — and in particular, compassion — allow groups of workers who engage in partnerships to partially mitigate the free-rider problem.

In contrast, the literature which studies loss aversion often features an analysis of optimal wage schemes in the presence of a rich performance measure, with a recurrent finding being that payments will typically be invariant to the performance signal over some regions. This unresponsiveness results from the firm's

[77] The finding that desert preferences lead agents to undertake similar efforts is contrary to the findings of Gill and Stone (2010) in the context of tournaments, discussed in Section 4.3, where effort differentiation results. Intuitively, agents with desert concerns prefer outcomes in which relative wages are reflective of relative effort provision. Since wages are necessarily unequal in tournaments, agents are then motivated toward unequal efforts. Likewise, equal sharing rules in partnerships typically induce equal efforts.

desire to shield workers from payments below the reference point and therefore reduce the size of the necessary loss premium. Accordingly, several authors have noted that reference-dependent preferences may provide an explanation for the real-world prevalence of simple wage schemes which typically feature a small number of distinct payments (see for instance Herweg et al., 2010 and the discussion therein).

These central findings have been established numerous times and in many different frameworks. However, there are several interesting questions which have received relatively little attention, or have yet to be studied, when parties are either inequity averse or loss averse. For instance, what are the implications of such preferences for long-term contracting in dynamic environments? How does the outcome of contracting change when workers are required to undertake several different productive tasks? Or when they can control the risk inherent in production? What if they can engage in undesirable activities in order to manipulate the performance measure?

Unfortunately, investigation of such issues is often impeded by our lack of knowledge regarding the exact nature of individual preferences in complex environments. Insights regarding optimal dynamic contracts will crucially depend on workers' attitudes toward intertemporal wage inequalities, or how their reference points adjust over time.[78] There exists little evidence here. Likewise, Macera (2018*a*) notes that our understanding of the relationship between risk aversion and loss aversion is underdeveloped.

This same problem also hinders the study of further issues relating to organisational design. Inequity aversion will likely have implications for managerial decisions, such as the social proximity of workers within firms, wage secrecy policies and the optimal size of departments or teams.[79] However, there is no established theory of how inequity concerns vary as coworkers become more distant, or as the size of the reference group changes.[80] It also seems overly simplistic to assume that wage secrecy will prevent social comparisons from occurring, when workers are likely to be able to infer at least some of the relevant information.

Similarly, while a small amount of papers which assume reference-dependent preferences have investigated how firms may be able to influence the reference

[78]For instance, it is plausible that a worker who earns more than his colleagues in a particular month may not feel particularly bad about this if he received less than them in previous months, since 'long-term' inequality has been reduced.

[79]While we have previously discussed how social proximity can be influenced by the firm's choice of wage scheme, it may also be affected by other factors: whether workers share an office, work similar hours, have the same job title etc.

[80]Suppose a worker receives the low wage w_L, while all others in his reference group of size N receive the high wage w_H. Under the Fehr and Schmidt specification (28), the worker's utility is invariant to changes in N. Yet it seems doubtful that finishing behind one other person inspires identical feelings to finishing behind one hundred others.

point through channels such as design of the wage scheme, the framing of the contract or the firm's organisational culture, any resulting predictions are necessarily speculative due to the dearth of evidence for how reference points change in response to such actions. It seems, therefore, that more evidence regarding the nature of individual preferences is required before we are able to investigate the foregoing issues with any confidence.

In his overview of the broader behavioural contract theory literature, Kőszegi (2014) notes that a common tendency of researchers has been to focus on how the parameters of the *behavioural* theory affect predictions. Yet, economists are ultimately more interested in how outcomes react to changes in the economic environment, since this kind of analysis yields predictions which are both more economically relevant and easier to test. While the foregoing discussion has provided some examples in this direction, there are many other interesting and important issues which have yet to be investigated. Hopefully, subsequent studies will explore these areas and continue to shed light on the outcomes of contracting in various environments.

Finally, it is important to note that the relationship between the theoretical literature and evidence — experimental or empirical — runs in both directions. While future studies can provide data which will continue to inspire theoretical models and help us answer some of the foregoing questions, it will also be important to discern whether the findings of the literature are consistent with real world evidence. In doing so, we will make significant progress toward understanding the nature of incentive contracts and their impact on economic behaviour.

Chapter 2

1 Introduction

The previous chapter explored the existing literature which investigates incentive contracting when workers are averse to either inequity or losses. In the former case, we found that studies typically assume each worker's reference group is confined to others within the firm, such as coworkers, subordinates or superiors. However, as discussed in the introduction to this thesis, there is significant evidence that social comparisons often extend beyond the firm to wider groups in society. In this chapter, we study incentive contracting when preferences are characterised by relative income concerns and workers attempt to keep up with the Joneses.

In order to capture this notion, we introduce a stylised model where workers have an aversion to falling behind a reference wage, which is determined by the economy-wide average income.[1] Moreover, to allow for a detailed comparative static analysis, we assume that workers' preferences can be represented by a utility function which is piecewise linear around this reference wage. The importance of relative income concerns is then captured by the difference between the left-hand and right-hand marginal utilities at this point. Not only is this approach both simple and intuitive, there is also survey evidence to suggest that the average wage in an economy does indeed act as an important point of comparison. For instance, the studies by Solnick and Hemenway (1998), Johansson-Stenman et al. (2002) and Alpizar et al. (2005) present evidence that many individuals would be willing to sacrifice a significant amount of income in order to improve their relative position in society and, in particular, move above the average income level.[2]

[1]This chapter is an earlier version of the paper 'Moral Hazard and Keeping up with the Joneses' (Demougin and Upton, 2019) and was cowritten with Dominique Demougin.

[2]These papers each ask individuals to express their preferences over hypothetical outcomes, which differ with respect to both their own income and the average income in society. For instance, Solnick and Hemenway (1998) ask respondents questions of the following nature:

> Would you rather live in Society A (you earn $50,000; others earn $25,000) or Society B (you earn $100,000; others earn $200,000)?

It is explained that prices are at their current level and that the purchasing power of money is equal in each society. They find that approximately half of respondents would prefer to live

The first part of this chapter considers the contracting problem of a single firm-worker pair who take the reference wage as exogenously given. Specifically, we solve for the wage scheme which minimises the firm's costs of implementing a particular effort level in the presence of moral hazard. The resulting contract is found to take either a binary or ternary form. The former is shown to be optimal when all payments are below the worker's reference wage, in which case utility is linear over the relevant region and the resulting contract behaves essentially as in the standard risk neutral case. We find that when the distribution function satisfies an additional regularity requirement, this occurs precisely when effort is sufficiently low. In the ternary case, depending on the outcome of a continuous performance measure, the principal either pays the smallest possible payment given limited liability, exactly the worker's reference wage or a bonus which is strictly greater than both. The frequency of the respective payments depends on both the degree of the worker's relative income concerns and on the quality of the performance measure.

Next, we present a detailed comparative static analysis and show that the firm benefits from contracting with workers who either have stronger relative income concerns, or evaluate their earnings relative to a higher reference wage. Intuitively, aversion to falling behind a reference wage creates a positive incentive effect, similar to the cases of inequity aversion and loss aversion. When limited liability constraints imply that the worker extracts a rent from the relationship, any participation effect plays no role so that, altogether, effort can be implemented at a lower cost for the firm. We also show that the firm's costs are increasing in the effort to be implemented, but decreasing in the quality of monitoring.

The last part of the chapter embeds the employment relationship into a replica economy populated by a continuum of identical firm-worker pairs. The reference wage is then assumed to be determined endogenously by the average income of workers, allowing us to further the idea of keeping up with the Joneses. We establish some additional comparative static results, before letting effort be chosen endogenously by firms to maximise profit. Using a series of numerical examples, we then examine how changes in the parameters of the model affect the economy's equilibrium. In particular, we highlight an externality effect whereby firms do not take into account the impact of their contracting decisions on the economy's average wage; it follows that firms would be able to increase their profits if they were to undertake these decisions collectively. In addition, we find that higher relative income concerns are associated with a lower average wage, as well as a reduced level of inequality as measured by the Gini Coefficient.

Our analysis is closely related to papers which consider incentive contracting

in a society in which they have 50% less income, so long as their relative income is high (i.e. expressing a preference for Society A in the foregoing example).

when a firm employs several inequity averse workers who dislike wage inequalities. As discussed in the previous chapter, a key issue in this literature is the extent to which remuneration of one worker should depend on the performance of others. Many studies investigate how contracts can be designed in order to either exacerbate or eliminate inequality, using wage schemes such as team contracts (Itoh, 2004; Demougin and Fluet, 2006; Goel and Thakor, 2006), relative performance evaluation (Bartling, 2011; 2012*a*) or tournaments (Grund and Sliwka, 2005; Demougin and Fluet, 2003). In contrast to these works, we are interested in wider social comparisons where the reference group consists of a large number of others and extends beyond the firm. Moreover, we limit attention to independent contracts where a worker's wage is dependent only on his own performance level.

Since the first part of the chapter considers contracting between a firm-worker pair where the reference wage is taken as exogenously given, our analysis also bears a resemblance to papers which study incentive contracting when workers' preferences are characterised by reference-dependence and loss aversion. The main focus of these models has been the study of the optimal wage scheme when the firm has access to a rich performance measure, with loss aversion implying that wage schemes exhibit some degree of payment insensitivity, violating Holmström's (1979) sufficient statistic result (de Meza and Webb, 2007; Herweg et al., 2010). However, this literature typically assumes that an individual's reference point is solely determined by his contract, rather than as a result of any wider social comparisons.

The remainder of the chapter is structured as follows. Section 2 introduces the model, while Section 3 describes the firm's design problem and shows that the optimal contract takes either a binary or ternary form. Section 4 describes the key properties of these contracts and analyses when each will be used. Section 5 provides a comparative static analysis of the ternary contract, while Section 6 extends the model to allow for an endogenous reference wage. Section 7 concludes. All proofs are relegated to the appendix of the chapter.

2 Setup

We consider the contracting problem between a risk-neutral firm (the principal) and a worker whose preferences are characterised by relative income concerns (the agent). The firm owns a production technology and wishes to employ the worker to undertake a certain level of effort, denoted by $a > 0$.[3] We solve for the firm's optimal contract which implements the desired effort level while minimising the expected wage. The worker's preferences are represented by a utility function which

[3] In Section 6, we introduce a function $v(a)$ which measures the firm's value of the worker's effort and let a be determined endogenously.

is separable in wage payment (w) and effort, $U(w, a; \alpha, w^R) = u(w; \alpha, w^R) - c(a)$, where w^R is an exogenously given reference wage. The function $c(a)$ captures the worker's disutility of undertaking effort a and is assumed to satisfy the standard regularity requirements: $c(0) = c'(0) = 0$, $c' > 0$ and $c'' > 0$. The worker's preferences over wage payments are represented by a piecewise linear function around w^R:

$$u(w; \alpha, w^R) = \begin{cases} w + \alpha(w - w^R) & \text{if } w < w^R \\ \\ w & \text{if } w \geq w^R \end{cases} \tag{1}$$

where the parameter $\alpha > 0$ measures the worker's aversion to falling behind the reference wage, referred to hereafter as the KUJ parameter. We follow the standard literature and assume that the firm holds all of the bargaining power, offering the worker a take-it-or-leave-it contract. If the worker rejects the contract, he undertakes no effort and receives no wage implying the utility level $U(0, 0; \alpha, w^R) = -\alpha w^R$. In addition, the worker is financially constrained so that wage payments are restricted to be non-negative in all states of the world.

Moral hazard occurs because effort is non-verifiable. Instead, the firm has access to the following monitoring technology: after the worker has produced effort a, monitoring generates a proxy variable $x \in [0, 1]$ with an exogenous probability $m \in (0, 1)$ and no information, denoted by $\emptyset$, with probability $(1 - m)$.[4] It is common knowledge that the realisation of x is drawn from a thrice-differentiable distribution function $F(x; a)$ with density $f(x; a) > 0$ over the support.[5] In order to guarantee the validity of the first-order approach, we require that F satisfies the strict Monotone Likelihood Ratio Property (MLRP) and the strict Convexity of the Distribution Function Condition (CDFC), as is standard in the moral hazard literature (Rogerson, 1985).[6]

If monitoring does not generate information, all parties observe $\emptyset$. Otherwise, the firm can choose to make the private signal x available, in which case it becomes verifiable. Alternatively, the firm can hide the realisation of x so that all other parties observe $\emptyset$ (*i.e.* the worker and, potentially, third parties cannot distinguish between situations where the monitoring technology did not generate any information and those in which the firm hides a disfavourable outcome). This

[4]For instance, suppose x is the output of an electronic device and $(1-m)$ the probability that the device breaks down. A similar model with a different focus appears in Bental et al. (2016), aimed at analysing the implications of incomplete contracts for legal design.

[5]We require that F has third derivatives for the comparative static analysis.

[6]Strict MLRP states that $\frac{f_a}{f}(x; a)$ is strictly increasing in x for all $x \in (0, 1)$; it also implies first-order stochastic dominance (*i.e.* $F_a(x; a) \leq 0$ for all x). Strict CDFC states that $F_{aa}(x; a) > 0$ for all $x \in (0, 1)$. See LiCalzi and Spaeter (2003) for a discussion of distributions which satisfy these conditions.

introduces an additional moral hazard issue on the side of the firm, who will announce the realisation of x only if it is advantageous to do so, thereby restricting the set of feasible contracts as described in the subsequent section.

Before solving the model, a couple of remarks are in order. First, with respect to the specification of monitoring and, in particular, the assumption $m < 1$. As is well known from the principal-agent literature, when the agent is risk-neutral and the principal has access to information of the form $x \in [a, b]$, the optimal contract is binary with a critical value equal to b; see e.g. Kim (1995; 1997). That is, the bonus is paid only for the highest possible performance level, which in the case of a continuous density function occurs with probability zero. In order to satisfy incentive compatibility, this in turn implies an unbounded bonus. This peculiar result also occurs when the wage-utility function is piecewise-linear; assuming that monitoring may fail eliminates this possibility.

Second, our specification implies that the worker's outside utility level varies with both the reference wage and the worker's KUJ parameter. We view this as being consistent with our interpretation of relative income concerns, as a worker who chose to reject the contract would continue to compare their earnings with the reference wage. This is also coherent with our specification in Section 6, where the reference wage is endogenously determined by the equilibrium average wage in the economy. Our formalisation is in contrast to the literatures which consider incentive contracting in the presence of inequity aversion or loss aversion, whereby workers either make intra-firm wage comparisons, or compare their wage to some reference point which is typically determined endogenously by the outcome of contracting. Accordingly, in both of these cases, it is logically consistent to consider an exogenously given outside utility.

3 Incentive Feasible Contracts

In this environment, an *incentive feasible* contract is a triplet $\mathcal{C} = \{a, w_\emptyset, w(x)\}$ with the following characteristics. First, it specifies the required effort level $a > 0$. Second, it conditions the worker's remuneration $w(x)$ on the verifiable monitoring output $x \in [0, 1]$ and stipulates the wage $w_\emptyset$ to be paid when this information is not generated or is hidden by the firm. Third, the contract satisfies the following conditions; (*i*) all wage payments are non-negative; (*ii*) $\mathcal{C}$ induces the worker to participate; (*iii*) $\mathcal{C}$ induces the worker to exert effort a, anticipating that the firm will not conceal the outcome of monitoring; (*iv*) $\mathcal{C}$ motivates the firm to never conceal the outcome of monitoring. With respect to the last condition, note that in its absence, the firm would only announce the realisation of x if $w(x) \leq w_\emptyset$. Anticipating this, the worker understands that any wage $w(x) > w_\emptyset$ would never actually be paid. Instead, for any realisation x, the parties would expect the wage $\min\{w(x), w_\emptyset\}$ to be paid. Clearly this wage scheme satisfies (iv) so that we can *without loss of generality* restrict attention to wage schemes such that $w(x) \leq w_\emptyset$, $\forall x \in [0, 1]$.

Suppose the firm wishes to implement a given effort level $a > 0$. Accordingly, they will choose the incentive feasible contract $\mathcal{C}$ which minimises expected costs. Mathematically, that contract is the solution to the optimisation problem:

$$C^F(a; w^R) = \min_{w_\emptyset, w(x)} \ (1-m)w_\emptyset + m\int_0^1 w(x)f(x;a)dx \tag{I}$$

$$a = \arg\max_{\hat{a}} \ (1-m)u(w_\emptyset; \alpha, w^R) + m\int_0^1 u(w(x); \alpha, w^R)f(x;\hat{a})dx - c(\hat{a}) \tag{IC}$$

$$(1-m)u(w_\emptyset; \alpha, w^R) + m\int_0^1 u(w(x); \alpha, w^R)f(x;a)dx - c(a) \geq -\alpha w^R \tag{PC}$$

$$w_\emptyset \geq w(x), \quad \forall x \in [0, 1] \tag{FFC}$$

$$w_\emptyset, w(x) \geq 0, \quad \forall x \in [0, 1] \tag{WFC}$$

where the (IC), (PC) conditions and the Worker's Financial Constraint (WFC) respectively ensure that the worker undertakes the effort a, that he accepts the contract and that payments are non-negative. The Firm's Feasibility Constraint (FFC) guarantees that the outcome of monitoring is never concealed. Problem (I) is solved in the appendix where we maximise the negative of the objective function and apply the first-order approach (*i.e.* we substitute the first-order condition of the worker's optimisation problem for IC). Moreover, we show that (PC) is never binding due to (WFC) and can thus be ignored.

Proposition 1 summarises the solution to (I) and introduces the notation applied in the remainder of the paper. We distinguish between three different contracts, for the cases where the highest wage payment is strictly less than, equal to, and strictly greater than the worker's reference wage, respectively.

Proposition 1. *For any $a > 0$, $\alpha > 0$, $m \in (0, 1)$ and $w^R > 0$, there exists an optimal contract which takes one of three possible forms:*

- *A* ***strict binary*** *contract with*

$$w_B(x) = \begin{cases} 0 & 0 \leq x < z_B \\ B_B & z_B \leq x \leq 1 \end{cases} \tag{2}$$

where z_B *is a critical value which partitions the support into two subintervals,* B_B *satisfies* $0 < B_B < w^R$ *and* $w_\emptyset = B_B$*.*

- *An* ***intermediate*** *contract with*

$$w_I(x) = \begin{cases} 0 & 0 \leq x < z_I \\ w^R & z_I \leq x \leq 1 \end{cases} \tag{3}$$

where z_I *is a critical value which partitions the support into two subintervals and* $w_\emptyset = w^R$*.*

- *A* ***strict ternary*** *contract with*

$$w_T(x) = \begin{cases} 0 & 0 \leq x < z_1 \\ w^R & z_1 \leq x < z_2 \\ B_T & z_2 \leq x \leq 1 \end{cases} \tag{4}$$

where $z_1 < z_2$ *are critical values which split the support into three subintervals,* B_T *satisfies* $w^R < B_T$ *and* $w_\emptyset = B_T$*.*

Note that while the intermediate contract can be thought of as the limit case of either the strict binary or the strict ternary contract, we introduce the distinction for the sake of the ensuing comparative static analysis.

The use of a *strict binary* contract has a straightforward interpretation in terms of the existing literature when the bonus required in order to align incentives is less than w^R, *i.e.* when effort is sufficiently small. In these cases, the worker's utility function is linear over the relevant region and the firm's optimisation program becomes similar to that of the standard risk-neutral agency problem.[7] Applying the intuition from Kim (1995), risk-neutrality induces a binary contract characterised by a critical z_B where low performance, defined by $x < z_B$, is rewarded with a zero payment, while the worker receives a bonus payment B_B in the case of high performance (*i.e.* $x \geq z_B$).

On the other hand, when effort is sufficiently large, the bonus payment is necessarily strictly greater than w^R, thereby crossing the worker's reference wage so that the relevant portion of the wage-utility function is piecewise linear, and

[7] The key difference is that in the current case, marginal utility is equal to $(1 + \alpha)$ instead of 1 for the standard risk-neutral case.

non-differentiable when the wage is equal to w^R. Solving (I) therefore requires subdifferential calculus to address this issue.[8] We show in the appendix that the relevant first-order condition with respect to $w(x)$ takes the form

$$-mf(x;a) + \lambda m \left[1 + \gamma(x)\alpha\right] f_a(x;a) + \xi(x) - \zeta(x) = 0, \ \forall x \tag{5}$$

where $\xi(x)$, $\zeta(x)$ are Lagrange variables associated with (WFC) and (FFC) while $\gamma(x)$ is a multiplier dealing with the non-differentiability at w^R.[9] $\gamma(x)$ is equal to 1 when $w(x) < w^R$, equal to 0 when $w(x) > w^R$ and can take any value over the unit interval at $w(x) = w^R$.[10] Taking into account that the Lagrange variables are non-negative, (5) implies that the wage prescribed for any given x depends upon the direction of the inequality

$$\lambda \left[1 + \gamma(x)\alpha\right] \frac{f_a(x;a)}{f(x;a)} \lesseqgtr 1 \tag{6}$$

Figure 1 provides a graphical interpretation of this condition. First, consider the case where $\xi(x) > 0$, which immediately implies $w(x) = 0$ so that $\zeta(x) = 0$ and $\gamma(x) = 1$. Substituting these values into (5) and rearranging implies that the LHS of (6) becomes strictly less than the RHS. Geometrically, this requires $x < z_1$. A symmetric argument applies when $\zeta(x) > 0$; accordingly, $w(x) = B_T > w^R$. Here, the LHS is strictly larger than the RHS and $\gamma(x) = 0$. Graphically, this is only feasible for $x > z_2$. Finally, when $\xi(x) = \zeta(x) = 0$, (6) holds with equality. In Figure 1, this corresponds to realisations of x at the intersection of the horizontal bold line and the grey shaded area. Analytically, that intersection is characterised by $0 < \gamma(x) < 1$ *a.e.* so that $w(x) = w^R$.[11]

From the foregoing, we know that the firm will use a strict binary contract for sufficiently low effort levels and a strict ternary contract for sufficiently high effort levels. By continuity, there must be at least one intermediate effort level where the bonus takes the value w^R. In Figure 1, it was assumed that $\lambda \frac{f_a(1;a)}{f(1;a)} > 1$. Suppose now that $\lambda \frac{f_a(1;a)}{f(1;a)} \leq 1$ so that the bold line does not intersect with $\lambda \frac{f_a(x;a)}{f(x;a)}$ over the support of x. In this case, the only possible wage payments are 0 or w^R and the result of this is the intermediate contract described in Proposition 1. Although strictly speaking this contract is binary, we differentiate it from the strict binary

[8]Subdifferential calculus is a technique for solving convex optimisation problems in the presence of non-differentiability. For a general reference, see Nesterov (2004).

[9]For a similar approach, see de Meza and Webb (2007).

[10]The expression $1 + \gamma(x)\alpha$ therefore represents the superdifferential of $u(w;\alpha,w^R)$. For $w \neq w^R$ it is exactly equal to the usual derivative. However, for $w = w^R$ it can take any value in the closed interval $[1, 1+\alpha]$ and is a convex combination of the left and right derivatives at that point; accordingly, $\gamma(x) \in [0,1]$.

[11]$x = z_1$ and $x = z_2$ are measure zero events and the value of $w(x)$ at these two points is inconsequential. The case where $\xi(x) > 0$, $\zeta(x) > 0$ is shown in the appendix to never occur.

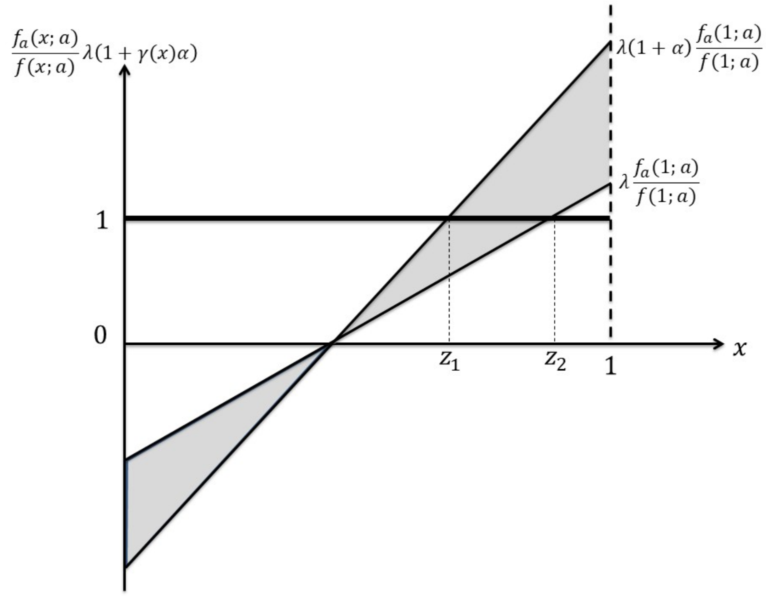

Figure 1: The feasible region of $\lambda\left[1+\gamma(x)\alpha\right]\frac{f_a(x;a)}{f(x;a)}$

case since the determination of the critical value z_I and the comparative statics of the contract are different.

4 Optimal Contracts

Proposition 1 tells us that when looking for the firm's optimal contract to implement a, we can restrict our search to three different types of wage scheme. In this section, we solve for the optimal payments and associated critical value(s) for each contractual form on a case-by-case basis. We proceed as follows. For the *strict binary* and *strict ternary* cases, we solve for the respective bonus payments B_B and B_T using the worker's incentive compatibility constraint and determine the corresponding critical values of the performance level in order to minimise costs. For the *intermediate* contract, the critical value z_I must be chosen to satisfy the incentive compatibility constraint since wage payments are fixed at zero and w^R. We then return to the question of when each contract will be chosen.

4.1 Strict Binary Contract

In this subsection, we suppose that the firm uses a strict binary contract (z_B, B_B). In this case, the worker's incentive compatibility requirement becomes:

$$a = \arg\max_{\widehat{a}} \; -mF(z_B;\widehat{a})\alpha w^R + \left[1-mF(z_B;\widehat{a})\right]\left[(1+\alpha)B_B - \alpha w^R\right] - c(\widehat{a}) \quad (7)$$

Hence, for any z_B, implementing effort a requires that the bonus payment satisfies:

$$B_B = \frac{c'(a)}{-mF_a(z_B; a)(1+\alpha)} \tag{8}$$

Accordingly, the firm will select z_B in order to minimise expected costs. However, this is only compatible with the worker's preferences if the resulting bonus satisfies $B_B < w^R$. Altogether, the firm's expected costs are given by:

$$\begin{aligned} C_B^F\left(a; w^R\right) = \min_{z_B} \quad & \frac{1 - mF(z_B; a)}{-mF_a(z_B; a)(1+\alpha)} c'(a) \\ \text{s.t.} \quad & \frac{c'(a)}{-mF_a(z_B; a)(1+\alpha)} < w^R \end{aligned} \tag{9}$$

Whenever the minimisation problem (9) has a solution, the optimal critical value z_B^* is implicitly defined by its first order condition:

$$mF_a(z_B; a) f(z_B; a) + \left[1 - mF(z_B; a)\right] f_a(z_B; a) = 0 \tag{10}$$

In the alternative case, we set $C_B^F\left(a; w^R\right) = +\infty$. This ensures that $C_B^F\left(a; w^R\right)$ is well-defined over $\mathbb{R}_{++}$ for all values of a. Following a similar approach for the intermediate and strict ternary cases, the firm's optimal wage scheme can then be found by minimising across the three contractual forms, whereby existence of a finite solution is guaranteed by Proposition 1.

4.2 Intermediate Contract

By definition, the intermediate contract can be thought of as a limit case of either the strict binary contract with $B_B = w^R$ or of the strict ternary contract with $B_T = w^R$. With either interpretation, the bonus is exogenously fixed at w^R resulting in the worker's incentive compatibility requirement:

$$a = \arg\max_{\widehat{a}} \; -mF(z_I; \widehat{a})\alpha w^R + [1 - mF(z_I; \widehat{a})] w^R - c(\widehat{a}) \tag{11}$$

Hence, implementing effort a with the intermediate contract necessitates defining z_I^* as the implicit solution to the equation:

$$c'(a) + mF_a(z_I; a)\,(1+\alpha)\, w^R = 0 \tag{12}$$

If (12) has a solution, the firm's costs of implementing effort become:

$$C_I^F\left(a; w^R\right) = \left[1 - mF\left(z_I^*; a\right)\right] w^R \tag{13}$$

Otherwise, when z_I^* is not well-defined, applying a similar logic to the foregoing subsection we define $C_I^F\left(a; w^R\right) = +\infty$.

4.3 Strict Ternary Contract

Finally, suppose the firm uses a strict ternary contract with $B_T > w^R$. The worker's incentive compatibility constraint is then:

$$a = \arg\max_{\widehat{a}} \; -mF(z_1;\widehat{a})\alpha w^R + m[F(z_2;\widehat{a}) - F(z_1;\widehat{a})]w^R + [1 - mF(z_2;\widehat{a})]B_T - c(\widehat{a}) \tag{14}$$

Therefore, given critical values z_1 and z_2 the bonus payment takes the form:

$$B_T = w^R + \frac{c'(a) + m(1+\alpha)F_a(z_1;a)w^R}{-mF_a(z_2;a)} \tag{15}$$

which implies that the cost of inducing effort becomes:

$$C_T^F\left(a; w^R\right) = \min_{z_1,z_2} \; [1 - mF(z_1;a)]w^R + \frac{1 - mF(z_2;a)}{-mF_a(z_2;a)}\left[c'(a) + m(1+\alpha)F_a(z_1;a)w^R\right]$$

$$\text{s.t.} \quad c'(a) + m(1+\alpha)F_a(z_1;a)w^R > 0 \tag{16}$$

$$z_2 > z_1$$

Following the same convention as above, we define $C_T^F\left(a; w^R\right) = +\infty$ in the case that (16) does not have a solution. Alternatively, the critical values z_1^* and z_2^* are implicitly defined by the following block-recursive equation system:

$$mF_a(z_2;a)f(z_1;a) + [1 - mF(z_2;a)](1+\alpha)f_a(z_1;a) = 0 \tag{17}$$

$$mF_a(z_2;a)f(z_2;a) + [1 - mF(z_2;a)]f_a(z_2;a) = 0 \tag{18}$$

Note that the equations (17) and (18) are independent of w^R. Combining them yields the following:[12]

$$(1+\alpha)\frac{f_a(z_1^*;a)}{f(z_1^*;a)} = \frac{f_a(z_2^*;a)}{f(z_2^*;a)} \tag{19}$$

[12] Equation (19) can also derived directly from the equations (31) and (32) in the proof of Proposition 1.

The conditions (17)–(19) allow us to make some observations. First, strict stochastic dominance – $F_a < 0$ for all $x \in (0,1)$ – implies $f_a(z_1^*; a),\ f_a(z_2^*; a) > 0$, which verifies our earlier representation in Figure 1. Second, the block-recursive structure of the equation system means that z_2^* is determined independently of the worker's KUJ parameter α. Third, consider the case $\alpha = 0$. Strict MLRP and (19) then yields $z_1^* = z_2^*$ which verifies the well-known result that in the absence of relative income concerns the optimal contract is binary.

To conclude this subsection, note that as discussed above the intermediate contract can also be interpreted as the limit case of the strict ternary contract. In that case, substituting $c'(a) + m(1+\alpha)F_a(z_1; a)w^R \geq 0$ for the constraint in (16), the boundary case yields $z_1^* = z_I^*$ and from (15) $B_T = w^R$. Even though z_2^* plays no role here, it can nevertheless be defined by (18). In the remainder, we refer to a contract as being *weakly ternary* if it takes either a strict ternary or an intermediate form.

4.4 The Firm's Costs of Inducing Effort

By the results of Proposition 1 and the firm's respective costs of inducing effort using the three types of contract defined in the previous subsections, we can immediately conclude

$$C^F\left(a; w^R\right) = \min\left\{C_B^F\left(a; w^R\right), C_I^F\left(a; w^R\right), C_T^F\left(a; w^R\right)\right\} \tag{20}$$

While this definition is very intuitive, it is also quite cumbersome; it requires solving three minimisation problems and comparing the ensuing costs for each possible effort level. A more practical approach would be to find a characterisation of the respective sets of effort levels associated with each contract type. With this in mind, let $\mathcal{A}_\mathcal{B}$ denote the set $\left\{a > 0 \mid C^F\left(a; w^R\right) = C_B^F\left(a; w^R\right)\right\}$ and define $\mathcal{A}_\mathcal{I}$, $\mathcal{A}_\mathcal{T}$ analogously.

Lemma 1. *The sets $\mathcal{A}_\mathcal{B}$, $\mathcal{A}_\mathcal{I}$ and $\mathcal{A}_\mathcal{T}$ satisfy:*

i) $\mathcal{A}_\mathcal{B} = \left\{a > 0 \mid C_B^F\left(a; w^R\right) < +\infty\right\}$,

ii) $\mathcal{A}_\mathcal{T} = \left\{a > 0 \mid C_T^F\left(a; w^R\right) < +\infty\right\}$,

iii) $\mathcal{A}_\mathcal{B}, \mathcal{A}_\mathcal{I}, \mathcal{A}_\mathcal{T}$ *partition* $\mathbb{R}^{++}$.

Recall that $C_B^F\left(a; w^R\right) < +\infty$ whenever the associated optimisation problem has a solution. The key insight behind part *i)* of the result is that the firm prefers to offer a bonus less than w^R whenever possible. Intuitively, this obtains because from the worker's point of view, the resulting impact of a marginal increase in the bonus is $(1+\alpha)$ rather than 1. Hence, the marginal power of the additional

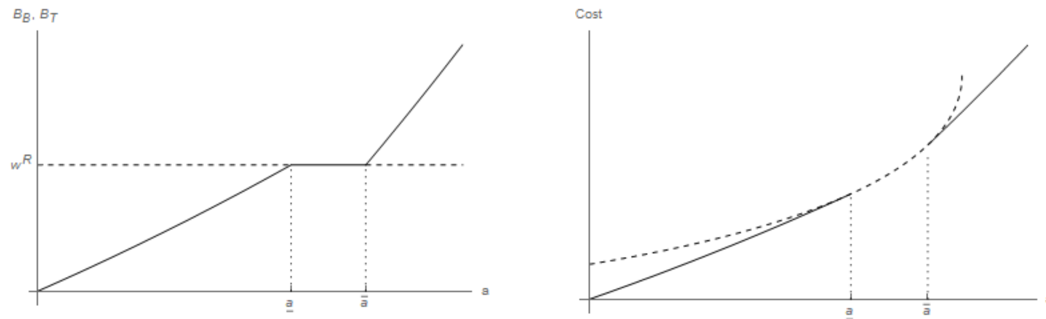

Figure 2: The optimal bonus payment and the function $C^F(a; w^R)$ as effort increases.

payment is larger when the bonus is less than w^R. With this in mind, part *i*) simply states that the set of effort levels for which the strict binary contract defined by (8) and (10) solves the original problem (I) is exactly equal to the set of effort levels for which (9) has a solution. In the proof of the Lemma, we show that a symmetric result holds for the set of strict ternary contracts. Together, these results imply that the two sets are disjoint and also allows us to conclude:

$$\mathcal{A}_{\mathcal{I}} = \left\{ a > 0 \mid C_B^F \left(a; w^R\right) = C_T^F \left(a; w^R\right) = +\infty \right\}$$

In line with the remark at the end of the previous subsection, we also introduce $\mathcal{A}_{\mathcal{WT}} = \mathcal{A}_{\mathcal{I}} \cup \mathcal{A}_{\mathcal{T}}$ as the set of effort levels associated with a weak ternary contract.

One would like to better understand the structure of these sets and in particular determine whether they are intervals. Unfortunately, without further restrictions on the distribution function, we cannot say how the critical values and bonuses change with the level of effort. To see why this is the case, observe that the respective equations which define the critical values contain the function f_a. Hence, evaluating the impact of a variation in a necessarily involves third derivatives of F. The next result provides one example of such a condition.[13]

Lemma 2. *Suppose the distribution function satisfies* $\frac{\partial}{\partial a}\left(\frac{F_a}{f_a}\right) \leq 0$ *for all* $x \in [0, 1]$. *Then there exists a critical effort level* $\underline{a}$ *such that* $a \in \mathcal{A}_{\mathcal{B}}$ *if* $a < \underline{a}$ *and* $a \in \mathcal{A}_{\mathcal{WT}}$ *otherwise.*

Under the condition of the Lemma, we show in the appendix that the bonus payment for a strict binary contract is increasing in a. Hence, such a contract is only valid for low effort levels, up until $\underline{a}$, at which point the bonus payment becomes equal to w^R. Accordingly, for $a \geq \underline{a}$ a weak ternary contract must be used.

[13]This condition is for instance satisfied by both classes of distribution functions described by LiCalzi and Spaeter (2003).

Figure 2 shows the typical evolution of the contract as a increases. The graphics have been derived for the case where $F(x;a) = x + \frac{x-x^2}{a+1}$ and $c(a) = \frac{1}{2}a^2$. The parameters chosen are $\alpha = 1$, $w^R = 5$ and $m = 0.8$. The left hand panel shows the contract's highest payment. Specifically, for $a < \underline{a}$ the optimal contract is a strictly binary with a bonus payment B_B less than w^R. From the foregoing, B_B is increasing until it becomes equal to w^R at the point $a = \underline{a}$. Between the effort levels $\underline{a}$ and $\overline{a}$, the intermediate contract is optimal. Hence, over that region the bonus payment is constant at w^R and variations in effort incentives are provided by adjusting the critical value z_I^*; we return to this point in the next section. For $a > \overline{a}$, minimising incentive costs entails a strict ternary contract with a bonus payment B_T larger than w^R.

The right hand panel plots $C_B^F(a;w^R)$, $C_I^F(a;w^R)$ and $C_T^F(a;w^R)$ for all values of a where the associated contracts are feasible. Specifically, from the foregoing $C_B^F(a;w^R)$ exists only for $a < \underline{a}$. Similarly, $C_T^F(a;w^R)$ is only feasible for $a > \overline{a}$. In the graphic, these functions are represented by the solid curve segments. Finally, the dashed curve illustrates all effort levels for which an intermediate contract is feasible and plots the associated costs $C_I^F(a;w^R)$. By definition, $C^F(a;w^R)$ is the lower envelope of these three curves. The graphics exemplify the findings of Lemma 1; whenever the strict binary or strict ternary contracts exist, they will minimise costs. In all other cases, the firm will use the intermediate contract.

5 Comparative Statics

In this section, we provide some comparative static results for the model. We omit the analysis of the strict binary contract since this case behaves similarly to the risk-neutral agency model which has been previously studied in the existing literature (see for instance Demougin and Fluet, 2001). The next four Propositions consider situations where the optimal contract takes the strict ternary form throughout the comparative statics exercise and examine the respective impacts of variations in the underlying parameters. The last result of this section provides analogous findings for the intermediate contract.

Proposition 2. *An increase in the worker's KUJ parameter α allows the firm to implement the desired level of effort at a lower cost. Moreover, z_1^* decreases while z_2^* remains unchanged. Finally, the bonus payment is reduced.*

As emphasised earlier, equation (18) is independent of α so that z_2^* must be invariant to changes in the worker's KUJ parameter. This invariance implies that the RHS of (19) remains constant. Accordingly, an increase in α on the LHS of that equation must be compensated by a reduction in the likelihood ratio at the point z_1^*. By strict MLRP, this implies that z_1^* decreases. The effect of a variation in α on the bonus can be understood as follows. Applying the envelope theorem to (16) yields $\frac{\partial C_T^F}{\partial \alpha} < 0$. However, since z_2^* is constant, the bonus is paid as often as before whereas the change in z_1^* implies that the reference wage is paid more often. Accordingly, the reduction in costs implies that the bonus payment must decrease. Intuitively, this is required in order to keep the worker's effort incentives constant.

Proposition 3. *An increase in the worker's reference wage w^R allows the firm to implement the desired level of effort at a lower cost. The optimal critical values z_1^* and z_2^* are unchanged, while the bonus payment is reduced.*

From the equations (17) and (18), it is immediate that the critical values z_1^* and z_2^*, and therefore the respective probabilities of obtaining the reference wage and the bonus, are independent of w^R. With this in mind, *ceteris paribus* an increase in w^R raises the worker's effort incentives because it marginally reduces the utility associated with a bad outcome ($x < z_1^*$) while increasing utility in the case of an intermediary result ($z_1^* \leq x < z_2^*$). Therefore, keeping the worker's motivation constant requires a reduction in the bonus. Technically, the proof shows that the second effect dominates the former so that expected wage costs go down.

Intuitively, the worker's wage-utility function $u(w; \alpha, w^R)$ is strictly decreasing in both α and w^R below the reference wage and unchanged elsewhere. These changes therefore enhance the firm's ability to punish the worker for poor performance at no extra cost, since he either becomes less satisfied from not meeting his reference wage following an increase in α, or finds himself further behind the higher reference wage following an increase in w^R. This allows for a reduction in payments elsewhere. As such, in our model, relative income concerns result in a positive incentive effect, similar to previous findings in the context of inequity aversion and loss aversion.[14] Moreover, since the worker's financial constraint implies that he will extract a rent from the relationship, any effect on the participation constraint plays no role. Accordingly, the positive incentive effect results in a reduction of the firm's costs.

While the above findings are consistent with models which assume either inequity averse or loss averse preferences, it should be noted that this is no longer necessarily the case if workers face unlimited liability. In this event, the participation constraint will bind; as discussed in the previous chapter, this leads to a negative participation effect following increases in either inequity aversion or loss aversion. However, since in our framework the worker's outside utility is decreasing in both α and w^R, we predict that the participation effect would be positive

[14] See, for instance, Bartling and von Siemens (2010*b*) and Herweg et al. (2010) respectively, as well as the discussion in the previous chapter.

in our environment, leading to a further reduction in the firm's costs. Intuitively, since social comparisons extend beyond the firm and therefore beyond the employment relationship, changes in the worker's preferences over relative income also impact the utility associated with his outside option, relaxing the participation constraint and lowering the firm's costs.

Proposition 4. *An increase in m, which measures the probability that monitoring generates information, allows the firm to implement the desired level of effort at a lower cost. The optimal critical values z_1^* and z_2^* increase, while the change in the bonus payment is ambiguous.*

Mathematically, the effects on the critical values z_1^* and z_2^* follow from applying the implicit function theorem on the equation system (17)-(18). To provide an intuition for this result, we first consider the case of a strict binary contract. In that case, when the firm chooses the critical value z_B^*, they face a simple trade-off. On the one hand, rewarding only the highest levels of performance is the most effective way to induce effort, but this requires a high bonus payment in order to satisfy incentive compatibility. On the other hand, due to imperfect monitoring, there is a possibility that no information is observed in which case the firm's feasibility requirement imposes paying the high reward. This waters down the effectiveness of the wage scheme and increases costs, creating incentives to reduce the critical value and lower the bonus. The same mechanism is at work with a strict ternary contract, though the intuition becomes more cumbersome because there are now two rewards, w^R and B_T, associated with the respective critical values z_1^* and z_2^*.

With respect to the firm's expected costs, it is straightforward to see that they cannot increase in the frequency with which the worker is monitored. For an intuitive justification, consider the case where m increases to $m + \epsilon$. The firm could commit to ignoring any received information with probability ϵ, in which case the mechanism would become equivalent to the one induced by the original contract, prior to the increase in the monitoring variable. Finally, we observe that the change in the bonus payment is ambiguous. For m close to zero, the effort level a can only be induced by offering a relatively large reward, since there is a low probability that effort influences the worker's wage. As m increases, this reward can be reduced. At the other extreme, when m converges to one, the increase in the critical values z_1^* and z_2^* implies that the bonus will become infinite.

Proposition 5. *An increase in the level of effort induced requires higher costs on the part of the firm.*

As emphasised in the discussion immediately preceding Lemma 2, without additional restrictions on the distribution function we are unable to describe the effects of an increase in effort on the bonus payment and the optimal critical values. To prove that the firm's costs must be increasing, we directly apply the envelope theorem to (16). For an intuitive explanation, note that the firm's costs can also be written as:

$$C_T^F(a; w^R) = [1 - mF(z_1; a)]w^R + [1 - mF(z_2; a)](B_T - w^R) \tag{21}$$

where $B_T - w^R$ denotes the additional bonus paid on top of the reference wage if $x \geq z_2$. By stochastic dominance, we know that $F_a(x; a) \leq 0$. Accordingly an increase in a raises the respective probabilities of paying the reference wage and the additional bonus. Moreover, simple differentiation verifies that the additional bonus itself is increasing in a.

Finally, we consider the comparative statics for the intermediate contract, applying a similar restriction as above whereby the form of the optimal contract is taken to remain unchanged throughout the exercise.

Proposition 6. *Suppose the optimal contract is an intermediate contract. Then the firm's expected wage costs are i) decreasing in α, ii) decreasing in m, iii) weakly decreasing in w^R and iv) increasing in a.*

Proposition 6 states that similar results and intuitions also hold for the intermediate contract. Wage payments are fixed at zero and w^R. Hence, for changes in α, m and a the critical value z_I^* defined by (12) must adjust in order for the contract to remain incentive compatible. For α and m, the critical value z_I^* increases such that the reference wage is paid less often and costs are reduced. The converse is true for a. When w^R increases, there are two countervailing effects on the firm's costs. While the incentive payment is higher, it must also be paid less often in order for effort incentives to remain constant: z_I^* therefore decreases. In the proof of Proposition 6, we show that the latter effect weakly dominates and an increase in the worker's reference wage cannot increase expected costs, similar to our findings for the strict ternary contract in Proposition 3.

6 Endogenous Reference Wage

The foregoing analysis investigated the contracting problem of a single firm-worker pair who took the reference wage as given. In this section, we allow w^R to be determined endogenously. Specifically, we consider an economy populated by a continuum of identical firm-worker pairs, each of whom contract with one another as described above. Consistent with the idea of keeping up with the Joneses, we then assume that each worker's reference wage is determined by the equilibrium average wage in this economy.[15]

Formally, we denote the economy's average wage by $\overline{w}$ and consider the equilibrium in which $w^R = \overline{w}$. While $\overline{w}$ is endogenous to the economy, due to our assumption of a continuum it cannot be influenced by any single firm-worker pair, each of whom take w^R as given when contracting. Moreover, due to the homogeneity in the economy, all contracts between pairs will be identical, leading to a common wage scheme. As a result of our assumption regarding the determination of the reference wage, this wage scheme must then be strictly ternary with $0 < \overline{w} < B_T$.

We first consider the costs associated with an exogenously given effort level a. Specifically, we derive $C_T^F(a; \overline{w}[a])$, assuming that all firm-worker pairs implement the same effort level, where $\overline{w}[a]$ results endogenously as described above.[16] Using this notation, our homogeneity assumption implies that in equilibrium each firm faces identical costs, which are equal to the average wage:

$$C_T^F(a; \overline{w}[a]) = \overline{w}[a] \tag{22}$$

Recall that the bonus payment and the firm's expected costs for the strict ternary contract are given by (15) and (16) respectively. It can be seen from (17) and (18) that the optimal critical values z_1^* and z_2^* are independent of the reference wage and therefore are unaffected by variations in $\overline{w}[a]$. Combining (16) and (22) together with $\overline{w}[a] = w^R$ and rearranging terms, we obtain:

$$w^R = \frac{[1 - mF(z_2; a)]c'(a)}{-m\left[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)(1 + \alpha)\right]} \tag{23}$$

which also represents the firm's costs of implementing a. Substituting (23) into (15) then yields the bonus payment for an endogenously determined w^R:

$$B_T = \frac{[1 - mF(z_2; a) + mF(z_1; a)]\, c'(a)}{-m\left[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)(1 + \alpha)\right]} \tag{24}$$

Clearly, the comparative static results for z_1 and z_2 are unchanged from the case of an exogenous reference wage. Proposition 7 establishes similar results to Propositions 2, 4 and 5 for the firm's costs of implementing effort and bonus payment, for the case of an endogenous reference wage. The basic intuition for these results is analogous to that for an exogenous reference wage as discussed in

[15]Implicitly, we apply the standard general equilibrium approach to the determination of the economy's average wage and assume that, at the time of contracting, firm-worker pairs have rational expectations regarding this equilibrium.

[16]The square brackets in $\overline{w}[a]$ are used to emphasise that the argument of $\overline{w}[\cdot]$ is not an individual choice of the firm-worker pair, but results from the assumption that all pairs are identical and implement the same effort level.

the previous section and is therefore omitted here.

Proposition 7. *Suppose the reference wage is determined endogenously as the average wage in the economy. Then:*

i) An increase in the KUJ parameter, α, reduces both the bonus payment and the firm's cost of implementing effort.

ii) An increase in the quality of monitoring, m, reduces the firm's cost of implementing effort, while the change in the bonus payment is ambiguous.

iii) An increase in the worker's effort, a, raises the firm's cost of implementing effort.

Throughout this paper, we have so far considered the firm's optimal contractual choice when implementing a given effort level a. In order to endogenise effort, we further extend the model by introducing an increasing concave function $v(a)$, which represents the value accruing to each firm associated with its worker's effort. The optimal a then results endogenously as the effort level which maximises each firm's profit. Given our restrictions on $v(a)$ and $c(a)$, this optimal value is implicitly defined by equalising a firm's marginal benefit to their marginal cost of implementing effort, $\frac{\partial C_T^F}{\partial a}(a; \overline{w}\,[a])$, for the equilibrium value of $\overline{w}\,[a]$. This, along with (17) and (18), then defines an equation system which can be solved for a^*, z_1^* and z_2^*.

Keeping in mind that each firm decides on their effort choice taking the reference wage as given, they will each equalise the marginal benefit of effort with the partial derivative of the cost function, rather than the total derivative with respect to effort as given by (25).

$$\frac{dC_T^F}{da}(a; \overline{w}\,[a]) = \frac{\partial C_T^F}{\partial a}(a; \overline{w}\,[a]) + \frac{\partial C_T^F}{\partial w^R}(a; \overline{w}\,[a]) \cdot \overline{w}'\,[a] \tag{25}$$

Moreover, we know from Propositions 3 and 7 that the second term on the right hand side of (25) is negative. Accordingly, all firms could increase their profits if they were to collectively decide on how much effort to implement, thereby taking into account the impact of their decision on the economy's average wage.

Figure 3 shows this phenomenon for the foregoing example, $F(x; a) = x + \frac{x - x^2}{a+1}$ and $c(a) = \frac{1}{2}a^2$, assuming $v(a) = 10a$, $m = 0.8$ and $\alpha = 1$. In equilibrium, given the reference wage $\overline{w}[a^*]$, each firm optimally chooses the effort level a^* which maximises profits, here given by $v(a^*) - \overline{w}[a^*]$. However, this is less than the profit level $v(a^{**}) - \overline{w}[a^{**}]$, which could be attained by all firms if they were to agree to implement the higher level of effort a^{**}. Intuitively, the increase from a^* to a^{**} leads to a rise in the average wage in the economy ($\overline{w}'\,[a] > 0$), from

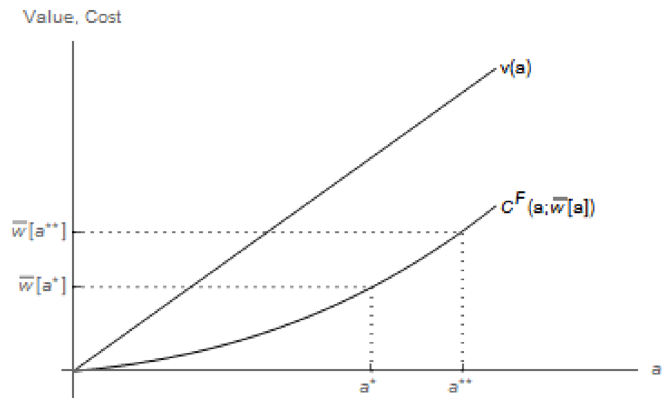

Figure 3: The value of effort and the firm's associated costs.

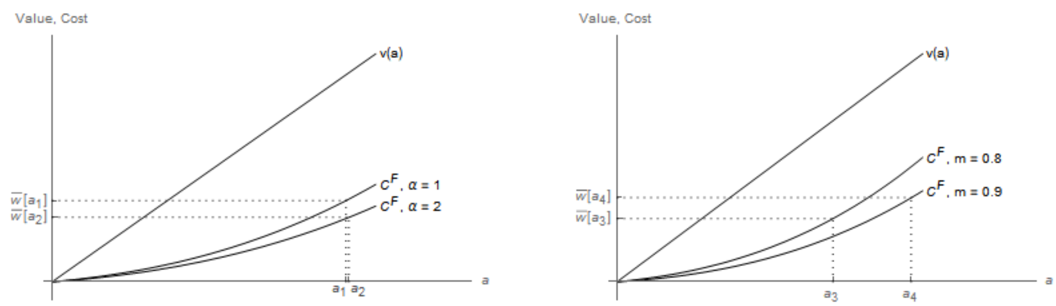

Figure 4: The impact of increases in α and m on the economy's equilibrium.

$\overline{w}[a^*]$ to $\overline{w}[a^{**}]$. Since an increase in the worker's reference wage reduces the costs associated with implementing a given effort level ($\frac{\partial C_T^F}{\partial w^R}(a; \overline{w}\,[a]) < 0$), this increase yields higher profits for each firm. However, given a reference wage of $\overline{w}[a^{**}]$, a^{**} is not the profit maximising effort choice for an individual firm, so that this cannot be sustained as an equilibrium in the economy. Numerically, in this example, under collective decision making by all firms there is a 30% increase in the level of effort implemented and a 7% increase in profits, with the average wage of workers rising by 69%.

We next consider the impact of variations in the parameters α and m for the above numerical example. The left-hand panel of Figure 4 shows how a firm's cost function $C^F(a; \overline{w}\,[a])$ changes following an increase in α from 1 to 2. A higher KUJ parameter has two effects on the firm's equilibrium costs. First, there is a direct effect which reduces the costs of implementing each effort level, in line with Proposition 7. Second, there is an additional indirect effect, since an increase in α also reduces the marginal cost of a; accordingly, the optimal effort level increases from a_1 to a_2 which leads to a countervailing effect on the firm's costs. The right-hand panel of Figure 4 then shows similar effects following an increase in monitoring quality from $m = 0.8$ to $m = 0.9$, leading to an increase in the optimal effort from a_3 to a_4.

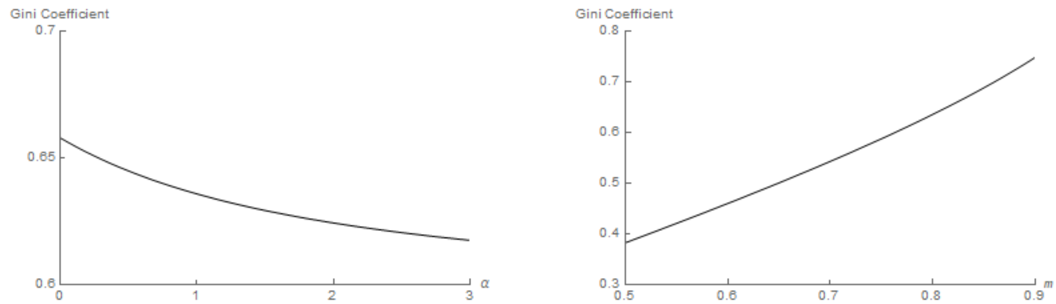

Figure 5: The impact of variations in α and m on the economy's Gini Coefficient.

While it is clear that the firm is better off following an increase in either the KUJ parameter or the quality of monitoring, since the ensuing profits will be higher, the impact on the worker's average income is ambiguous. For instance, in the numerical case represented by Figure 4, the direct effect dominates following an increase in α from 1 to 2 so that the average wage payment decreases from $\overline{w}[a_1]$ to $\overline{w}[a_2]$. In contrast, when we increase m from 0.8 to 0.9, the indirect effect becomes dominant so that the average wage rises from $\overline{w}[a_3]$ to $\overline{w}[a_4]$.

Finally, we evaluate the response of the Gini Coefficient amongst workers in the economy to variations in α and m. Specifically, for variations in α, we fix m at 0.8 and solve the foregoing example for values of $\alpha \in [0, 3]$. For each α, we then obtain the optimal contract with wage schedule $(0, \overline{w}, B_T)$, effort a and associated payment frequencies $mF(z_1; a)$, $m\,[F(z_2; a) - F(z_1; a)]$ and $[1 - mF(z_2; a)]$. We use these variables to calculate the resulting Gini Coefficient. A similar approach is followed for variations in $m \in [0.5, 0.9]$, keeping α fixed at 1. Figure 5 plots the respective Gini Coefficients. The left-hand panel shows a reduction in inequality following increases in the KUJ parameter. Intuitively, this obtains because a higher α allows for lower powered incentives, leading to a compressed wage schedule. In contrast, the right-hand panel shows that inequality increases as the quality of monitoring rises. Intuitively, a higher m leads to higher powered incentive pay in order to implement a given effort level. Moreover, the bonus is further increased due to the higher effort now chosen by firms. Since both effects are associated with a rise in the average wage $\overline{w}$, the intermediate payment is also higher, so that the wage schedule becomes more dispersed.

While the above results seem intuitive, they do not necessarily extend to more general settings. To see why this is the case, observe that an analytical approach requires deriving the impact of a change in a on the payment scheme. However, as discussed in Section 4 this is not possible without further restrictions on the distribution function.

7 Conclusion

In this chapter, we study optimal incentive contracting under moral hazard when workers have relative income concerns and attempt to keep up with the Joneses. The first part of the chapter considers the contracting problem of a single firm-worker pair who take the reference wage as exogenously given and shows that the optimal wage scheme takes either a binary or ternary form. We also show that firms benefit from relative income concerns, since any given effort level can be implemented at a lower cost. The second part of the chapter assumes that the reference wage is endogenously determined by the economy's equilibrium average income. In this case, contracting between pairs becomes interdependent via the reference wage, so that externality effects can arise. It then follows that firms could benefit from collective decision making when deciding on how much effort to implement. Moreover, using a series of numerical examples, we show that an economy with higher relative income concerns has a lower average wage and reduced inequality, as measured by the Gini Coefficient.

There are a number of ways in which our findings complement the existing literature. First, the result that the optimal wage scheme features either two or three distinct payments is consistent with the literature which considers incentive contracting in the presence of loss aversion. A key result in this literature is that the wage scheme becomes partially unresponsive at the reference point, since this shields the worker from losses. Similarly, in our framework, the optimal contract features a payment which is equal to the reference wage. Second, we find that relative income concerns induce a positive incentive effect, similar to both inequity aversion and loss aversion, which allows the firm to implement effort at a lower cost when the worker is financially constrained. However, we argue that in the alternative case — where the worker faces unlimited liability — our findings would differ from these literatures since the participation effect would be positive. This is due to the fact that, in our framework, workers continue to make social comparisons even when choosing to reject the contract. Third, in the latter part of the chapter, we show that our formalisation of other-regarding preferences continues to induce a positive effect on incentives when the reference wage is determined endogenously by the average income in the economy.

To conclude, some comments with respect to our specification of the workers' preferences are in order. First, imposing a piecewise linear wage-utility function significantly simplifies our analysis, since the optimal contractual form features either two or three wages. As is well known from the literature, a considerably more complex wage scheme would emerge were we to introduce strict concavity everywhere. We introduce this simplification since it allows us to investigate some fundamental effects of relative income concerns, while retaining tractability.

Second, we assume that all individuals share a common reference wage which is equal to the average worker's income in the economy. We believe that this is consistent with both the idea of keeping up with the Joneses and with existing scientific evidence, as discussed in the introduction. Moreover, the assumption seems especially appropriate since all workers in our model are identical.

We believe that there are a number of interesting ways in which our analysis could be further developed. A natural extension would be to allow for heterogeneity between firm-worker pairs with respect to productivity, leading to differences in contracting between pairs. One could then examine how changes in the composition of the overall economy influence optimal contracts. Moreover, in such an environment, heterogeneous workers could have relative income concerns with respect to disparate reference wages. For instance, individuals may attempt to keep up with the Joneses within a subset of the economy, such as those with a comparable ability, education etc.

Alternatively, recall that the optimal contract typically features three distinct payments; one possible interpretation of this result is a wage scheme which dismisses the worker for low performance. The remaining payment structure could then be thought of as offering a basic payment equal to the reference wage, together with the promise of a bonus payment for high performance.[17] One could then analyse how incentives can be created via both bonus pay and the threat of dismissal when workers have relative income concerns, as well as the implications for labour market policies. A detailed investigation of this issue is provided in the following chapter of this thesis.

8 Appendix

8.1 Optimal Contractual Form

Proof of Proposition 1. For $0 < w_\emptyset < w^R$, (FFC) implies $w(x) < w^R$ for all x. Hence, $u(w;\alpha,w^R)$ is linear over the relevant region. We leave it to the reader to verify that the contract will take the *strict binary* form described in the main text.[18]

Next, for $w_\emptyset > w^R$ we show that the optimal contract takes the *strict ternary* form. We proceed as follows; (i) we maximise the negative of the objective function; (ii) we initially ignore (PC) verifying at the end of the proof that it is satisfied; (iii) given the restrictions on $F(x;a)$ and $c(a)$, we apply the first-order

[17] Dittmann et al. (2010) and Iantchev (2009) also derive optimal payment schemes which are constant at the smallest possible wage for the lowest levels of performance and offer a similar interpretation.

[18] This is well known from the risk-neutral case; see e.g. Kim (1995; 1997). For our specific model, see also Bental et al. (2016).

approach. Altogether, the Lagrangian of the simplified problem becomes:

$$
\begin{aligned}
\mathcal{L} = & -(1-m)w_{\emptyset} - m\int_0^1 w(x)f(x;a)dx \\
& + \lambda\left(m\int_0^1 \left[w(x)+\theta\alpha(w(x)-w^R)\right] f_a(x;a)dx - c'(a)\right) \\
& + \int_0^1 \xi(x)w(x)dx + \int_0^1 \zeta(x)\left(w_{\emptyset} - w(x)\right)dx
\end{aligned}
\tag{26}
$$

where $u(w;\alpha,w^R) = w + \theta\alpha(w - w^R)$ and θ is an indicator function taking the value 1 if $w < w^R$ and 0 otherwise. In order to deal with the slackness in the first-order condition at w^R, we introduce a multiplier $\gamma(x)$ that takes the value 1 when $w(x) < w^R$, the value 0 when $w(x) > w^R$ and requires $0 \le \gamma(x) \le 1$ at $w(x) = w^R$. The necessary conditions for maximisation of (26) are:

$$
\begin{cases}
-(1-m) + \int_0^1 \zeta(x)dx & = 0 \\
-mf(x;a) + \lambda m\left[1+\gamma(x)\alpha\right] f_a(x;a) + \xi(x) - \zeta(x) & = 0, \ \forall x \\
\xi(x)w(x) & = 0, \ \forall x \\
\zeta(x)(w_{\emptyset} - w(x)) & = 0, \ \forall x \\
\xi(x), \zeta(x) & \ge 0, \ \forall x
\end{cases}
\tag{27}
$$

For any $x \in [0,1]$, there are in principle four possible cases: $\xi(x) = \zeta(x) = 0$; $\xi(x) = 0$, $\zeta(x) > 0$; $\xi(x) > 0$, $\zeta(x) = 0$ and $\xi(x)$, $\zeta(x) > 0$. Clearly, with $a > 0$ the last possibility can never occur (otherwise by complementary slackness $w_{\emptyset} = 0$, and by (FFC) and (WFC) $w(x) = 0$ for all x so that (IC) can never be satisfied). Associated with the remaining cases, we define three sets:

- $X^I = \{x \in [0,1] \,|\, \xi(x) > 0, \zeta(x) = 0\}$. Accordingly, $\forall x \in X^I$, we have $w(x) = 0 < w^R$ and thus $\theta = 1$. Hence, the second condition in (27) implies:

$$
\forall x \in X^I, \ \lambda\left[1+\alpha\right]\frac{f_a(x;a)}{f(x;a)} < 1 \tag{28}
$$

– $X^{II} = \{x \in [0,1] \,|\, \xi(x) = 0, \zeta(x) = 0\}$. Hence, we have:

$$\forall x \in X^{II},\ \lambda\,[1 + \gamma(x)\alpha]\,\frac{f_a(x;a)}{f(x;a)} = 1 \tag{29}$$

Observe that by strict MLRP, (29) can only be satisfied for one distinct x for $\gamma(x) = 0$ and similarly for $\gamma(x) = 1$, so that $w(x) = w^R$ *a.e.* in X^{II}.

– $X^{III} = \{x \in [0,1] \,|\, \xi(x) = 0, \zeta(x) > 0\}$. For any $x \in X^{III}$, we have $w(x) = w_\emptyset > w^R$ and therefore $\theta = 0$, hence:

$$\forall x \in X^{III},\ \lambda\frac{f_a(x;a)}{f(x;a)} > 1 \tag{30}$$

Observe that for any λ, the sets X^I, X^{II}, X^{III} form a partition of $[0,1]$. Clearly, $\lambda = 0$ is not possible since otherwise $X^I = [0,1]$ from (28) which violates (IC). Suppose $\lambda > 0$. We define z_1 and z_2 as the solutions to the respective equations:

$$\lambda(1+\alpha)\frac{f_a(x;a)}{f(x;a)} = 1 \tag{31}$$

and

$$\lambda\frac{f_a(x;a)}{f(x;a)} = 1 \tag{32}$$

Claim 1. $X^I = [0, z_1)$; $X^{II} = [z_1, z_2]$ and $X^{III} = (z_2, 1]$ where $z_1, z_2 \in (0,1)$.

Proof. We first show that $z_1, z_2 \in (0,1)$. By strict MLRP, if (31) and (32) have solutions, they are unique with $0 < z_1 < z_2$. Contrary to the claim, suppose $z_2 = 1$ or does not exist. Accordingly, the set $X^{III} = \emptyset$. Ignoring z_1 and z_2 since they are of measure zero, this implies that either $w(x) = 0$ or $w(x) = w^R$ for all x. But then the firm can reduce $w_\emptyset$, leading to a contradiction. Accordingly, $z_2 < 1$ which also ensures $z_1 < 1$.

$[0, z_1) \subseteq X^I$ follows directly from Figure 1 since for all $x < z_1$ we have $\lambda\,[1 + \gamma(x)\alpha]\,\frac{f_a(x;a)}{f(x;a)} < 1$. To verify $[0, z_1) \supseteq X^I$, suppose to the contrary $x \in X^I$, but $x \geq z_1$. Since $x \in X^I$, we know $w(x) = 0$. Then by strict MLRP $\lambda\,[1+\alpha]\,\frac{f_a(x;a)}{f(x;a)} \geq 1$ contradicting (28). Hence, $X^I = [0, z_1)$. A similar argument can be made to show that $X^{III} = (z_2, 1]$, from which $X^{II} = [z_1, z_2]$ follows. □

Since the wages paid at the exact points z_1 and z_2 are irrelevant to the optimisation problem, the above discussion implies that without loss of generality the optimal wage scheme takes the strict ternary form given by (4) in the proposition.

Using similar arguments with $\lambda < 0$ would imply a monotonically decreasing wage leading to a contradiction with (IC). Next, we verify that (PC) holds. Given the contractual form, (PC) requires:

$$
\begin{aligned}
&- mF(z_1;a)\alpha w^R + m[F(z_2;a) - F(z_1;a)]w^R \\
&\qquad\qquad + [1 - mF(z_2;a)]B_T - c(a) \geq -\alpha w^R \quad (33)
\end{aligned}
$$

Clearly (33) is satisfied at $a = 0$, so that the worker can always guarantee expected utility weakly greater than $-\alpha w^R$ by undertaking zero effort. Moreover, by (IC), expected utility from exerting the firm's desired effort level must be weakly greater than this; accordingly, the contract satisfies (PC).

Finally, with $w_\emptyset = w^R$ the proof follows along similar lines to the above, except that either z_2 does not exist or is just equal to 1. Hence, any realization x is either in X^I or X^{II} and the solution is the *intermediate contract* described by Proposition 1. □

8.2 Optimal Contracts

Proof of Lemma 1. We define $S_B = \{a > 0 \mid C_B^F(a;w^R) < +\infty\}$,
$S_T = \{a > 0 \mid C_T^F(a;w^R) < +\infty\}$ and proceed in four steps. Point 1 establishes a subsidiary result which is used during the remainder of the proof. Parts i), ii) and iii) of the Lemma are then shown in points 2, 3 and 4 respectively.

1. $S_B \cap S_T = \emptyset$. Consider any $a \in S_B$. Accordingly, (9) has a solution for a with $B_B < w^R$. By (8), this implies that $c'(a) + mF_a(z_B^*;a)(1+\alpha)w^R < 0$. By contradiction, suppose $a \in S_T$. Therefore (16) has a solution for a with $B_T > w^R$. By (15), this requires $c'(a) + m(1+\alpha)F_a(z_1^*;a)w^R > 0$. Taken together, these inequalities imply:

$$c'(a) + mF_a(z_B^*;a)(1+\alpha)w^R < c'(a) + m(1+\alpha)F_a(z_1^*;a)w^R$$

$$\iff \quad F_a(z_B^*;a) < F_a(z_1^*;a)$$

$$\iff \quad F_a(z_2^*;a) < F_a(z_1^*;a)$$

   where the third inequality follows by (10) and (18). Moreover, (17) and (18) imply $f_a(z_1^*;a)$, $f_a(z_2^*;a) > 0$; combined with strict MLRP, this implies that for all $x \in [z_1^*, z_2^*]$, we have $f_a(x;a) > 0$. Hence, for $x \in [z_1^*, z_2^*]$, $F_a(x;a)$ is increasing in x so that $z_2^* < z_1^*$. This yields a contradiction to (19) by MLRP and $\alpha > 0$. Hence, $S_B \cap S_T = \emptyset$.

2. $\mathcal{A}_\mathcal{B} = S_B$. Clearly, $\mathcal{A}_\mathcal{B} \subseteq S_B$ since for any $a > 0$, we have $C^F(a;w^R) < +\infty$. To verify $S_B \subseteq \mathcal{A}_\mathcal{B}$, consider $a \in S_B$. By point 1 of the proof, we know

$C_B^F\left(a;w^R\right) < C_T^F\left(a;w^R\right)$. Moreover, $a \in S_B$ implies $C_B^F\left(a;w^R\right) < +\infty$ so that the constraint in (9) is irrelevant and the optimisation problem simplifies to:

$$\min_{z_B} \frac{1 - mF(z_B;a)}{-mF_a(z_B;a)(1+\alpha)} c'(a) \tag{34}$$

Observe that (34) contains the intermediate contract as a special case (when $z_B = z_I$). Hence by convexity of the problem $C_B^F\left(a;w^R\right) < C_I^F\left(a;w^R\right)$ and therefore $C_B^F\left(a;w^R\right) = \min\left\{C_B^F\left(a;w^R\right), C_I^F\left(a;w^R\right), C_T^F\left(a;w^R\right)\right\}$ and $a \in \mathcal{A}_\mathcal{B}$.

3. $\mathcal{A}_\mathcal{T} = S_T$. This follows by a similar argument to point 2.

4. $\mathcal{A}_\mathcal{B}, \mathcal{A}_\mathcal{I}, \mathcal{A}_\mathcal{T}$ *partition* $\mathbb{R}_{++}$. By Proposition 1, $\mathcal{A}_\mathcal{B} \cup \mathcal{A}_\mathcal{I} \cup \mathcal{A}_\mathcal{T} = \mathbb{R}_{++}$. $\mathcal{A}_\mathcal{B} \cap \mathcal{A}_\mathcal{T} = \emptyset$ follows immediately from points 1-3. $\mathcal{A}_\mathcal{B} \cap \mathcal{A}_\mathcal{I} = \emptyset$ can be seen from the arguments of point 2. $\mathcal{A}_\mathcal{I} \cap \mathcal{A}_\mathcal{T} = \emptyset$ then follows symmetrically.

□

Proof of Lemma 2. We proceed in three steps.

1. z_B^* *is weakly increasing in* a. Let $\varphi(z_B, a, m)$ represent the LHS of (10) so that $\frac{\partial z_B^*}{\partial a} = -\frac{\varphi_a(z_B,a,m)}{\varphi_z(z_B,a,m)}$ where taking the derivative with respect to z_B and substituting for $[1 - mF(z_B;a)]$ using (10) yields:

$$\frac{\partial \varphi}{\partial z_B} = \frac{mF_a(z_B;a)}{f_a(z_B;a)} \left[f_a(z_B;a) f_x(z_B;a) - f(z_B;a) f_{ax}(z_B;a)\right] > 0 \tag{35}$$

To see the sign, observe that the fraction and the square bracket are both negative by (10) and since $\frac{\partial}{\partial x}\left(\frac{f_a}{f}\right) = \frac{1}{f^2}\left[f_{ax} f - f_a f_x\right] > 0$ by strict MLRP. Similarly, taking the derivative of φ in a and again substituting for $[1 - mF(z_B;a)]$ yields:

$$\frac{\partial \varphi}{\partial a} = \frac{mf(z_B;a)}{f_a(z_B;a)} \left[f_a(z_B;a) F_{aa}(z_B;a) - F_a(z_B;a) f_{aa}(z_B;a)\right] \leq 0 \tag{36}$$

The result obtains since $\frac{\partial}{\partial a}\left(\frac{F_a}{f_a}\right) = \frac{1}{f_a^2}\left[f_a F_{aa} - F_a f_{aa}\right] \leq 0$ verifying the result.

2. $\lim_{a\to 0} B_B = 0$. First, note $\lim_{a\to 0} z_B^*(a) \neq 0, 1$. Indeed, by (10) we must have $f_a(z_B;a) > 0$. However, by strict MLRP we know $f_a(x;a) < 0$ as $x \to 0$. Moreover, z_B^* being weakly increasing in a implies that the limit cannot converge to 1. Hence, (8) and $\lim_{a\to 0} c'(a) = 0$ directly verify the claim.

3. Taking into account $z_B^{*\prime}(a) \geq 0$ and the restrictions on $F(x;a)$, (8) implies by total differentiation:

$$\frac{dB_B}{da} = \frac{-F_a(z_B^*;a)c''(a) + c'(a)\left[f_a(z_B^*;a)z_B^{*\prime}(a) + F_{aa}(z_B^*;a)\right]}{m(1+\alpha)\left[F_a(z_B^*;a)\right]^2} > 0$$

Altogether, the strict binary contract obtains for effort levels close to zero. As effort increases, the strict binary contract remains feasible (and therefore by part *i*) of Lemma 1 also optimal) until it requires setting $B_B = w^R$; we denote the associated effort level by $\underline{a}$. Finally, it follows by Proposition 1 that for all $a \geq \underline{a}$, we have $a \in \mathcal{A}_{\mathcal{WT}}$, thus concluding the proof.

□

8.3 Comparative Statics

Proof of Proposition 2. By the envelope theorem, we have:

$$\frac{\partial C_T^F}{\partial \alpha} = -[1 - mF(z_2;a)]\frac{F_a(z_1;a)}{F_a(z_2;a)}w^R < 0$$

which implies that the firm's costs decrease. It can be seen from (18) that z_2^* does not vary in α; (19) and strict MLRP imply that z_1^* is decreasing in α. The previous two results verify that B_T decreases in α. □

Proof of Proposition 3.

1. *The firm's costs are reduced.* Let $w^R = \widehat{w}^R$ and denote the optimal contract by $\widehat{\mathcal{C}} = \left(0, \widehat{w}^R, \widehat{B}, \widehat{z}_1, \widehat{z}_2\right)$. Next, let the reference wage increase to $\widetilde{w}^R > \widehat{w}^R$ and consider the (non-optimal) contract $\widetilde{\mathcal{C}} = \left(0, \widehat{w}^R, \widetilde{B}, \widehat{z}_1, \widehat{z}_2\right)$ where the bonus $\widetilde{B}$ has been adjusted to maintain incentive compatibility, while the rest of the contract remains unchanged. Using the first-order condition of the worker's incentive compatibility constraint, it follows that:[19]

$$\widetilde{B} = \widehat{w}^R + \alpha(\widehat{w}^R - \widetilde{w}^R) + \frac{c'(a) + (1+\alpha)m\widehat{w}^R F_a(\widehat{z}_1;a)}{-mF_a(\widehat{z}_2;a)} \tag{37}$$

[19] To be clear, $\widehat{w}^R$ appears in (37) since this is the payment prescribed for intermediate performance. $\widetilde{w}^R$ appears directly from the worker's utility function. In particular, (37) can be derived by differentiating the LHS of (33) and setting the result equal to zero. However, one must first account for the fact that since $\widehat{w}^R < \widetilde{w}^R$, in this case we have $u(\widehat{w}^R;\alpha,\widetilde{w}^R) = \widehat{w}^R + \alpha(\widehat{w}^R - \widetilde{w}^R)$. This explains the additional term in (37) which does not appear in (15).

Next, we use (15) and (37) in order to calculate:

$$\left[\widetilde{B} - \widehat{B}\right] = \alpha \left[\widehat{w}^R - \widetilde{w}^R\right] < 0$$

Accordingly, since all other aspects of the contract are unchanged, the firm's costs are lower. Even though $\widetilde{\mathcal{C}}$ is not the cost minimal incentive scheme, this nevertheless verifies that the optimal contract reduces costs.

2. *z_1 and z_2 unchanged.* Follows immediately from (17) and (18).

3. *Bonus payment decreases.* Follows from points 1 and 2 and (21).

□

Proof of Proposition 4. By the envelope theorem, we have:

$$\frac{\partial C_T^F}{\partial m} = \frac{c'(a) + F(z_2; a)(1+\alpha)F_a(z_1; a)m^2 w^R}{m^2 F_a(z_2; a)} - F(z_1; a)w^R$$

which is negative if $c'(a) + F(z_2; a)(1+\alpha)F_a(z_1; a)m^2 w^R > 0$. This is satisfied since $c'(a) + m(1+\alpha)F_a(z_1; a)w^R > 0$ (else we would have $B_T \leq w^R$) and clearly $mF(z_2; a) < 1$. This implies that the firm's costs decrease. Next, we proceed as in the proof of Lemma 2 and let $\varphi(z_B, a, m)$ represent the LHS of (10) so that $\frac{\partial z_B^*}{\partial m} = -\frac{\varphi_m(z_B,a,m)}{\varphi_z(z_B,a,m)}$. Taking the derivative with respect to m, we obtain:

$$\frac{\partial \varphi}{\partial m} = F_a(z_B)f(z_B) - F(z_B)f_a(z_B) < 0$$

where the sign can be seen from (10). This along with (35) implies $\frac{\partial z_B^*}{\partial m} > 0$. This also implies $\frac{\partial z_2^*}{\partial m} > 0$ since $z_B^* = z_2^*$ by (10) and (18). Moreover, $\frac{\partial z_1^*}{\partial m} > 0$ follows from (19) and strict MLRP. Finally, the change in the bonus payment remains ambiguous since $\frac{\partial B_T}{\partial z_1}, \frac{\partial B_T}{\partial z_2} > 0$ while $\frac{\partial B_T}{\partial m} < 0$. □

Proof of Proposition 5. By the envelope theorem, we have:

$$\frac{\partial C_T^F}{\partial a} = -\, mF_a(z_1; a)w^R - mF_a(z_2; a)(B_T - w^R) + [1 - mF(z_2; a)]\frac{\partial B_T}{\partial a}$$

where $\frac{\partial B_T}{\partial a} > 0$ from (15). Hence the firm's costs are increasing in a. □

Proof of Proposition 6. For points *i),ii)* and *iv*), we proceed in two steps. First, we apply the implicit function theorem to (12), which yields:

$$\frac{\partial z_I^*}{\partial \alpha} = -\frac{F_a(z_I; a)}{f_a(z_I; a)\,(1+\alpha)} > 0$$

$$\frac{\partial z_I^*}{\partial m} = -\frac{F_a(z_I; a)}{m f_a(z_I; a)} > 0$$

$$\frac{\partial z_I^*}{\partial a} = -\frac{c''(a) + m F_{aa}(z_I; a)\,(1+\alpha)\,w^R}{m f_a(z_I; a)\,(1+\alpha)\,w^R} < 0$$

where the sign of each derivative follows from $f_a(z_I^*; a) > 0$ as shown in the proof of Proposition 1.[20] Second, we totally differentiate (13) with respect to each variable:

$$\frac{dC_I^F}{d\alpha} = \left[-m f\left(z_I^*; a\right) w^R\right] \frac{\partial z_I^*}{\partial \alpha} < 0$$

$$\frac{dC_I^F}{dm} = -\left[F\left(z_I^*; a\right) + m f\left(z_I^*; a\right) \frac{\partial z_I^*}{\partial m}\right] w^R < 0$$

$$\frac{dC_I^F}{da} = -\left[F_a\left(z_I^*; a\right) + f\left(z_I^*; a\right) \frac{\partial z_I^*}{\partial a}\right] m w^R > 0$$

which verifies the claims.

For point *iii)*, suppose that initially the reference wage is given by w^R with an optimal contract defined by $\mathcal{C} = (0, w^R, z_I)$. Given this, the effort level induced is implicitly defined by (12). Consider a small increase of ϵ in the reference wage, such that the optimal contract remains intermediate. Next, suppose that instead of applying the optimal (intermediate) contract, the firm were to leave the incentive scheme unchanged. Since the bonus payment (w^R) is then strictly less than the reference wage ($w^R + \epsilon$), the worker's incentive compatibility constraint must be taken from (7) i.e.:

$$a = \arg\max_{\widehat{a}} \; -mF(z_I; \widehat{a})\alpha\left(w^R + \epsilon\right) + \left[1 - mF(z_I; \widehat{a})\right]\left[(1+\alpha)\,w^R - \alpha\left(w^R + \epsilon\right)\right] - c(\widehat{a})$$

Differentiating this and rearranging yields (12), implying that the unchanged contract would implement the original effort level at identical costs. Clearly, adjus-

[20] Recall that in the intermediate case, the first critical value continues to be defined by equation (31). $\lambda > 0$ implies that the x which solves this equation must then satisfy $f_a(x; a) > 0$.

ting the contract optimally must therefore generate weakly lower expected costs, verifying the claim. □

Proof of Proposition 7.

i) The change in the bonus payment follows from differentiating (24) with respect to α and z_1, which yields:

$$\frac{\partial B_T}{\partial \alpha} = \frac{[1 - mF(z_2; a) + mF(z_1; a)]\, c'(a)[1 - mF(z_2; a)]F_a(z_1; a)}{m\,[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)\,(1 + \alpha)]^2} < 0$$

and

$$\frac{\partial B_T}{\partial z_1} = \frac{c'(a)\,[1 - mF(z_2; a)]\,[F_a(z_2; a) - F_a(z_1; a)\,(1 + \alpha)]\, mf(z_1; a)}{m\,[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)\,(1 + \alpha)]^2}$$

$$+ \frac{c'(a)\,[1 - mF(z_2; a)]\,[1 - mF(z_2; a) + mF(z_1; a)]\, f_a(z_1; a)\,(1 + \alpha)}{m\,[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)\,(1 + \alpha)]^2} > 0$$

respectively. Along with our previous findings that z_1 is decreasing in α and the value of z_2 is independent of α, these results imply that (24) is decreasing in α. To see that $F_a(z_2; a) - F_a(z_1; a)\,(1 + \alpha) > 0$, note that for $x \in [z_1^*, z_2^*]$, $F_a(x; a)$ is increasing in x, as shown in the proof of Lemma 1. Since $F_a < 0$, we have $|F_a(z_2; a)| < |F_a(z_1; a)|$ from which $F_a(z_2; a) - F_a(z_1; a) > 0$ follows. For the change in the firm's costs, we first establish the following from differentiation of (23):

$$\frac{\partial C_T^F}{\partial z_1} = \frac{[1 - mF(z_2; a)]c'(a)\,[mF_a(z_2; a)f(z_1; a) + [1 - mF(z_2; a)]f_a(z_1; a)\,(1 + \alpha)]}{m\,[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)\,(1 + \alpha)]^2} = 0$$

$$\frac{\partial C_T^F}{\partial z_2} = \frac{c'(a)F(z_1; a)\,[mF_a(z_2; a)f(z_2; a) + [1 - mF(z_2; a)]f_a(z_2; a)]}{[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)\,(1 + \alpha)]^2} = 0$$

where equality to zero follows from (17) and (18) so that small changes in z_1 and z_2 do not influence (23). Taking the partial derivative of (23) with respect to α then yields:

$$\frac{\partial C_T^F}{\partial \alpha} = \frac{[1 - mF(z_2; a)]^2 c'(a)F_a(z_1; a)}{m\,[mF_a(z_2; a)F(z_1; a) + [1 - mF(z_2; a)]F_a(z_1; a)\,(1 + \alpha)]^2} < 0$$

as required.

ii) Since $\frac{\partial C_T^F}{\partial z_1} = \frac{\partial C_T^F}{\partial z_2} = 0$, the decrease in the firm's costs follows from partial differentiation of (23):

$$\frac{\partial C_T^F}{\partial m} = c'(a)\frac{mF(z_1;a)F_a(z_2;a)\left[2-mF(z_2;a)\right]+(1+\alpha)\left[1-mF(z_2;a)\right]^2F_a(z_1;a)}{m^2\left[mF_a(z_2;a)F(z_1;a)+[1-mF(z_2;a)]F_a(z_1;a)\left(1+\alpha\right)\right]^2} < 0$$

as required.

iii) Since $\frac{\partial C_T^F}{\partial z_1} = \frac{\partial C_T^F}{\partial z_2} = 0$, the increase in the firm's costs follows from partial differentiation of (23):

$$\frac{\partial C_T^F}{\partial a} = \frac{c'(a)mF(z_1;a)\left[mF_a(z_2;a)^2+[1-mF(z_2;a)]F_{aa}(z_2;a)\right]}{m\left[mF_a(z_2;a)F(z_1;a)+[1-mF(z_2;a)]F_a(z_1;a)\left(1+\alpha\right)\right]^2}$$

$$+\frac{c'(a)[1-mF(z_2;a)]\left[mF_a(z_2;a)F_a(z_1;a)+[1-mF(z_2;a)]\left(1+\alpha\right)F_{aa}(z_1;a)\right]}{m\left[mF_a(z_2;a)F(z_1;a)+[1-mF(z_2;a)]F_a(z_1;a)\left(1+\alpha\right)\right]^2}$$

$$-\frac{c''(a)[1-mF(z_2;a)]\left[mF_a(z_2;a)F(z_1;a)+[1-mF(z_2;a)]F_a(z_1;a)\left(1+\alpha\right)\right]}{m\left[mF_a(z_2;a)F(z_1;a)+[1-mF(z_2;a)]F_a(z_1;a)\left(1+\alpha\right)\right]^2} > 0$$

as required.

□

Chapter 3

1 Introduction

The foregoing chapter studied incentive contracting in an environment where workers' preferences were characterised by relative income concerns, with a key finding being that the optimal wage scheme featured three distinct payments. As discussed previously, since the payment associated with poor performance was set by the firm to be as low as possible, one possible interpretation of this result is a binary wage scheme with the additional option of dismissing the worker. Although there are then only two distinct payments, the contract would be perceived by the worker as ternary, since there are three distinct outcomes (i.e. dismissal and a base wage; retention and a base wage; retention, a base wage and a bonus payment) associated with poor, intermediate and good performance, respectively.

Clearly, dismissal can only be an effective method of creating incentives if a worker experiences a reduction in expected utility following termination of the employment relationship. Whether this is the case will depend on the outcomes associated with dismissal and, in particular, labour market conditions. For instance, a worker who expects to be unemployed for a long period of time following dismissal, resulting in a significant loss of income, would anticipate a large reduction in expected utility and, ex ante, become more inclined to undertake costly effort in order to avoid this possibility.

It follows that labour market policies, which influence the payoffs of both employed and unemployed workers, will have further implications for employment relationships when firms must provide workers with effort incentives. In recent years, there has been a renewed focus on policies which attempt to reduce inequalities; these include minimum wages, unemployment benefits and regulation aimed at realigning the bargaining power between firms and workers. Such policies are particularly relevant if individuals care about how their incomes compare to those of others. However, given the foregoing discussion, it is important when evaluating these policies to also consider how they might impact employment relationships and the ability of firms to provide effort incentives.

In this chapter, we explore this issue by considering an economy populated by infinitely lived firms and workers to be matched with one another, all of whom are homogeneous and atomistic in nature. Firms are assumed to be self-interested and risk neutral, while workers' preferences exhibit relative income concerns. As in the previous chapter, we formalise this notion by assuming that each worker's utility function is piecewise linear around a reference wage, which, in equilibrium, is determined by the average income of employed workers in the economy.[1] The measure of workers is assumed to be constant, whereas the measure of firms is determined in equilibrium by freedom of market entry.

All parties operate in a frictional labour market, so that both firms and workers who are unmatched face a positive probability of remaining unmatched in the subsequent period. This is formalised using the standard textbook model of 'labour market matching' (see e.g. Pissarides, 2000, Chapter 1) in discrete time, albeit with two key differences. First, each firm's revenue depends on an unobservable effort input by its corresponding worker, resulting in moral hazard. In accordance with the foregoing discussion, each firm creates incentives through use of both a single explicit bonus payment and the threat of dismissal for poor performance, so that there are three distinct outcomes for a worker depending on the realisation of a performance measure. Such a contract is then perceived by workers as being ternary in nature, as previously discussed. Second, the possibility that a worker is dismissed due to poor performance results in an endogenous proportion of matches being dissolved; this creates an additional flow into unemployment.

In this environment, we solve for a firm's choice of contract in order to implement a given effort level. For simplicity, we assume that each firm-worker pair who are matched in a given time period negotiate a new single-period contract, regardless of whether they were matched in previous time periods, so that multi-period contracts between pairs are not allowed for. We show that the equilibrium critical value of the performance measure associated with dismissal depends on the ratio between the worker's relative benefits of being employed and the firm's relative benefits of employing a worker. Intuitively, from a firm's point of view, the threat of dismissal creates incentives for the worker, allowing for a lower explicit bonus payment and therefore reducing expected wage costs. The size of these *dismissal incentives* is determined by the worker's relative benefit of being employed. On the other hand, the 'costs' for a firm of using dismissal are the possibility of reduced profits in future periods, should they terminate a worker's employment following poor performance. The firm trades off these costs and benefits when

[1] This assumption, that all workers — including those who are unemployed — compare their income to the average wage payment of employed workers in a given time period, is further discussed in the conclusion to this chapter.

deciding on the optimal critical value.[2]

This trade-off will, in particular, depend on two exogenous parameters in our model which relate to labour market policy. First, the minimum wage determines how low firms can set their base wages and therefore has an impact on both the income of employed workers and the profits of firms. Second, unemployment benefits determine the income of unmatched workers and therefore affect the relative benefits of employment. Using a series of numerical examples, we explore the implications of exogenous variations in these parameters for the probability of dismissal, the optimal contract and the steady-state equilibrium labour market outcomes.

In our framework, the minimum wage acts as an *efficiency wage*, which improves the relative benefits for a worker of being employed. Accordingly, we find that dismissal is more effective following a raise in the minimum wage and is therefore used more often. This, in turn, allows for a lower bonus payment. It follows that while firms are worse off from such an increase since they must pay a higher base wage, they are partially cushioned from the impact due to this positive incentive effect, with the overall result being a more compressed wage schedule. Additionally, we find that, in our example, the unemployment rate rises following an increase in the minimum wage due to two distinct effects. First, the increased rate of dismissal means that more firm-worker matches are terminated in each period, with the frictional nature of the labour market implying that some of this unemployment is persistent. Second, the reduction in profits leads to less active firms in the market in each period, lowering the rate at which pairs are matched.

The effect of an increase in unemployment benefits, in contrast, is to reduce the relative value for a worker of being employed. This makes the threat of dismissal less effective, so that the bonus payment in this case must be increased. Moreover, since dismissal becomes less effective, it is used less often. Overall, the firm's costs of implementing effort increase, while the wage schedule becomes more dispersed. We find that a rise in unemployment benefits induces two opposing effects on the unemployment rate. Since wage costs increase, firms make lower profits and as a result some are forced out of the market, meaning that it becomes harder for unemployed workers to find match. However, as dismissal is used less often to create incentives, existing matches last for longer on average. In our example, the latter effect is dominant so that unemployment is reduced in equilibrium.

We show that while these effects exist for the case where workers are self-interested, they are amplified in the presence of relative income concerns. Intui-

[2] In the framework of Wang and Yang (2015), workers who produce a sequence of low outputs become increasingly poor and therefore difficult to motivate. Similar to our analysis, firms then trade off the benefits of hiring a new worker with the potential costs in terms of foregone future profits, with dismissal only occurring when workers become sufficiently poor.

tively, this results because both the minimum wage and unemployment benefit payments are less than the average wage in the economy, which acts as a reference wage for workers. It follows that even small changes then have large impacts on utility when individuals care about their relative income, due to the high marginal utility below the reference wage.

This chapter is closely related to three distinct areas. First, our analysis is part of a wider literature which studies employment relationships characterised by moral hazard in a frictional labour market (see for instance Rocheteau, 2001; Demougin and Helm, 2011; Moen and Rosén, 2011, 2013; Starmans, 2017). Perhaps most closely related to this chapter is the recent article by Wang and Yang (2015), who study dynamic contracts featuring endogenous termination, motivated by the optimal provision of incentives and risk sharing.[3] They argue that their model generates predictions consistent with empirical observations relating to wage and employment dynamics, with severance compensation playing a key role in their framework. In contrast to their paper, our primary focus is the equilibrium contract between a firm-worker pair rather than labour market outcomes. Moreover, their analysis differs from ours in several important respects: they allow for multi-period contracts, study non-participation in the labour force as well as consumption and savings decisions and pairs are assumed to bargain over the employment contract. In particular, a key difference is that in their setup workers are assumed to be risk averse but self-interested, so that relative income concerns play no role.

Second, our analysis is related to a number of studies which consider the labour market impact of minimum wages, typically in the context of search-and-matching models (for instance, Van den Berg and Ridder, 1998; Acemoglu, 2001 and Flinn, 2006). Moreover, our results are consistent with previous findings that minimum wages can act as efficiency wages in environments where the effort of workers is unobservable (Shapiro and Stiglitz, 1984; Rebitzer and Taylor, 1995). The chapter is also related to the vast literature which investigates — both theoretically and empirically — the impact of changes in minimum wages on employment; see Neumark and Wascher (2007) for a survey and Meer and West (2016) for a recent contribution.[4]

Third, our analysis relates to papers which study the role of unemployment benefits in labour markets with matching frictions. In the benchmark model, unemployment benefits determine the outside option of workers and therefore influence wages through bargaining (Pissarides, 2000). Some authors have furt-

[3]Studies which consider the incentive effects of dismissal in the absence of labour market frictions include Stiglitz and Weiss (1983), Spear and Wang (2005), Sannikov (2008) and Wang (2011, 2013).

[4]Our analysis is also related to studies which explore the spillover effects of minimum wages; see for instance Katz and Krueger (1992).

her extended this model in order to study the additional impact on the search incentives of workers (see for instance Acemoglu and Shimer, 1999; recent contributions include Mitman and Rabinovich, 2015, Boadway and Cuff, 2018 and Landais et al., 2018). Our analysis also contributes to a wider literature which analyses the effects of unemployment benefits on labour market outcomes; see Schmieder and von Wachter (2016) for a recent survey article.

The remainder of the chapter is organised as follows. Section 2 introduces the model and outlines parties' preferences, the labour market and the employment relationship. Section 3 then analyses the contracting problem of a single firm-worker pair, while Section 4 discusses the model's equilibrium. In Section 5, we perform some numerical exercises to study the impact of changes in the minimum wage and unemployment benefits. Section 6 concludes.

2 Setup

2.1 Preferences

We consider a discrete-time economy populated by many firms (principals) and many workers (agents), all of whom are infinitely lived and atomistic in nature. All firms are identical, as are all workers. The measure of workers active in the market is constant and normalised to one, while the measure of active firms at time $t \in \mathbb{N}$ will be determined endogenously by freedom of market entry.

Each firm is risk neutral and owns a production technology, which, in each period, requires the effort input a_t of a single worker and creates revenue $\Gamma(a_t)$, where $\Gamma(0) = 0$, $\Gamma' > 0$ and $\Gamma'' \leq 0$. The period t profit of a firm that employs a worker is then simply $\Gamma(a_t)$, net of any wage payment w_t, so that:

$$\pi(a_t, w_t) = \Gamma(a_t) - w_t \tag{1}$$

Alternatively, as is standard in the matching literature, firms who do not employ a worker at time t but choose to remain in the market must pay hiring costs h associated with maintaining an open job vacancy.

Workers' preferences are assumed to be characterised by relative income concerns and can be represented by a utility function which is separable in wages and effort, $U(w_t, a_t; \alpha, w_t^R) = u(w_t; \alpha, w_t^R) - c(a_t)$. The function $u(w_t; \alpha, w_t^R)$ is piecewise linear around a reference wage w_t^R:

$$u(w_t; \alpha, w_t^R) = \begin{cases} w_t + \alpha\left(w_t - w_t^R\right) & \text{if } w_t < w_t^R \\ \\ w_t & \text{if } w_t \geq w_t^R \end{cases} \tag{2}$$

where the parameter $\alpha \geq 0$ measures the extent to which the workers are averse to falling behind the reference wage.[5] Note that setting $\alpha = 0$ yields standard risk-neutrality as a special case. The reference wage is common to all workers and is determined endogenously in equilibrium as the average wage payment in the economy at time t. However, due to their atomistic natures, w_t^R cannot be influenced by any *single* firm or worker and as such is taken by all as given during contracting.[6] The function $c(a_t)$ represents a worker's disutility of undertaking effort and satisfies $c' > 0$, $c'' >$, $c(0) = 0$ and $\lim_{a_t \to 0} c'(a_t) = 0$. Firms and workers both discount the future at a common and constant rate $\beta \in (0, 1)$ and behave as to maximise their lifetime discounted expected profits and utilities, respectively.[7]

Finally, the state imposes a minimum wage, such that any payments from firms to workers cannot be lower than $M \geq 0$. This also prohibits transfers from workers to firms, or negative wage payments. Moreover, the state pays a benefit b to workers who are unemployed, where $0 \leq b \leq M$. For simplicity, we assume that these unemployment benefits are exogenously given and do not consider the mechanism by which they are financed.[8]

2.2 Labour Market Environment

Contracting takes place in a labour market characterised by frictions, which is captured by a simple model of matching (see for instance Pissarides, 2000). In a given time period, each firm is matched with at most one worker, while each worker is matched with at most one firm. At time t, the proportion of unmatched workers is denoted by ρ_t^u. These workers are referred to as being unemployed, with ρ_t^u then representing the economy's unemployment rate. The number of unmatched firms per worker is denoted by ρ_t^v.

The number of new matches in each time period t is determined by the function:

$$\mu\left(\rho_t^u, \rho_t^v\right) \tag{3}$$

which is increasing in both arguments, concave, exhibits constant returns to scale and satisfies $\mu(0, \rho_t^v) = \mu\left(\rho_t^u, 0\right) = 0$.[9] Workers and firms are selected randomly

[5] In the following, for notational convenience we suppress the dependence of $u(\cdot)$ on α and w_t^R where appropriate and simply write $u(w_t)$.

[6] We retain our assumption from the previous chapter regarding the rational-expectations determination of the reference wage in each period.

[7] For simplicity, we assume that workers consume all of their income in each time period and as such do not accumulate wealth over time.

[8] One could rectify this, for example, by imposing a fixed tax upon each employed worker. We also assume that these benefits are independent of employment history.

[9] These are standard requirements for a matching function, see Petrongolo and Pissarides (2001). We assume that only unmatched firms and workers engage in matching; that is, we do not allow for movements from employment to employment. Moreover, we do not consider

from the pools of unemployed workers and vacant jobs, with our assumption of homogeneity ensuring that each match is identical in nature. It is convenient to follow the matching literature and let:

$$\theta_t \equiv \frac{\rho_t^v}{\rho_t^u} \tag{4}$$

denote the ratio of vacant jobs to unemployed workers, or *labour market tightness*. The number of new matches per vacant job then follows from constant returns to scale of the matching function:

$$\frac{\mu\left(\rho_t^u, \rho_t^v\right)}{\rho_t^v} = \mu\left(\frac{\rho_t^u}{\rho_t^v}, 1\right) = \mu\left(\frac{1}{\theta_t}, 1\right) = q(\theta_t) \tag{5}$$

which also represents the probability of a firm with an open vacancy being matched at time t. Similarly, we let the number of new matches per unemployed worker be denoted by:

$$\frac{\mu\left(\rho_t^u, \rho_t^v\right)}{\rho_t^u} = \mu\left(1, \frac{\rho_t^v}{\rho_t^u}\right) = \mu\left(1, \theta_t\right) = \theta_t q(\theta_t) \tag{6}$$

which is also the probability of an unemployed worker finding a match in a given time period.[10]

The flow out of employment in each time period t consists of two elements. First, as is standard in the matching literature, an exogenous proportion φ of the existing matches are destroyed, perhaps due to changes in demand or shocks to productivity. Second, an endogenous proportion δ_t are dissolved due to dismissal of the worker for poor performance; the variable δ_t is therefore determined by the contractual agreements between firm-worker pairs. The total flow out of employment at time t is then equal to $\left(\varphi + (1-\varphi)\,\delta_t\right)(1-\rho_t^u)$. Altogether, the flows into and out of employment are illustrated by Figure 1, along with the quantity of workers and firms who are matched or unmatched at any given time t.

In each time period t, a sequence of three events occurs. First, each firm-worker pair who enter the period matched undergo a time-independent employment relationship, to be discussed subsequently.[11] This results in an instantaneous expected

the case where unemployed workers can expend costly search effort in order to improve their prospects of being matched.

[10]The last equality follows from:

$$\theta_t q(\theta_t) = \theta_t \mu\left(\frac{1}{\theta_t}, 1\right) = \mu\left(1, \theta_t\right)$$

[11]Specifically, for the sake of parsimony, we restrict attention to single-period contracts which do not depend upon the histories of either the firm or worker. It follows that all matches at time t are identical in nature and depend only on the environment in which parties interact at that

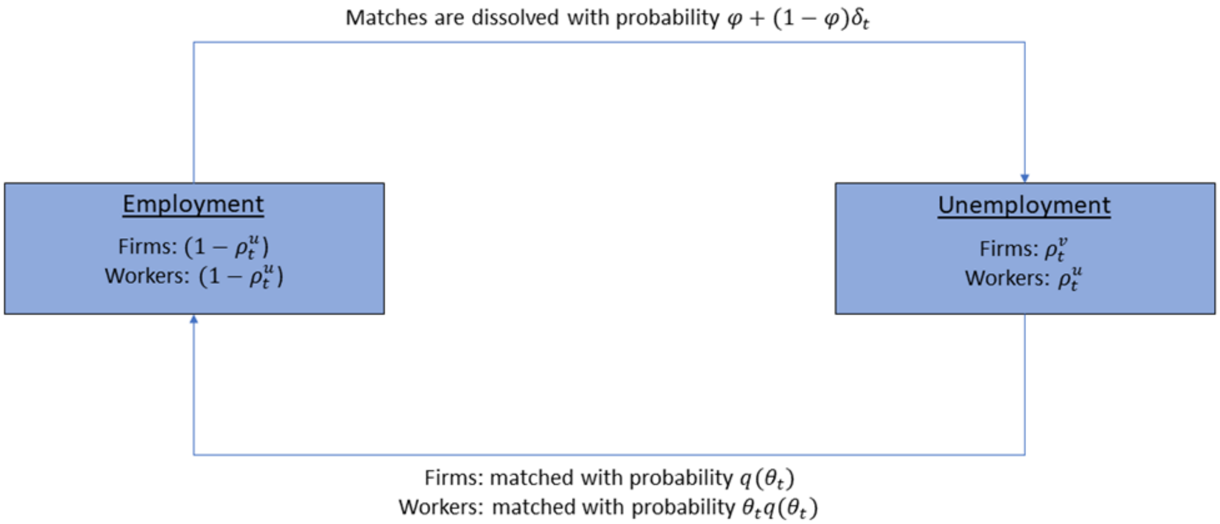

Figure 1: The flows into and out of employment.

utility v_t^e and instantaneous expected profit π_t^e for the worker and firm, respectively. The respective instantaneous utilities and profits of unmatched workers and firms are v_t^u and π_t^u. Second, a proportion $\varphi + (1-\varphi)\,\delta_t$ of the existing matches are dissolved. Firms and workers who exit a match, for either reason, will be unmatched with certainty in period $t+1$. Finally, a number of the firms and workers who have been unmatched throughout period t are chosen to be matched at random, according to the process described above; they will then enter period $t+1$ as matches.

The foregoing discussion allows us to write the lifetime discounted expected profits of a matched and unmatched firm at time t, denoted hereafter by Π_t^e and Π_t^u, respectively:

$$\Pi_t^e = \pi_t^e + \beta\left[(1-\varphi)\,(1-\delta_t)\,\Pi_{t+1}^e + (\varphi + \delta_t - \varphi\delta_t)\,\Pi_{t+1}^u\right] \tag{7}$$

$$\Pi_t^u = \pi_t^u + \beta\left[q(\theta_t)\Pi_{t+1}^e + (1-q(\theta_t))\,\Pi_{t+1}^u\right] \tag{8}$$

The first term in (7) is a firm's instantaneous expected profit associated with employing a worker; the second term is then the discounted future expected profit. The square bracket in (7) is explained as follows. With probability $(1-\varphi)\,(1-\delta_t)$, the firm will remain matched in the subsequent period $t+1$, which, at that time, yields the discounted expected profit Π_{t+1}^e. Alternatively, with probability $\varphi + \delta_t - \varphi\delta_t$, the firm's match is dissolved so that it is unmatched in period $t+1$, yielding discounted expected profit Π_{t+1}^u. The equation (8) can be explained similarly, with $q(\theta_t)$ representing the probability of a firm with a vacancy in period

time.

t being matched and therefore employing a worker in period $t+1$.

By a similar logic, the respective lifetime discounted expected utilities of an employed and unemployed worker at time t are:

$$V_t^e = v_t^e + \beta \left[(1-\varphi)(1-\delta_t) V_{t+1}^e + (\varphi + \delta_t - \varphi\delta_t) V_{t+1}^u \right] \tag{9}$$

$$V_t^u = v_t^u + \beta \left[\theta_t q(\theta_t) V_{t+1}^e + (1 - \theta_t q(\theta_t)) V_{t+1}^u \right] \tag{10}$$

The intuitions for the expressions (9) and (10) are similar to the foregoing case, with the main difference being that an unemployed worker's probability of becoming matched is $\theta_t q(\theta_t)$, rather than $q(\theta_t)$ for an unmatched firm.

Finally, we are able to express the change in the unemployment rate at time t as:

$$\frac{d\rho_t^u}{dt} = (\varphi + \delta_t - \varphi\delta_t)(1-\rho_t^u) - \rho_t^u \theta_t q(\theta_t) \tag{11}$$

where the first and second terms represent the flows into and out of unemployment respectively, as described previously. A steady-state equilibrium in the labour market is characterised by two conditions. First, the unemployment level is constant, so that $\frac{d\rho_t^u}{dt} = 0$.[12] Rearranging (11) then yields the steady-state unemployment level:

$$\rho_t^u = \frac{\varphi + \delta_t - \varphi\delta_t}{\varphi + \delta_t - \varphi\delta_t + \theta_t q(\theta_t)} \tag{12}$$

Second, freedom of market entry requires that the lifetime discounted expected profit of a firm which chooses to open a vacancy is zero. Mathematically, this condition requires that:

$$\Pi_t^u = 0 \tag{13}$$

Our subsequent analysis will restrict attention to the steady-state equilibrium in the labour market. Along with our assumption that employment relationships between firms and workers are time-independent, this ensures that all variables become constant over time in equilibrium. Accordingly, for simplicity, we omit all time subscripts t for the remainder of the chapter.

2.3 The Employment Relationship

Employment relationships are time-independent and similar to those discussed in the previous chapter, with two key differences: all wage payments must be greater

[12] By (6), we have $\rho_t^u \theta_t q(\theta_t) = \mu(\rho_t^u, \rho_t^v)$ so that the second term on the RHS of (11) is simply the number of new matches at time t. The stationarity condition $\frac{d\rho_t^u}{dt} = 0$ therefore implies that ρ_t^v also becomes constant over time in equilibrium.

than the economy's minimum wage and the firm can additionally use dismissal to create incentives. As before, the worker is paid a wage in return for exerting effort, with moral hazard occurring because this effort is non-verifiable. Instead, the firm has access to a monitoring technology which operates as follows. Once the worker has produced effort a, a verifiable proxy variable $x \in [0, 1]$ is generated with exogenous probability $m \in (0, 1)$. In this case, we say that monitoring has succeeded. It is common knowledge that the realisation of x is drawn from a time-independent distribution function $F(x; a)$, with associated density $f(x; a) > 0$ over the support. In addition, F is assumed to satisfy the *strict Monotone Likelihood Ratio Property* (MLRP) and the *strict Convexity of the Distribution Function Condition* (CDFC).[13] Alternatively, with probability $(1 - m)$ monitoring fails and no information (denoted by $\emptyset$) is produced.[14]

The firm has two available channels by which the monitoring process can be used to create effort incentives. First, the worker can be dismissed for poor performance, *i.e.* a low realisation of x. As long as an employed worker receives a higher discounted expected utility than an unemployed worker, this will create incentives to undertake effort. Second, as is standard in the moral hazard literature, the firm can condition wages on the information (if any) that is observed. We also make the standard assumption that the firm is assumed to hold all of the bargaining power.

The order of the game for the employment relationship is then as follows. First, the firm offers a take-it-or-leave-it contract to the worker. If the contract is rejected, then the match is immediately dissolved and both parties receive instantaneous payoffs equal to those of firms and workers who were not initially matched (*i.e.* π^u and v^u respectively). If the contract is accepted, the worker then chooses a level of effort to undertake. Next, the monitoring process either succeeds and produces information x, or fails and produces no information $\emptyset$. Finally, based on the outcome of monitoring, the firm pays the appropriate wage and possibly terminates the worker's employment, as outlined by the contract. It is assumed that the firm can feasibly commit to dismissal for poor performance and cannot renege on this obligation once effort has been undertaken.

Finally, as outlined previously, unemployed workers receive a benefit payment b, while firms who do not employ a worker but continue to remain in the market must pay hiring costs h. This implies the following instantaneous utility and profit

[13] Strict MLRP states that $\frac{f_a}{f}(x; a)$ is strictly increasing in x for all $x \in (0, 1)$, while strict CDFC requires $F_{aa}(x; a) > 0$ for all $x \in (0, 1)$. See also the discussion in the previous chapter.

[14] See the previous chapter for a detailed discussion of this monitoring process.

levels:

$$v^u = (1+\alpha)\, b - \alpha w^R \tag{14}$$

$$\pi^u = -h \tag{15}$$

where (14) follows from (2) and the fact that w^R will be determined by the average wage in the economy and, therefore, $b \leq M \leq w^R$.

3 Contracting

The previous chapter derived the optimal wage scheme for a worker whose preferences can be represented by the utility function (2). When the reference wage is defined by the average wage in the economy, it was shown that this contract was ternary, with three distinct wage payments. While that environment did not explicitly feature unemployment, we discussed that one possible interpretation of this result could be a binary wage scheme (i.e. a base wage which is always paid, as well as a possible bonus payment for high performance) along with the possibility of dismissal for sufficiently poor performance. While such a contract features only two distinct possible wage payments, it is nonetheless perceived by workers as being ternary since, from their point of view, there are three distinct possible outcomes. In this chapter, we explore this possibility in an environment which does explicitly feature unemployment, allowing us to study how changes in labour market policies impact contracting between parties.

As before, we first study the firm's optimal single-period contractual choice in order to implement a given effort level a at the lowest possible cost, before later considering the firm's optimal effort choice. In line with the foregoing discussion, we restrict attention to wage schemes which take the form:

$$w(x) = \begin{cases} W & x \in [0, z_2) \\ \\ W + B & x \in [z_2, 1] \end{cases} \tag{16}$$

where the base wage W, the bonus payment B and the critical performance level z_2 are decision variables for the firm.[15] Moreover, the firm must also choose an additional critical performance level, z_1, which represents the 'lowest acceptable' performance for a worker. That is, outcomes $x < z_1$, observed with frequency $mF(z_1; a)$, will result in dismissal. The probability of a match being dissolved

[15] We do not investigate whether such a wage scheme would be optimal in the current framework.

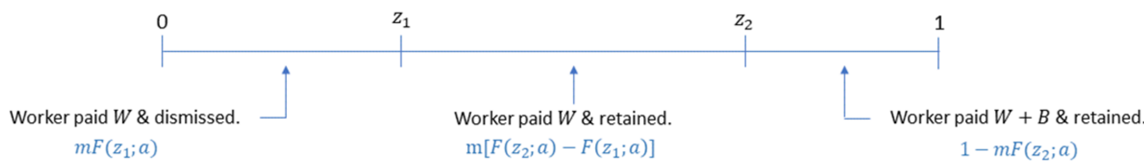

Figure 2: The payment and dismissal decision associated with each realisation of the performance signal $x \in [0, 1]$, along with their respective probabilities. In case of no information, which occurs with probability $(1 - m)$, the worker is paid $W + B$ and retained for employment in the next period.

due to dismissal of a worker is then:

$$\delta = mF(z_1; a) \tag{17}$$

As long as the worker's expected utility from being employed in each period is greater than that from unemployment, dismissal will create further effort incentives. It is assumed that the effort level a to be implemented is always sufficiently large that explicit incentive pay is required and therefore $B > 0$.[16] The case in which monitoring fails and no information is generated is treated identically to a realisation of $x \geq z_2$; that is, the employee is not dismissed and the wage received is $W + B$.[17] Altogether, the contractual form is summarised by Figure 2, which illustrates the payment and dismissal decision associated with each realisation of the performance measure $x \in [0, 1]$.

In this environment, an incentive feasible contract is a tuple $\mathcal{C} = \{W, B, z_1, z_2\}$ which induces the worker to undertake the firm's desired effort level a. The firm's objective is therefore to design the wage scheme which maximises lifetime discounted expected profits, subject to the following constraints:

1. The contract must induce the worker to undertake effort a. Since workers behave as to maximise their discounted expected lifetime utility, the incentive compatibility constraint then requires that a maximises V^e.

2. The contract must induce the worker to accept the contract. Since rejecting the contract results in unemployment for a worker, the participation constraint then requires that $V^e \geq V^u$.

3. All possible wage payments must be weakly greater than the minimum wage:

[16] For low effort levels, it may be possible to pay a fixed wage and provide adequate incentives solely via the threat of dismissal; we ignore these cases. In addition, note that we rule out the counter-intuitive case of a worker simultaneously receiving a bonus payment and being dismissed from employment; as we shall see later, such a result never occurs for the numerical examples we consider.

[17] This assumption is similar in nature to the requirement that $w(x) \leq w_\emptyset$ in the previous chapter. See the discussion there for an intuitive justification.

$W, W + B \geq M$.

We solve for the firm's choice of contract in three steps. First, we derive the bonus payment. From the foregoing discussion, an employed worker's instantaneous expected utility is given by:

$$v^e = mF(z_2; a)u(W) + [1 - mF(z_2; a)]\, u(W + B) - c(a) \tag{18}$$

In equilibrium, each worker's reference wage will be determined by the average wage in the economy. It then follows that $W \leq w^R \leq W + B$, so that $u(W) = (1 + \alpha)\, W - \alpha w^R$ while $u(W + B) = W + B$. Accordingly, Condition 1 requires that:

$$\begin{aligned} a = \arg\max_{\hat{a}}\ & mF(z_2; \hat{a}) \left((1 + \alpha)\, W - \alpha w^R\right) + [1 - mF(z_2; \hat{a})]\, (W + B) - c(\hat{a}) \\ & + \beta\, [(1 - \varphi)\, (1 - mF(z_1; \hat{a}))\, V^e + (\varphi + (1 - \varphi)\, mF(z_1; \hat{a}))\, V^u] \end{aligned} \tag{19}$$

The RHS of (19) represents the worker's discounted expected lifetime utility (9), where we have substituted δ and v^e using (17) and (18). Note that while a worker's effort at time t influences both their instantaneous utility and the *probability* of employment or unemployment in the subsequent period, it does not affect the *value* of this employment or unemployment. Accordingly, the terms V^e and V^u in (19) do not vary with the worker's effort and are therefore taken as given.[18] Our restrictions on the distribution function imply the validity of the first-order approach, so that we can substitute (19) for its first-order condition. Setting this equal to zero and rearranging then yields the necessary bonus payment:

$$B = \alpha \left(W - w^R\right) + \frac{c'(a) + \beta\, (1 - \varphi)\, mF_a(z_1; a)\, (V^e - V^u)}{-mF_a(z_2; a)} \tag{20}$$

Second, we show that the worker will always prefer to accept the contract for any $z_1 \in [0, 1]$, $z_2 \in [0, 1]$ and $W \geq M$, so that the participation constraint (and therefore condition 2) is satisfied.

Lemma 1. *The worker prefers to accept the contract for any base wage $W \geq M$.*

Proof. From (9) and (10), the steady-state equilibrium values of V^e and V^u are

[18] Slightly abusing notation and reintroducing time-subscripts for clarity, the constraint can be written as:

$$a_t = \arg\max_{\hat{a}_t}\ V_t^e(\hat{a}_t) = v_t^e(\hat{a}_t) + \beta\, [(1 - \varphi)\, (1 - \delta_t(\hat{a}_t))\, V_{t+1}^e + (\varphi + (1 - \varphi)\, \delta_t(\hat{a}_t))\, V_{t+1}^u]$$

where V_{t+1}^e and V_{t+1}^u clearly do not depend on the effort at time t, $\hat{a}_t$.

given by:

$$V^e = v^e + \beta \left[(1-\varphi)(1-\delta) V^e + (\varphi + \delta - \varphi\delta) V^u \right] \tag{21}$$

$$V^u = v^u + \beta \left[\theta q(\theta) V^e + (1 - \theta q(\theta)) V^u \right] \tag{22}$$

respectively. Rearranging these expressions and simplifying then yields:

$$V^e - V^u \geq 0 \tag{23}$$

$$\Leftrightarrow \qquad \frac{v^e - v^u}{1 - \beta(1-\delta)(1-\varphi) + \beta\theta q(\theta)} \geq 0 \tag{24}$$

so that $V^e - V^u \geq 0$ if and only if $v^e - v^u \geq 0$. Using (14) and (18), the latter inequality can be expressed as:

$$mF(z_2;a)\left((1+\alpha)W - \alpha w^R\right) + \left[1 - mF(z_2;a)\right](W + B) \\ - c(a) - \left[(1+\alpha) b - \alpha w^R\right] \geq 0 \tag{25}$$

Rearranging this then yields:

$$(1+\alpha)(W-b) + \left[1 - mF(z_2;a)\right]\alpha\left(w^R - W\right) \\ + \left[1 - mF(z_2;a)\right] B - c(a) \geq 0 \tag{26}$$

Note that the first three terms of (26) are positive, so that the inequality is satisfied in particular when the worker chooses to implement zero effort ($c(0) = 0$). Incentive compatibility then implies that the inequality must also hold when the worker implements the firm's desired effort level. □

Intuitively, the proof of Lemma 1 shows that $V^e \geq V^u$ is equivalent to the requirement that $v^e \geq v^u$. Since we have $W \geq M \geq b$, the minimum payment received by an employed worker cannot be less than the unemployment benefit b. This implies that the instantaneous utility of an employed worker who undertakes zero effort cannot be less than that of an unemployed worker. Employed workers who undertake the firm's desired effort level a must then also receive a weakly greater instantaneous utility, due to incentive compatibility, so that $v^e \geq v^u$ is satisfied.

Third, we derive the values of W, z_1 and z_2 which maximise the firm's discounted expected lifetime profit V^e, subject to the minimum wage constraint. Note that the firm's instantaneous profit π^e is simply the value of output produced,

$\Gamma(a)$, net of expected wage costs:

$$\pi^e = \Gamma(a) - W - [1 - mF(z_2;a)]\, B \tag{27}$$

Inserting (17), (20) and (27) into (9) then yields the firm's optimisation problem:

$$\begin{aligned}\min_{W,z_1,z_2} \; & \Gamma(a) - W - [1 - mF(z_2;a)]\, \alpha \left(W - w^R\right) \\ & - [1 - mF(z_2;a)] \frac{c'(a) + \beta\,(1-\varphi)\, mF_a(z_1;a)\,(V^e - V^u)}{-mF_a(z_2;a)} \\ & + \beta\,[(1-\varphi)\,(1 - mF(z_1;a))\,\Pi^e + (\varphi + (1-\varphi)\, mF(z_1;a))\,\Pi^u]\end{aligned} \tag{28}$$

s.t. $W \geq M$

Note that since the term W enters the firm's objective function negatively — and since the worker's participation is guaranteed for any $W \geq M$ — the firm will set W as low as possible so that the minimum wage constraint binds and $W = M$.[19] Using this result, the first order conditions with respect to z_1 and z_2 can be expressed as follows:

$$\frac{1 - mF(z_2;a)}{mF_a(z_2;a)} \cdot \frac{f_a(z_1;a)}{f(z_1;a)} - \frac{\Pi^e - \Pi^u}{V^e - V^u} = 0 \tag{29}$$

$$\begin{aligned}&[-mf(z_2;a)] \left[\alpha\left[W - w^R\right] + \frac{c'(a) + (1-\varphi)\,\beta m F_a(z_1;a)\,[V^e - V^u]}{-mF_a(z_2;a)}\right] \\ &+ [1 - mF(z_2;a)] \left[\frac{[c'(a) + (1-\varphi)\,\beta m F_a(z_1;a)\,[V^e - V^u]]\, mf_a(z_2;a)}{[mF_a(z_2;a)]^2}\right] = 0\end{aligned} \tag{30}$$

where the terms V^e, V^u, Π^e and Π^u are taken as given by the firm, analogous to the case of the worker's optimisation problem (19).[20]

4 Equilibrium

In equilibrium, the reference wage is determined by the average wage in the economy, $W + [1 - mF(z_2;a)]\, B$. Using (20), along with the fact that each firm will

[19]Intuitively, this results since the current period base wage cannot be used to create incentives, as all contracts are single-period in nature.

[20]Intuitively, these terms are unaffected by the firm's contractual decisions; see also footnote 18.

set $W = M$, we therefore have:

$$
\begin{aligned}
w^R = M &+ [1 - mF(z_2;a)]\,\alpha\left(M - w^R\right) \\
&+ [1 - mF(z_2;a)]\,\frac{c'(a) + \beta\,(1-\varphi)\,mF_a(z_1;a)\,(V^e - V^u)}{-mF_a(z_2;a)}
\end{aligned}
\tag{31}
$$

Rearranging this then yields:

$$
w^R = M + [1 - mF(z_2;a)]\,\frac{c'(a) + \beta\,(1-\varphi)\,mF_a(z_1;a)\,(V^e - V^u)}{-mF_a(z_2;a)\,[1 + \alpha\,[1 - mF(z_2;a)]]} \tag{32}
$$

Using (32), the first-order conditions associated with the firm's choice of critical values, (29) and (30), can be rewritten as:

$$
\frac{1 - mF(z_2;a)}{mF_a(z_2;a)} \cdot \frac{f_a(z_1;a)}{f(z_1;a)} - \frac{\Pi^e - \Pi^u}{V^e - V^u} = 0 \tag{33}
$$

$$
mf(z_2;a)F_a(z_2;a) + [1 + \alpha\,[1 - mF(z_2;a)]]\,[1 - mF(z_2;a)]\,f_a(z_2;a) = 0 \tag{34}
$$

The critical value z_2 does not affect the probability of dismissal; it is therefore chosen by the firm in order to minimise current-period expected wage costs and, from (34), is determined by the quality of monitoring m and the relative income concerns parameter α alone. In contrast, the firm's choice of z_1 will determine the dismissal probability and therefore by (33) depends on the variables V^e, V^u, Π^e and Π^u, each of which vary with the underlying labour market conditions.

From the firm's point of view, the possibility of the worker's dismissal creates two effects. First, dismissal entails a loss of discounted expected utility equal to $V^e - V^u$ for the worker. This creates positive effort incentives, since workers undertake higher effort in an attempt to avoid dismissal, allowing the firm to reduce the size of the bonus payment and therefore increase instantaneous expected profits. Second, dismissal of the worker causes the match to be dissolved. Due to the frictional nature of the labour market, the firm must then pay hiring costs and forego the profits from employment for an uncertain number of periods, until they are rematched with another worker. This leads to an expected reduction in discounted expected profits of $\Pi^e - \Pi^u$. These two effects can be respectively thought of as the benefits and costs of using dismissal to create incentives and will determine a firm's optimal choice of z_1. In particular, the effects will vary with both the minimum wage in the economy, M, and the magnitude of unemployment benefits, b.

The equations (7), (8), (9), (10), (14), (15), (17), (18), (20), (27) and (32) form a linear system which can be solved for the variables Π^e, Π^u, V^e, V^u, π^e, π^u,

v^e, v^u, B, w^R and δ. In addition, the values of z_1 and z_2 are determined by (33) and (34), while the steady state equilibrium unemployment rate ρ^u and vacancy per worker ρ^v are determined by the labour market equilibrium conditions (12) and (13). Altogether, these equations characterise the solution to the model.

5 Numerical Analysis

We wish to explore how variations in the underlying parameters of the model impact the economy's equilibrium. However, the complexity of the equation system prevents us from proceeding analytically. As such, we instead proceed by solving a series of numerical examples. To that end, we let $\Gamma(a) = \gamma a$, where γ is a measure of productivity, and assume that $c(a) = 0.5a^2$. In addition, we let the distribution function be defined by $F(x; a) = x^a$ and apply the transformation $y_1 = z_1^a$, which enables us to write $F_a(z_1; a) = \frac{1}{a}(y_1 \ln y_1)$. A similar transformation is applied for z_2; we then treat y_1 and y_2 as the firm's decision variables. Finally, we let the economy's matching function μ take the Cobb-Douglas form $\mu(\rho_t^u, \rho_t^v) = \overline{\mu}(\rho_t^u)^{0.5}(\rho_t^v)^{0.5}$, where $\overline{\mu}$ is a parameter which captures the efficiency of the matching process. It then follows that $q(\theta) = \overline{\mu}\theta^{-0.5}$.

While the previous section analysed a firm's contracting decision in order to implement a given effort level a, we slightly extend the model to allow effort to be determined endogenously. As with the variables W, z_1 and z_2, each firm will choose a in order to maximise discounted expected lifetime profits. Given the above specifications and the transformations $F(z_i; a) = y_i$ and $F_a(z_i; a) = \frac{1}{a}(y_i \ln y_i)$, the firm's optimisation problem (28) can be re-expressed as:

$$\begin{aligned} \max_{y_1, y_2, a} \quad & \gamma a - M - [1 - my_2]\,\alpha\left(M - w^R\right) \\ & - [1 - my_2]\,\frac{a^2 + \beta(1-\varphi)\,my_1 \ln y_1 (V^e - V^u)}{-my_2 \ln y_2} \\ & + \beta\left[(1-\varphi)(1 - my_1)\,\Pi^e + (\varphi + (1-\varphi)\,my_1)\,\Pi^u\right] \end{aligned} \tag{35}$$

where $W = M$. The first order condition associated with a is then:

$$\gamma - [1 - my_2]\,\frac{2a}{-my_2 \ln y_2} = 0 \tag{36}$$

Note that the optimal a is therefore determined solely by the parameters γ, m and α (via y_2); neither the minimum wage nor employment benefits play a role. Intuitively, a change in either of these parameters shifts the firm's costs of creating effort incentives, but leaves the marginal cost unchanged; accordingly, there is no effect on the optimal a. This is helpful, since it reduces the complexity of the model and allows us to delineate the effects of a change in either parameter with

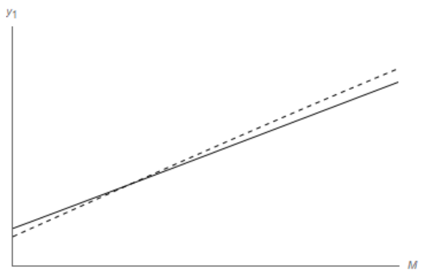

Figure 3: The impact of a change in M on y_1.

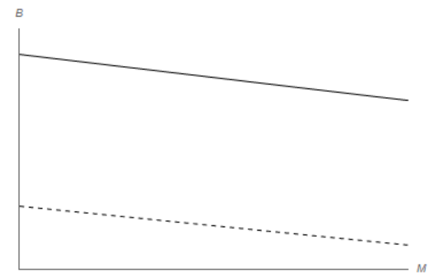

Figure 4: The impact of a change in M on B.

increased clarity. However, it is possible that with other distribution functions the optimal effort level may respond to changes in these parameters, so that this is not a general feature of the model.

In the following, we use Mathematica to solve the model for a series of numerical examples. Throughout, we assume that $m = 0.9$, $\beta = 0.8$, $h = 300$, $\gamma = 100$, $\varphi = 0.05$ and $\overline{\mu} = 0.2$. We then consider variations in M and b for the cases where $\alpha = 0$ (solid lines) and $\alpha = 1$ (dashed lines); this allows us to analyse the effect of relative income concerns on the economy's equilibrium. Specifically, as a benchmark we let $M = 800$ and $b = 100$. The first part of the analysis then considers variations of M from this benchmark between 700 and 1000, while the second part considers variations of b between 0 and 300.[21]

5.1 Changes in M

5.1.1 Contracting

We first consider the implications of an increase in M for contracting between a firm-worker pair. From (34) and (36), neither y_2 nor a are affected by such a change. As discussed previously, when deciding on the optimal value of y_1, the firm compares the worker's gain in discounted expected utility from employment, $V^e - V^u$, with its own increase in discounted expected profits from employing a worker, $\Pi^e - \Pi^u$. The ratio of these terms then determines the choice of y_1 by (33). As the worker's benefit from being employed increases, dismissal becomes more effective as a tool for creating incentives and is therefore used more often. On the other hand, the greater the firm's increase in profit from having an employee, the more expensive dismissal becomes as a way to induce effort (i.e. in terms of

[21]Note that since the purpose of the analysis is not to generate labour market outcomes consistent with real world data, these parameter values have not been chosen on the basis of empirical observations. Instead, our purpose is to demonstrate some possible implications of changes in labour market policies for incentive contracting when firms can use dismissal to create incentives.

foregone future profits) and therefore the probability of termination is lower.

With this in mind, there are two *ceteris paribus* effects of an increase in the economy's minimum wage. First, the term $V^e - V^u$ increases. Clearly, employed workers receive higher wages, raising the instantaneous expected utility associated with employment, v_e. Moreover, this effect is particularly strong when workers have relative income concerns. Since the reference wage of individuals is determined by the average payment to matched workers, which has increased, unemployed workers who receive an unchanged benefit b find themselves further behind w^R and therefore suffer a utility loss, reducing v^u. As such, an increase in M widens the discounted expected utility gap between employed and unemployed workers, so that dismissal becomes increasingly more painful for employees. Second, higher wages imply that a firm's profits when employing a worker are now reduced, whereas the payoff of an unmatched firm is unchanged. It follows that the term $\Pi^e - \Pi^u$ is reduced.

The above discussion implies that dismissal becomes both more effective and less expensive following an increase in M and therefore becomes a more efficient method of creating incentives. Accordingly, Figure 3 shows the increase in y_1 for our specific numerical example as M is increased, so that workers are dismissed more often.[22] This is true in the case of both standard preferences ($\alpha = 0$; solid line) and relative income concerns ($\alpha = 1$; dashed line). However, the curve is steeper — and therefore more sensitive to variations in M — in the $\alpha = 1$ case, due to the additional impact of relative income concerns on the instantaneous utility of unemployed workers, v^u, via the reference wage.

Figure 4 shows that in both cases, the contract's bonus payment is reduced. Since dismissal of workers becomes a more effective way to create incentives, and is therefore used more often, workers are willing to exert higher effort in an attempt to avoid termination of their employment. Moreover, as the effort level to be induced remains constant, this allows for lower explicit incentive pay, so that the bonus is reduced. Finally, note that the required bonus payment for any given value of M is lower in the relative income concerns case. Intuitively, this follows from the fact that aversion to falling behind the reference wage helps the firm to create effort incentives at a lower costs, as shown in the previous chapter.

5.1.2 Utility and Profit

Figure 5 illustrates the impact of an increase in M on the average wage in the economy and therefore the reference wage of workers. For our example, despite the decrease in the bonus, the rise in M is sufficient to increase the contract's expected wage payment. The change in the instantaneous utility of workers is then shown

[22]Note that since a is unchanged and $y_1 = F(z_1; a) = z_1^a$, this also implies that z_1 increases.

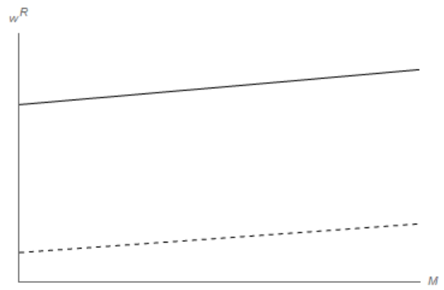

Figure 5: The impact of a change in M on w^R.

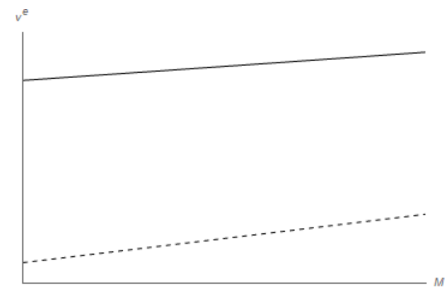

Figure 6: The impact of a change in M on v^e.

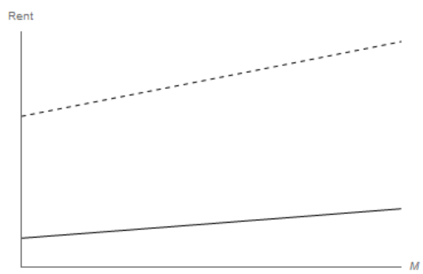

Figure 7: The impact of a change in M on the rent of an employed worker.

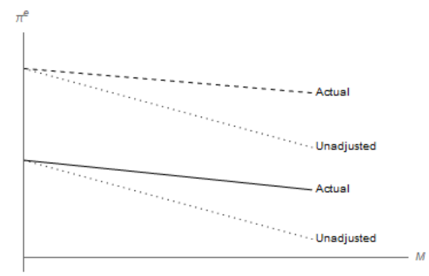

Figure 8: The impact of a change in M on π^e.

by Figure 6. In both cases, v^e is increasing. However, the increase is at a faster rate when $\alpha = 1$ due to the higher marginal utility associated with payments below the reference wage. Next, the change in a worker's relative benefit of being employed (i.e. the term $v^e - v^u$), hereafter referred to as an employed worker's *rent*, is illustrated by Figure 7. Since the income of unemployed workers is constant and equal to b, v^u is unchanged in the case of standard preferences; accordingly, the increase in rent is then driven solely by the increase in v^e outlined above. However, as discussed previously, when workers have relative income concerns, the rise in the reference wage w^R implies that v^u is reduced following the increase in M. This reduction, along with the increase in v^e, causes a much steeper rise in the rent of employed workers in the $\alpha = 1$ case.

Finally, we consider the implications for the profit of firms who employ a worker. Clearly, since the average wage payment w^R has increased and effort remains unchanged, firms will make lower profits following an increase in M. However, due to the positive effect on incentives, it should be noted that the magnitude of the marginal reduction in profit as M rises is less than unity. That is, for every £1 by which the minimum wage increases, the expected profit of a firm is reduced by less than £1. This is illustrated by Figure 8, which should be

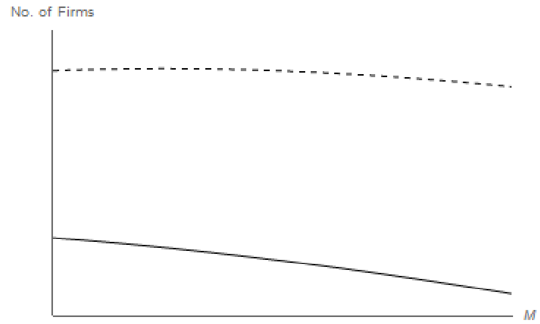

Figure 9: The impact of a change in M on the number of active firms in the economy.

read as follows. The solid and dashed lines show the actual reductions in profit, for the cases where $\alpha = 0$ and $\alpha = 1$, respectively. The dotted lines below then plot the profit levels in the absence of any incentive effect, so that the bonus payment remains unadjusted.[23] It follows that while minimum wages do indeed leave firms worse off, they are partially cushioned by the impact due to an efficiency wage effect, which allows them to reduce explicit incentive payments.

5.1.3 Labour Market

We now consider the impact of variations in M on equilibrium in the labour market. The number of active firms in the economy is determined by the freedom of market entry condition (13).[24] Combining (13) with (8) and (15) yields:

$$\frac{h}{\beta} = q(\theta)\Pi^e \tag{37}$$

Clearly, the left-hand side of (37) is invariant to changes in M, so that the change in the number of firms is determined by the right-hand side of the equation alone. Recall that $q(\theta)$ represents the probability of an unmatched firm finding a worker in a given time period, while Π^e denotes the discounted expected profit of a matched firm. With this in mind, note that an increase in M has two effects. First, as previously discussed, a raise in the minimum wage reduces the instantaneous profit of a matched firm, π^e. This also implies that Π^e decreases. Second, since firms increase y_1 after an raise in M, as shown by Figure 3, workers are dismissed with a higher probability; this increases the flow of workers into unemployment and therefore improves an unmatched firm's chances of finding a match. These two

[23]Technically, since we are considering variations in M between 700 and 1000, the origin of the graph in Figure 8 is associated with $M = 700$. Slightly abusing notation by writing profit as a function of the minimum wage, the lines labelled 'actual' plot $\pi^e(M)$ for the cases where $\alpha = 0$ and $\alpha = 1$, while the lines labelled 'unadjusted' plot $\pi^e(700) - (M - 700)$.

[24]The number of active firms in the economy is given by the term $(1 - \rho^u) + \rho^v$.

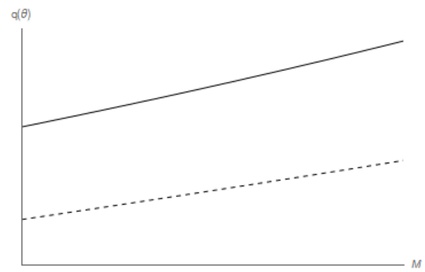

Figure 10: The impact of a change in M on $q(\theta)$.

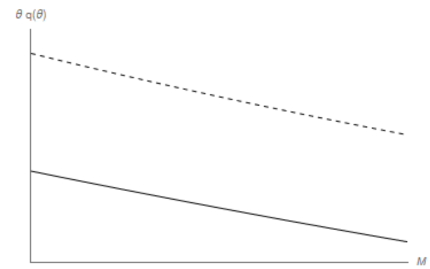

Figure 11: The impact of a change in M on $\theta q(\theta)$.

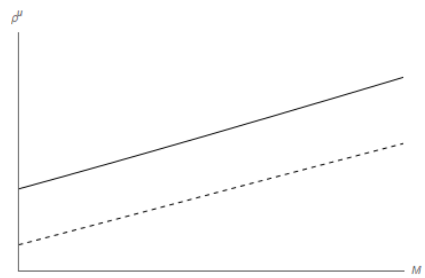

Figure 12: The impact of a change in M on ρ^u.

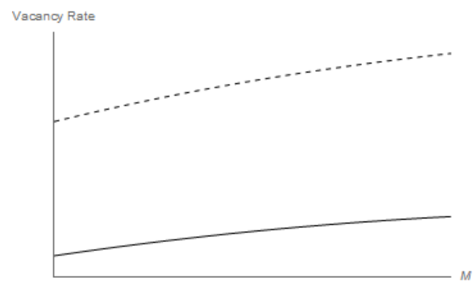

Figure 13: The impact of a change in M on the economy's vacancy rate.

effects have an opposing impact on the number of active firms in the economy. Figure 9 shows that in our example, the former effect is dominant so that the number of firms decreases. However, the decrease is much less pronounced in the $\alpha = 1$ case. Intuitively, this occurs since, by Figure 3, y_1 is more responsive to variations in M when workers have relative income concerns. Accordingly, the flow into unemployment is higher in this case, increasing the probability that an unmatched firm becomes matched (i.e. strengthening the latter of these two effects).

The above discussion implies that $q(\theta)$ must increase in response to a higher M; this is illustrated by Figure 10. Not only are matches now dissolved at a higher rate, there are also less firms in the economy to compete with. In contrast, the probability of an unmatched worker finding employment, $\theta q(\theta)$, falls following an increase in M, as shown by Figure 11. This follows from the fact that there are a fixed number of workers in the economy, while the number of firms has decreased.

Finally, Figures 12 and 13 illustrate the impact of a change in M on the economy's unemployment rate and vacancy rate, respectively.[25] Unemployment

[25]Note that ρ^u denotes the number of unemployed workers per worker and therefore captures the economy's unemployment rate. In contrast, ρ^v denotes the number of unmatched firms per worker. While ρ^v is commonly referred to as the vacancy rate in the literature, Figure 13 plots

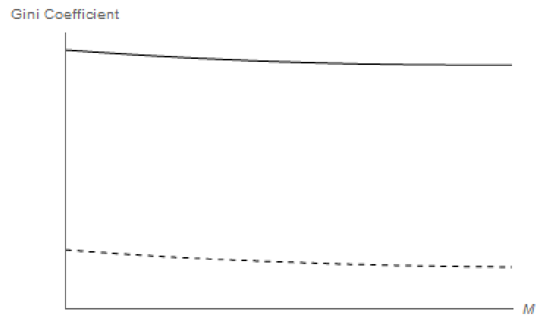

Figure 14: The impact of a change in M on the economy's Gini Coefficient.

clearly rises, since there are less firms active in the economy and existing matches are dissolved at a greater rate. Moreover, the same effects imply that the vacancy rate in the economy will also increase.

5.1.4 The Gini Coefficient

To conclude our analysis of the impact of a variation in M, we consider the economy's Gini Coefficient. Specifically, we consider three distinct groups in the economy: unemployed workers who receive an income b, employed workers who receive an income M and employed workers who receive an income $M + B$. The respective frequencies of these groups are given by ρ^u, $(1 - \rho^u)mF(z_2; a)$ and $(1 - \rho^u)[1 - mF(z_2; a)]$. Using this information, the Gini Coefficient amongst workers in the economy can be calculated for each particular value of M. Figure 14 then plots this graph.

In both the $\alpha = 0$ and $\alpha = 1$ cases, we can see that there is a marginal decrease in the Gini Coefficient as M rises. This is a result of several counteracting forces. While b remains unchanged, the increase in the unemployment rate ρ^u implies that more workers receive this level of income. In addition, the terms M and $M + B$ each increase. Both of these effects increase inequality. However, among employed workers, there is a decrease in inequality since the wage schedule becomes compressed due to the reduction in B. Accordingly, the distance between M and $M + B$ is reduced. This is shown by Figure 14 to be the dominant effect on the Gini Coefficient, which is reduced in both cases.

---

the number of unmatched firms per firm, i.e. the term:

$$\frac{\rho^v}{(1 - \rho^u) + \rho^v}$$

which we refer to throughout as the economy's vacancy rate.

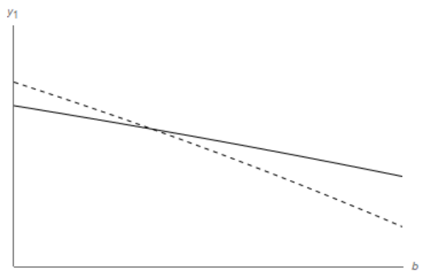

Figure 15: The impact of a change in b on y_1.

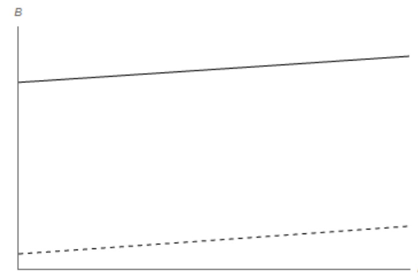

Figure 16: The impact of a change in b on B.

5.2 Changes in b

5.2.1 Contracting

As before, we begin our analysis of the impact of an increase in the unemployment benefit b by considering the implications for contracting between a firm-worker pair. Similar to the case of M, neither y_2 nor a are affected by such a change. The impact on a firm's choice of y_1 will again depend on how the terms $V^e - V^u$ and $\Pi^e - \Pi^u$ respond to variations. The key effect of an increase in b is to decrease the term $V^e - V^u$. Intuitively, unemployed workers receive a higher payment and as such the income difference between them and their employed counterparts is reduced. This effect is especially pronounced when workers have relative income concerns, due to the high marginal utility below the reference wage.

The result of the increase in b is therefore to reduce the effectiveness of dismissal as a tool for creating incentives. Accordingly, Figure 15 shows that firms reduce y_1 in our numerical example, so that workers face a lower probability of dismissal from employment.[26] As before, the cases of standard preferences ($\alpha = 0$) and relative income concerns ($\alpha = 1$) are represented by the solid and dashed lines, respectively. In either instance, y_1 is lowered. However, the curve is steeper when $\alpha = 1$, as y_1 becomes more sensitive to changes in b due to the aforementioned high marginal utility.

Figure 16 then shows that the contract's bonus payment is increased. Intuitively, this occurs since dismissal of workers is less effective and is therefore used less often. Workers are less willing to exert effort in order to avoid termination, as the utility associated with being unmatched has increased. Moreover, since the firm continues to induce the same amount of effort, explicit incentives (i.e. the bonus payment) must now be raised.

[26]This is in contrast to the findings of Wang and Yang (2015). In their model, workers are only dismissed once they become too poor to motivate. Increases in unemployment benefits *ceteris paribus* reduce incentives, so that the threshold for termination must be increased, resulting in a higher flow into unemployment.

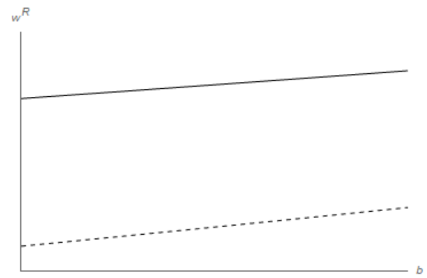

Figure 17: The impact of a change in b on w^R.

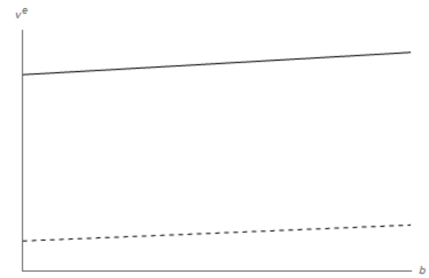

Figure 18: The impact of a change in b on v^e.

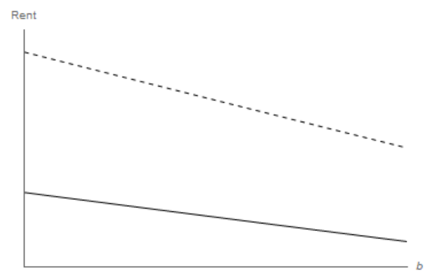

Figure 19: The impact of a change in b on the rent of an employed worker.

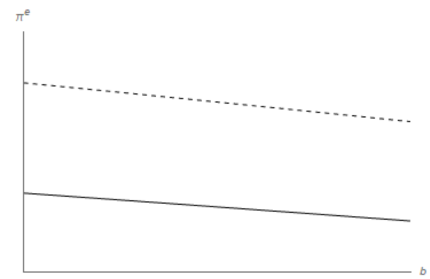

Figure 20: The impact of a change in b on π^e.

5.2.2 Utility and Profit

The above discussion implies that, since the base wage is unchanged and the bonus payment increases, the average payment to employed workers rises in response to a higher b. It follows that the reference wage of workers, w^R, becomes higher, as illustrated by Figure 17. The graph in Figure 18 then shows that the instantaneous utility of employed workers increases accordingly. An increase in unemployment benefits is therefore associated with higher utility for both employed and unemployed workers, due to the negative effect on incentives.[27] However, the rent of an unemployed worker (the term $v^e - v^u$) is decreasing, since v^u grows much faster. This is particularly true when workers have relative income concerns, due to the higher marginal utility below the reference point, so that the dashed line in Figure 19 is steeper.

Finally, we consider the implications of an increase in b for the profits of matched firms. Clearly, since the effort to be induced remains unchanged, while the bonus payment must increase, firms pay higher wages on average and therefore make lower instantaneous profits. This is illustrated for our example by Figure

[27]Wang and Yang (2015) also find that employed workers become better off following an increase in unemployment benefits, since outside options are increased and wages are determined by the outcome of bargaining.

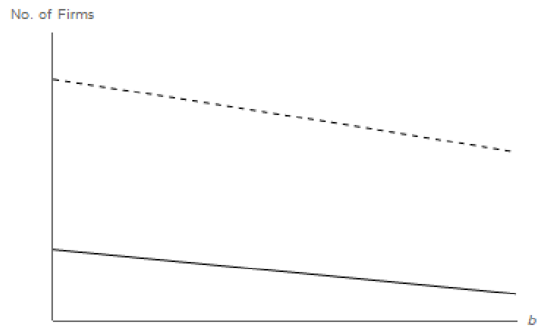

Figure 21: The impact of a change in b on the number of active firms in the economy.

20. It follows that even if higher unemployment benefits do not affect affect firms directly — for instance, due to difficulties in finding prospective workers or via the taxation required to fund such an increase — they may still suffer due to the negative effect on incentives.

5.2.3 Labour Market

We now turn our attention to the effects of variations in b on the labour market equilibrium. Recall that the change in the number of firms will be determined by the right-hand side of (37). An increase in unemployment benefits has two effects. First, as discussed in the foregoing, the negative impact of a higher b on incentives means that firms must pay a higher bonus and as such receive lower profits, which implies that discounted expected profits, Π^e, also decrease. Second, since workers are let go with a lower probability y_1 following the change, as shown by Figure 15, the flow of workers into unemployment is decreased. This reduces an unmatched firm's chances of finding an unemployed worker. These two effects work in the same direction to reduce the number of firms active in the market, as illustrated by Figure 21. Intuitively, less firms wish to expend money searching for a worker, since, *ceteris paribus*, i) the probability of finding one is lower and ii) expected profits in the case that they do find one are reduced.

From (37), it is clear that $q(\theta)$ must increase if Π^e falls. Figure 22 then illustrates this. Intuitively, the number of firms active in the labour market are reduced so that those who remain face a better chance of finding a worker. In contrast, the probability of an unemployed worker finding a match is reduced following an increase in b; this is shown by Figure 23. Not only is the rate at which matches are dissolved lower due to the decrease in y_1, but there are now also less firms actively searching for a worker.

Figures 24 and 25 illustrate the respective implications of an increase in b for the economy's unemployment rate and vacancy rate. In particular, there are

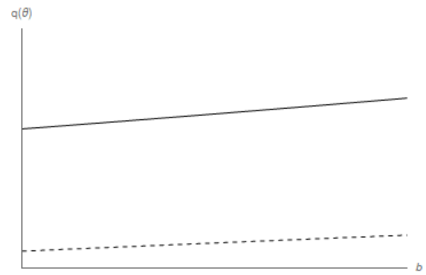

Figure 22: The impact of a change in b on $q(\theta)$.

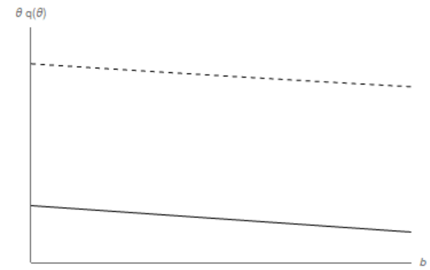

Figure 23: The impact of a change in b on $\theta q(\theta)$.

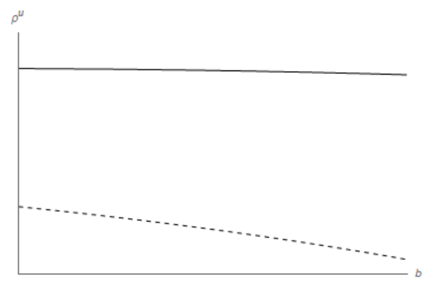

Figure 24: The impact of a change in b on ρ^u.

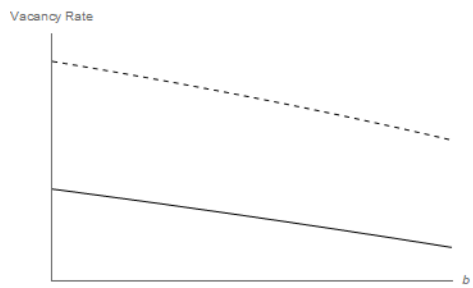

Figure 25: The impact of a change in b on the economy's vacancy rate.

two effects on the unemployment rate. First, workers are dismissed with a lower probability which means that the rate at which matches are dissolved is reduced; this leads to a lower ρ^u. Second, as discussed in the foregoing, the reduction in the amount of firms active in the economy lowers an unemployed worker's chances of finding a match, decreasing the flow out of unemployment and leading to a higher ρ^u. Figure 24 shows that in both cases of our example, the former effect is dominant so that unemployment is reduced. This is especially true when workers have relative income concerns, since the former effect is stronger due to the increased sensitivity of y_1 to variations in b. Finally, Figure 25 plots the economy's vacancy rate, which is falling in b. Not only are there less dismissals, so that existing matches last longer, an increase in b also leads to a reduction of unmatched firms in the market. Both of these effects lead to a lower vacancy rate.

5.2.4 The Gini Coefficient

Following the same approach as the previous subsection, we can calculate the Gini Coefficient associated with a particular value of b. Figure 26 plots this graph. In both the $\alpha = 0$ and $\alpha = 1$ cases, the Gini Coefficient is decreasing as b grows. The effects which lead to this change are as follows. First, there are less unemployed

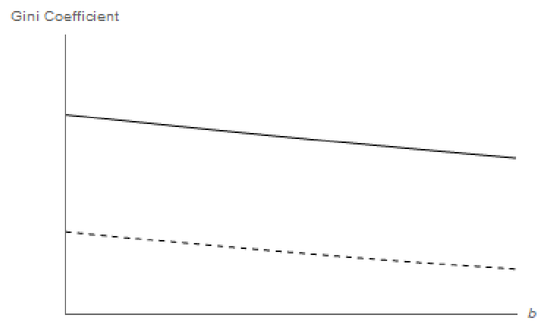

Figure 26: The impact of a change in b on the economy's Gini Coefficient.

workers due to the reduction in ρ^u, while those workers who are unemployed receive an increased income of b. Both of these effects reduce inequality. On the other hand, among employed workers, the increase in the bonus payment has the effect of increasing inequality. Figure 26 shows that for our example, the former effects are dominant so that overall inequality among workers in the economy is reduced.

6 Conclusion

In this chapter, we study incentive contracting when workers have relative income concerns and firms can use the threat of dismissal to create effort incentives. In order to formalise this idea, we extend the contracting environment outlined in the previous chapter to allow for a frictional labour market with unemployment. In this context, we use a series of numerical examples to analyse the impact of changes in the economy's minimum wage and unemployment benefit payments. We find that the minimum wage acts as an efficiency wage by raising the value of employment relative to unemployment. This induces workers to exert higher effort in an attempt to avoid the loss of income associated with dismissal, allowing firms to lower explicit incentive pay. It follows that since dismissal becomes more effective, it is used more often and firms terminate the employment of workers with a higher probability. However, expected wage payments rise following an increase, so that employed workers are better off on average and firms' expected profits are reduced. In contrast, an increase in unemployment benefits reduces the relative value of employment, so that the bonus payment must be increased if incentives are to stay constant. Dismissal becomes less efficient and is therefore used less often. The average wage payment to workers then increases, reducing the profits of firms. We show that while these effects exist when workers are self-interested, they are particularly strong in the presence of relative income concerns.

While we primarily consider the implications for contracting between firm-worker pairs, we also investigate how these changes affect labour market outcomes. Studies which analyse the impact of changes in the minimum wage in the context of search-and-matching models typically find two opposing effects on employment (Meer and West, 2016). First, by increasing a firm's costs of hiring a worker, an increase in the minimum wage reduces demand for labour and pushes some firms out of the market. Second, the higher minimum wage increases the expected returns to employment, which induces additional search effort from workers. This increases both the quantity of workers and the intensity of their search, leading to a positive effect on employment.[28] Our analysis shows that, even in the absence of search effects, this increase in returns to employed workers can impact the rate of unemployment via the efficiency wage effect, which increases the probability of dismissal and therefore raises the flow of workers into unemployment.

In the standard textbook model of matching, unemployment benefits dictate workers' outside options. Since wages are determined by Nash bargaining between firms and workers, it follows that a higher unemployment benefit increases equilibrium wages, reducing the profitability of firms and leading to a reduction in demand for labour. When workers must additionally undertake costly effort when searching for employment, many papers have also studied the potential discouragement effects of increases in unemployment benefits on this search effort and the resulting impact on unemployment (Boadway and Cuff, 2018; Landais et al., 2018). However, our analysis highlights the existence of a positive employment effect, since increases in the unemployment benefit reduce the efficiency of dismissal as a tool for creating incentives. Accordingly, firms dismiss workers less often, reducing the flow into unemployment.

To conclude the chapter, we briefly comment on our specification of workers' preferences. Throughout, we assumed that all workers, employed or unemployed, compared their incomes to the same reference wage, which was determined by the average wage payment to employed workers. The assumption of identical reference wages seems natural, since all workers in our model are ex ante homogeneous, while the assumption that the income of unemployed workers is not taken into account is consistent with evidence that social comparisons have a tendency to be upward-looking in nature (see e.g. Hecht, 2017). Moreover, this specification significantly simplifies the analysis, since it implies $b < W < w^R < W + B$ and therefore prevents situations whereby the reference wage becomes less than the economy's minimum wage. However, due to the lack of evidence for how social comparisons change when relevant others are in alternative 'states' (such as unemployment or retirement), it is difficult to ascertain the accuracy of this assumption.

[28] See Van den Berg and Ridder (1998), Acemoglu (2001) and Flinn (2006).

In addition, for the sake of tractability we assumed that each worker compared their income in a given time period with the average income in the economy in that same time period. Similarly, evidence for the exact nature of social comparisons in dynamic environments is scarce, so that alternative specifications are possible. For instance, one could allow for workers who accumulate wealth over time, which then acts as a point of comparison.

There are several other directions in which the model could be extended. It would be straightforward to allow for firms and workers to bargain over the terms of the contract; one could then consider the impact of variations in bargaining power on the outcomes of contracting. Alternatively, one could incorporate search effort into the model and consider how changes in the minimum wage or unemployment benefits affect search incentives when workers have relative income concerns. Finally, as mentioned in the text, one could allow for unemployment benefits to be funded endogenously by taxation; this would be particularly interesting, since taxes would also influence the ability of firms to provide incentives to workers. These developments, along with others, are left for future research.

Conclusion

This thesis examines the optimal provision of incentives within firms when workers' preferences are characterised by relative income concerns. The aim has been to provide some new insights which aid our understanding of the design of incentive contracts and their impact on individual behaviour.

We began in Chapter 1 by surveying the existing literatures which consider incentive provision with two related models of preferences: inequity aversion and loss aversion. We found that, in either case, introduction of such preferences typically resulted in two distinct effects due to the reduction in utility associated with low wage outcomes. First, firms benefit from an increased ability to create incentives, since a given incentive scheme now induces higher effort. Second, firms must pay higher wages on average to provide a worker with the same level of expected utility. The overall implications for a firm's costs of these incentive and participation effects were shown to depend on the specifics of the economic environment.

We also studied the implications of these effects for the structure of the optimal contract. With inequity aversion, we discussed how firms could benefit from designing wage schemes which are dependent on the performance of others within the reference group. For instance, depending on the economic environment, it may be beneficial to eliminate wage inequalities using team contracts, or exacerbate them through the use of relative performance evaluation or tournament schemes. When workers are loss averse, however, we saw that studies have mostly focused on the optimal wage scheme when firms have access to a rich performance measure and found that payments are typically insensitive to this signal over some regions, since firms wish to shield workers from losses. Several authors have discussed how the aforementioned findings are consistent with empirical observations, such as the fact that real world incentive schemes tend to be simple, less sensitive to performance than predicted by the standard theory and result in wage compression within organisations (see Herweg et al., 2010; Jensen and Murphy, 1990 and Prendergast, 1999, respectively).

Next, in Chapter 2, we began our investigation of incentive contracting when workers have relative income concerns. Initially considering the problem of a single firm-worker pair who take the reference wage as given, we found that the optimal wage scheme took a simple form and featured either two or three distinct payments. In the latter case, the contract featured a wage payment equal to the worker's reference wage, similar to findings in the context of loss aversion (de Meza and Webb, 2007). We also showed that, similar to the foregoing cases, relative income concerns induce a positive incentive effect, which allows firms to reduce the bonus payment associated with high performance. In our framework, since workers extract a positive rent, this implies that firms can induce effort at a lower cost. However, we also discussed a key difference to the cases of inequity aversion and loss aversion with respect to the participation effect, which we argue is positive when workers undertake wider social comparisons with others outside of the firm.

In the latter part of the chapter, we assumed that the reference wage was determined endogenously by the economy's equilibrium average income. In this framework, we showed that the interdependence between the contracting of firm-worker pairs (via the reference wage) resulted in an externality effect, so that firms could benefit from collective decision making. Clearly, such effects do not arise in the cases of inequity aversion and loss aversion since comparisons, social or otherwise, do not extend beyond the firm. Finally, we were also able to explore how changes in the underlying parameters of the model affected the wider economy, by analysing changes in average income and the Gini coefficient.

In Chapter 3, we extended the model in order to study the role of dismissal when workers care about their relative income. For this purpose, we embedded the aforementioned contracting framework into a dynamic environment featuring a frictional labour market and unemployment. We showed that dismissal acts as a substitute for explicit incentive pay, allowing firms to reduce the average wage paid to workers in each time period. However, due to the frictional nature of the labour market, in the event that a worker's employment is terminated, firms may spend several periods searching for a new match resulting in a loss of profits. Our analysis shows that firms trade-off these effects when deciding the frequency with which to dismiss workers.

Since the size of any incentives created by the possibility of dismissal necessarily depend on the underlying conditions of the labour market, it is also interesting to consider the implications of changes in policies pertaining to employment. To this end, we studied the effects of changes in the minimum wage and unemployment benefits. We found that a higher minimum wage aids the creation of incentives via an efficiency wage effect, which allows for a reduction in the bonus payment. In contrast, unemployment benefits have a negative effect on incentives so that the bonus payment must be increased. Our framework also allows for an analysis of how these changes impact the number of firms active in the labour market and the steady-state equilibrium rate of unemployment.

In addition to the above, we showed that the effects of such changes are strengthened when workers have relative income concerns in comparison to the self-interested case. One of the reasons for this result is that policies which affect the average income of workers in the economy have a direct impact on utility, via the reference wage. This is in stark contrast to the cases of inequity aversion or loss aversion, where any comparisons are intra-firm in nature and are therefore unaffected by broader changes in economy-wide outcomes.

When reflecting on the analysis contained in this thesis, a key issue is our formalisation of relative income concerns. Throughout, we assumed that workers' preferences could be represented by a piecewise linear function around a reference wage which is determined endogenously by the average wage in the economy. This approach is simple, captures the notion of relative income comparisons in an intuitive way and allows for a tractable analysis of the implications of such preferences for incentive contracting. Moreover, as discussed in Chapter 2 of this thesis, there is some evidence that the average wage in a society does indeed represent an important point of comparison.

The assumption that social comparisons extend to all individuals in an economy seems particularly appropriate in our framework, since all workers were assumed to be identical. However, as discussed in Chapter 1, many authors have argued that in environments where salient others are disparate, individuals will typically focus on those who they perceive as being most similar to themselves. It follows that, in the presence of heterogeneity, workers' social comparisons may be limited to a specific group within society.[1] More generally, one could think of individuals as undertaking comparisons with several different groups, with varying weights relating to the degree of social proximity. A related issue concerns the extent to which individuals might compare themselves with others who are in alternative 'states'; for instance, to what degree does an employed worker undertake comparisons with those who are involuntarily unemployed, retired or engaging in home production?

There are similar complexities in dynamic environments. In Chapter 3, we assumed that individuals were myopic in the sense that social comparisons were limited to the current time period. In reality, the true nature of intertemporal

[1] Hecht (2017) conducts a series of interviews with 30 UK-based individuals whose annual incomes are in the top 1% of the distribution, many of whom are also within the top 1% of the wealth distribution. She finds that several feel relatively disadvantaged: while they are aware of their own advantaged economic position within the general population, they undertake wide-ranging social comparisons with others including entrepreneurs, philanthropists, billionaires and sports stars. Accordingly, they often find themselves 'looking up' and as such do not always perceive themselves as earning a 'high income'.

relative income concerns is likely to be much more elaborate than this. Workers may care about long-term inequalities, or about accumulated wealth rather than income. Clark et al. (2008) discuss evidence that individuals may even undertake comparisons with their past selves and dislike falling behind previous levels of income; this notion shares several similarities with reference-dependence and loss aversion.

Even if one is able to pin down the correct nature of social comparisons in any given environment, a question still remains regarding which utility function best represents these preferences. Clearly, our assumption of piecewise linearity was a simplification, allowing for a parsimonious analysis of incentive contracting. However, the relationship between relative income concerns and risk preferences is an understudied issue, so that further development of our model of preferences is not straightforward. Ultimately, the question of which specification of preferences is most appropriate in any given environment is empirical. Hopefully, future studies will shed some light on this matter.

To conclude the thesis we briefly reflect on some promising future directions for the literature. With respect to the foregoing discussion, increased empirical evidence for the exact nature of social comparisons will help economists develop tractable models of preferences which can capture important aspects of individual behaviour; these models can then be applied to the study of incentive contracting in complex environments, improving the predictions of the theory. In addition, Chapters 2 and 3 of this work discussed several possible extensions to our analysis. These include heterogeneous workers, the introduction of bargaining power and considering contracting in more complex labour market environments. In particular, future studies should aim to investigate the implications of relative income concerns for economically interesting variables relating to policy, since a tendency of the existing literature has been to limit analysis to the parameters of the behavioural theory; this is consistent with the observation of Kőszegi (2014) regarding the wider behavioural contract theory literature. Chapter 3 of this thesis provided one example of a development along these lines, but there are several further possibilities yet to be explored. Finally, it will be interesting to see if the predictions of the existing literature hold up to empirical or experimental scrutiny. Such studies represent an important agenda for future research and will be invaluable for the development of further models which investigate incentive contracting with non-standard preferences.

Bibliography

Abdellaoui, M., Bleichrodt, H. and Paraschiv, C. (2007), 'Loss aversion under prospect theory: A parameter-free measurement', *Management Science* **53**(10), 1659–1674.

Abeler, J., Altmann, S., Kube, S. and Wibral, M. (2010), 'Gift exchange and workers' fairness concerns: When equality is unfair', *Journal of the European Economic Association* **8**(6), 1299–1324.

Abeler, J., Falk, A., Goette, L. and Huffman, D. (2011), 'Reference points and effort provision', *American Economic Review* **101**(2), 470–492.

Acemoglu, D. (2001), 'Good jobs versus bad jobs', *Journal of Labor Economics* **19**(1), 1–21.

Acemoglu, D. and Shimer, R. (1999), 'Efficient unemployment insurance', *Journal of Political Economy* **107**(5), 893–928.

Adams, J. S. (1963), 'Towards an understanding of inequity', *Journal of Abnormal and Social Psychology* **67**(5), 422–436.

Adams, J. S. (1965), Inequity in social exchange, *in* 'Advances in experimental social psychology', Vol. 2, Elsevier, pp. 267–299.

Agell, J. and Lundborg, P. (1995), 'Theories of pay and unemployment: Survey evidence from Swedish manufacturing firms', *Scandinavian Journal of Economics* **97**(2), 295–307.

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

Alesina, A., Di Tella, R. and MacCulloch, R. (2004), 'Inequality and happiness: Are Europeans and Americans different?', *Journal of Public Economics* **88**(9-10), 2009–2042.

Allais, M. (1953), 'Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole Américaine', *Econometrica* **21**(4), 503–546.

Alpizar, F., Carlsson, F. and Johansson-Stenman, O. (2005), 'How much do we care about absolute versus relative income and consumption?', *Journal of Economic Behavior & Organization* **56**(3), 405–421.

Arkes, H. R., Hirshleifer, D., Jiang, D. and Lim, S. (2008), 'Reference point adaptation: Tests in the domain of security trading', *Organizational Behavior and Human Decision Processes* **105**(1), 67–81.

Arkes, H. R., Hirshleifer, D., Jiang, D. and Lim, S. S. (2010), 'A cross-cultural study of reference point adaptation: Evidence from China, Korea, and the US', *Organizational Behavior and Human Decision Processes* **112**(2), 99–111.

Armantier, O. and Boly, A. (2015), 'Framing of incentives and effort provision', *International Economic Review* **56**(3), 917–938.

Babcock, L., Wang, X. and Loewenstein, G. (1996), 'Choosing the wrong pond: Social comparisons in negotiations that reflect a self-serving bias', *The Quarterly Journal of Economics* **111**(1), 1–19.

Balmaceda, F. (2018), 'Optimal task assignments with loss-averse agents', *European Economic Review* **105**, 1–26.

Banerjee, S. and Sarkar, M. (2017), 'Other-regarding principal and moral hazard: The single-agent case', *International Journal of Economic Theory* **13**(3), 313–326.

Banerji, A. and Gupta, N. (2014), 'Detection, identification, and estimation of loss aversion: Evidence from an auction experiment', *American Economic Journal: Microeconomics* **6**(1), 91–133.

Barberis, N. C. (2013), 'Thirty years of prospect theory in economics: A review and assessment', *Journal of Economic Perspectives* **27**(1), 173–195.

Bartling, B. (2011), 'Relative performance or team evaluation? Optimal contracts for other-regarding agents', *Journal of Economic Behavior & Organization* **79**(3), 183–193.

Bartling, B. (2012*a*), 'Endogenous social comparisons and the internal organization of firms', *Managerial and Decision Economics* **33**(1), 11–17.

Bartling, B. (2012*b*), 'Multi-tasking and inequity aversion in the linear–exponential–normal moral hazard model', *Economics Letters* **116**(3), 523–525.

Bartling, B. and von Siemens, F. A. (2010*a*), 'Equal sharing rules in partnerships', *Journal of Institutional and Theoretical Economics* **166**(2), 299–320.

Bartling, B. and von Siemens, F. A. (2010*b*), 'The intensity of incentives in firms and markets: Moral hazard with envious agents', *Labour Economics* **17**(3), 598–607.

Baucells, M., Weber, M. and Welfens, F. (2011), 'Reference-point formation and updating', *Management Science* **57**(3), 506–519.

Bell, D. E. (1985), 'Disappointment in decision making under uncertainty', *Operations Research* **33**(1), 1–27.

Bental, B., Deffains, B. and Demougin, D. (2016), Purposive approach and non-comprehensive incentive contracts. Working paper.

Bergerhoff, J. and Vosen, A. (2015), Can being behind get you ahead? Reference dependence and asymmetric equilibria in an unfair tournament. Working paper.

Bewley, T. (2004), Fairness, reciprocity, and wage rigidity. IZA Discussion paper series, No. 1137.

Bewley, T. F. (1998), 'Why not cut pay?', *European Economic Review* **42**(3-5), 459–490.

Boadway, R. and Cuff, K. (2018), 'Optimal unemployment insurance and redistribution', *Journal of Public Economic Theory* **20**(3), 303–324.

Bolton, G. E. and Ockenfels, A. (2000), 'ERC: A theory of equity, reciprocity and competition', *American Economic Review* **90**(1), 166–193.

Bolton, P. and Dewatripont, M. (2005), *Contract Theory*, MIT Press.

Booij, A. S. and Van de Kuilen, G. (2009), 'A parameter-free analysis of the utility of money for the general population under prospect theory', *Journal of Economic Psychology* **30**(4), 651–666.

Booij, A. S., Van Praag, B. M. and Van De Kuilen, G. (2010), 'A parametric analysis of prospect theory's functionals for the general population', *Theory and Decision* **68**(1-2), 115–148.

Bowman, D., Minehart, D. and Rabin, M. (1999), 'Loss aversion in a consumption–savings model', *Journal of Economic Behavior & Organization* **38**(2), 155–178.

Brink, A. G. and Rankin, F. W. (2013), 'The effects of risk preference and loss aversion on individual behavior under bonus, penalty, and combined contract frames', *Behavioral Research in Accounting* **25**(2), 145–170.

Brosnan, S. F. and De Waal, F. B. (2003), 'Monkeys reject unequal pay', *Nature* **425**(6955), 297–299.

Buckingham, J. T. and Alicke, M. D. (2002), 'The influence of individual versus aggregate social comparison and the presence of others on self-evaluations', *Journal of Personality and Social Psychology* **83**(5), 1117–1130.

Camerer, C., Babcock, L., Loewenstein, G. and Thaler, R. (1997), 'Labor supply of New York City cabdrivers: One day at a time', *The Quarterly Journal of Economics* **112**(2), 407–441.

Camerer, C. F. (2000), Prospect theory in the wild: Evidence from the field, *in* D. Kahneman and A. Tversky, eds, 'Choices, Values and Frames', Cambridge University Press and Russell Sage Foundation, pp. 148–161.

Cappelli, P. and Sherer, P. D. (1988), 'Satisfaction, market wages, & labor relations: An airline study', *Industrial Relations: A Journal of Economy and Society* **27**(1), 56–73.

Card, D. and Dahl, G. B. (2011), 'Family violence and football: The effect of unexpected emotional cues on violent behavior', *The Quarterly Journal of Economics* **126**(1), 103–143.

Card, D., Mas, A., Moretti, E. and Saez, E. (2012), 'Inequality at work: The effect of peer salaries on job satisfaction', *American Economic Review* **102**(6), 2981–3003.

Casciaro, T. and Lobo, M. S. (2008), 'When competence is irrelevant: The role of interpersonal affect in task-related ties', *Administrative Science Quarterly* **53**(4), 655–684.

Cato, S. and Ebina, T. (2014), Inequality aversion in long-term contracts. Working paper.

Chen, M. K., Lakshminarayanan, V. and Santos, L. R. (2006), 'How basic are behavioral biases? Evidence from capuchin monkey trading behavior', *Journal of Political Economy* **114**(3), 517–537.

Chen, M. K. and Santos, L. R. (2006), 'Some thoughts on the adaptive function of inequity aversion: An alternative to Brosnan's social hypothesis', *Social Justice Research* **19**(2), 201–207.

Clark, A. E., Frijters, P. and Shields, M. A. (2008), 'Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles', *Journal of Economic Literature* **46**(1), 95–144.

Clark, A. and Oswald, A. (1996), 'Satisfaction and comparison income', *Journal of Public Economics* **61**(3), 359–381.

Copeland, P. V. and Cuccia, A. D. (2002), 'Multiple determinants of framing referents in tax reporting and compliance', *Organizational Behavior and Human Decision Processes* **88**(1), 499–526.

Corgnet, B., Gómez-Miñambres, J. and Hernán-González, R. (2018), 'Goal setting in the principal-agent model: Weak incentives for strong performance', *Games and Economic Behavior* **109**, 311–326.

Crawford, V. P. and Meng, J. (2011), 'New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income', *American Economic Review* **101**(5), 1912–32.

Daido, K. and Itoh, H. (2007), The Pygmalion and Galatea effects: An agency model with reference-dependent preferences and applications to self-fulfilling prophecy. Working paper.

Daido, K., Morita, K., Murooka, T. and Ogawa, H. (2013), 'Task assignment under agent loss aversion', *Economics Letters* **121**(1), 35–38.

Daido, K. and Murooka, T. (2016), 'Team incentives and reference-dependent preferences', *Journal of Economics & Management Strategy* **25**(4), 958–989.

Dato, S., Grunewald, A. and Müller, D. (Forthcoming), Expectation-based loss aversion and rank-order tournaments. Forthcoming in Economic Theory.

de Meza, D. and Webb, D. C. (2007), 'Incentive design under loss aversion', *Journal of the European Economic Association* **5**(1), 66–92.

DellaVigna, S. (2009), 'Psychology and economics: Evidence from the field', *Journal of Economic Literature* **47**(2), 315–372.

DellaVigna, S., Lindner, A., Reizer, B. and Schmieder, J. F. (2017), 'Reference-dependent job search: Evidence from Hungary', *The Quarterly Journal of Economics* **132**(4), 1969–2018.

DeMarzo, P. and Kaniel, R. (2017), Relative pay for non-relative performance: Keeping up with the Joneses with optimal contracts. Working paper.

Demougin, D. and Fluet, C. (1998), 'Mechanism sufficient statistic in the risk-neutral agency problem', *Journal of Institutional and Theoretical Economics* **154**(4), 622–639.

Demougin, D. and Fluet, C. (2001), 'Monitoring versus incentives', *European Economic Review* **45**(9), 1741–1764.

Demougin, D. and Fluet, C. (2003), Inequity aversion in tournaments. Working Paper 03-22, CIRPEE.

Demougin, D. and Fluet, C. (2006), Group vs. individual performance pay when workers are envious, *in* D. Demougin and C. Schade, eds, 'Contributions to Entrepreneurship and Economics - First Haniel-Kreis Meeting on Entrepreneurial Research', Duncker and Humblot Verlag, pp. 39–47.

Demougin, D., Fluet, C. and Helm, C. (2006), 'Output and wages with inequality averse agents', *Canadian Journal of Economics* **39**(2), 399–413.

Demougin, D. and Helm, C. (2011), 'Job matching when employment contracts suffer from moral hazard', *European Economic Review* **55**(7), 964–979.

Demougin, D. and Upton, H. (2019), Moral hazard and keeping up with the Joneses. Working paper.

Dittmann, I., Maug, E. and Spalt, O. (2010), 'Sticks or carrots? Optimal CEO compensation when managers are loss averse', *Journal of Finance* **65**(6), 2015–2050.

Dubey, P., Geanakoplos, J. and Haimanko, O. (2013), 'Prizes versus wages with envy and pride', *Japanese Economic Review* **64**(1), 98–121.

Dupor, B. and Liu, W.-F. (2003), 'Jealousy and equilibrium overconsumption', *American Economic Review* **93**(1), 423–428.

Dur, R. and Glazer, A. (2007), 'Optimal contracts when a worker envies his boss', *Journal of Law, Economics, & Organization* **24**(1), 120–137.

Ederer, F. and Patacconi, A. (2010), 'Interpersonal comparison, status and ambition in organizations', *Journal of Economic Behavior & Organization* **75**(2), 348–363.

Elster, J. (1991), Envy in social life, *in* R. J. Zeckhauser, ed., 'Strategy and choice', The MIT Press, Cambridge, MA, and London, pp. 49–82.

Engelmann, D. and Strobel, M. (2004), 'Inequality aversion, efficiency, and maximin preferences in simple distribution experiments', *American Economic Review* **94**(4), 857–869.

Englmaier, F. (2005), Moral hazard, contracts and social preferences: A survey, *in* B. Agarwal and A. Vercelli, eds, 'Psychology, Rationality and Economic Behaviour: Challenging Standard Assumptions', Palgrave Macmillan UK, London, pp. 125–139.

Englmaier, F. and Wambach, A. (2010), 'Optimal incentive contracts under inequity aversion', *Games and Economic Behavior* **69**(2), 312–328.

Ericson, K. M. and Fuster, A. (2011), 'Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments', *The Quarterly Journal of Economics* **126**(4), 1879–1907.

Falk, A. and Knell, M. (2004), 'Choosing the Joneses: Endogenous goals and reference standards', *Scandinavian Journal of Economics* **106**(3), 417–435.

Fehr, E. and Goette, L. (2007), 'Do workers work more if wages are high? Evidence from a randomized field experiment', *American Economic Review* **97**(1), 298–317.

Fehr, E. and Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics* **114**(3), 817–868.

Fehr, E. and Schmidt, K. M. (2006), The economics of fairness, reciprocity and altruism–experimental evidence and new theories, *in* S.-C. Kolm and J. M. Ythier, eds, 'Handbook of the Economics of Giving, Altruism and Reciprocity', Vol. 1, Elsevier, pp. 615–691.

Fershtman, C. (2008), Economics and social status, *in* S. N. Durlauf and L. E. Blume, eds, 'The New Palgrave Dictionary of Economics: Volume 1 – 8', Palgrave Macmillan UK, London, pp. 6077–6080.

Fershtman, C., Gneezy, U. and List, J. A. (2012), 'Equity aversion: Social norms and the desire to be ahead', *American Economic Journal: Microeconomics* **4**(4), 131–44.

Fershtman, C., Hvide, H. K. and Weiss, Y. (2003), 'A behavioral explanation of the relative performance evaluation puzzle', *Annales d'economie et de statistique* **71/72**, 349–361.

Festinger, L. (1954), 'A theory of social comparison processes', *Human Relations* **7**(2), 117–140.

Fischbacher, U. and Gachter, S. (2010), 'Social preferences, beliefs, and the dynamics of free riding in public goods experiments', *American Economic Review* **100**(1), 541–56.

Flinn, C. J. (2006), 'Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates', *Econometrica* **74**(4), 1013–1062.

Frank, R. H. (1985), *Choosing the right pond: Human behavior and the quest for status*, Oxford University Press.

Frederick, S. and Loewenstein, G. (1999), Hedonic adaptation, *in* D. Kahneman, E. Diener and N. Schwarz, eds, 'Well-Being: The foundations of Hedonic Psychology', Russell Sage Foundation, New York, pp. 302–329.

Frohlich, N., Oppenheimer, J. and Kurki, A. (2004), 'Modeling other-regarding preferences and an experimental test', *Public Choice* **119**(1-2), 91–117.

Gamba, A., Manzoni, E. and Stanca, L. (2017), 'Social comparison and risk taking behavior', *Theory and Decision* **82**(2), 221–248.

Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989), 'Psychological games and sequential rationality', *Games and Economic Behavior* **1**(1), 60–79.

Gibbons, R. (1998), 'Incentives in organizations', *Journal of Economic Perspectives* **12**(4), 115–132.

Gill, D. and Prowse, V. (2012), 'A structural analysis of disappointment aversion in a real effort competition', *American Economic Review* **102**(1), 469–503.

Gill, D. and Stone, R. (2010), 'Fairness and desert in tournaments', *Games and Economic Behavior* **69**(2), 346–364.

Gill, D. and Stone, R. (2015), 'Desert and inequity aversion in teams', *Journal of Public Economics* **123**, 42–54.

Gneezy, U. (2005), Updating the reference level: Experimental evidence, *in* A. Rapoport and R. Zwick, eds, 'Experimental Business Research', Vol. 3, Springer, pp. 263–284.

Gneezy, U., Goette, L., Sprenger, C. and Zimmermann, F. (2017), 'The limits of expectations-based reference dependence', *Journal of the European Economic Association* **15**(4), 861–876.

Goel, A. M. and Thakor, A. V. (2006), Optimal contracts when agents envy each other. Working paper.

Goette, L., Cerulli-Harms, A. and Sprenger, C. (Forthcoming), Randomizing endowments: An experimental study of rational expectations and reference-dependent preferences. Forthcoming in the American Economic Journal: Microeconomics.

Goette, L., Graeber, T., Kellogg, A. and Sprenger, C. (2018), Heterogeneity of loss aversion and expectations-based reference points. Working paper.

Goukasian, L. and Wan, X. (2010), 'Optimal incentive contracts under relative income concerns', *Mathematics and Financial Economics* **4**(1), 57–86.

Graham, C. and Felton, A. (2006), 'Inequality and happiness: Insights from Latin America', *Journal of Economic Inequality* **4**(1), 107–122.

Green, J. R. and Stokey, N. L. (1983), 'A comparison of tournaments and contracts', *Journal of Political Economy* **91**(3), 349–364.

Greenberg, J. (1993), The social side of fairness: Interpersonal and informational classes of organizational justice, *in* R. Cropanzano, ed., 'Justice in the workplace: Approaching fairness in human resource management', Erlbaum, Hillsdale, NJ.

Grund, C. and Przemeck, J. (2012), 'Subjective performance appraisal and inequality aversion', *Applied Economics* **44**(17), 2149–2155.

Grund, C. and Sliwka, D. (2005), 'Envy and compassion in tournaments', *Journal of Economics and Management Strategy* **14**(1), 187–207.

Gul, F. (1991), 'A theory of disappointment aversion', *Econometrica* **59**(3), 667–86.

Heath, C., Larrick, R. P. and Wu, G. (1999), 'Goals as reference points', *Cognitive Psychology* **38**(1), 79–109.

Hecht, K. (2017), A relational analysis of top incomes and wealth: Economic evaluation, relative (dis) advantage and the service to capital. LSE Department of Sociology Working Paper 11.

Heffetz, O. (2018), Are reference points merely lagged beliefs over probabilities? NBER Working Paper 24721.

Heffetz, O. and List, J. A. (2014), 'Is the endowment effect an expectations effect?', *Journal of the European Economic Association* **12**(5), 1396–1422.

Heidhues, P. and Kőszegi, B. (2014), 'Regular prices and sales', *Theoretical Economics* **9**(1), 217–251.

Helliwell, J. F. and Huang, H. (2010), 'How's the job? Well-being and social capital in the workplace', *ILR Review* **63**(2), 205–227.

Herweg, F., Müller, D. and Weinschenk, P. (2010), 'Binary payment schemes: Moral hazard and loss aversion', *American Economic Review* **100**(5), 2451–2477.

Hilken, K., De Jaegher, K. and Jegers, M. (2013), 'Strategic framing in contracts', *Discussion Paper Series, Tjalling C. Koopmans Research Institute* **13**(04).

Holmström, B. (1979), 'Moral hazard and observability', *The Bell Journal of Economics* **10**(1), 74–91.

Holmström, B. (1982), 'Moral hazard in teams', *The Bell Journal of Economics* **13**(2), 324–340.

Holmström, B. and Milgrom, P. (1991), 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design', *Journal of Law, Economics, & Organization* **7**, 24–52.

Hori, K. and Osano, H. (2014), Dynamic compensation contracts and capital structure under loss aversion. Working paper.

Huck, S. and Rey-Biel, P. (2006), 'Endogenous leadership in teams', *Journal of Institutional and Theoretical Economics* **162**(2), 253–261.

Iantchev, E. P. (2009), Risk or loss aversion? Evidence from personnel records. Working paper.

Innes, R. D. (1990), 'Limited liability and incentive contracting with ex-ante action choices', *Journal of Economic Theory* **52**(1), 45–67.

Itoh, H. (2004), 'Moral hazard and other-regarding preferences', *Japanese Economic Review* **55**(1), 18–45.

Jensen, M. C. and Murphy, K. J. (1990), 'Performance pay and top-management incentives', *Journal of Political Economy* **98**(2), 225–264.

Jofre, A., Moroni, S. and Repetto, A. (2015), Dynamic contracts under loss aversion. Working paper.

Johansson-Stenman, O., Carlsson, F. and Daruvala, D. (2002), 'Measuring future grandparents' preferences for equality and relative standing', *The Economic Journal* **112**(479), 362–383.

Just, D. R. and Wu, S. Y. (2005), Loss aversion and reference points in contracts. Working paper.

Kachelmeier, S. J., Shehata, M. et al. (1992), 'Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China', *American Economic Review* **82**(5), 1120–1141.

Kahneman, D. (2003), 'A perspective on judgment and choice: Mapping bounded rationality', *American Psychologist* **58**(9), 697–720.

Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1991), 'Anomalies: The endowment effect, loss aversion, and status quo bias', *Journal of Economic Perspectives* **5**(1), 193–206.

Kahneman, D. and Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**(2), 263–292.

Kanbur, R., Pirttilä, J. and Tuomala, M. (2008), 'Moral hazard, income taxation and prospect theory', *Scandinavian Journal of Economics* **110**(2), 321–337.

Kandel, E. and Lazear, E. P. (1992), 'Peer pressure and partnerships', *Journal of Political Economy* **100**(4), 801–817.

Karle, H., Kirchsteiger, G. and Peitz, M. (2015), 'Loss aversion and consumption choice: Theory and experimental evidence', *American Economic Journal: Microeconomics* **7**(2), 101–20.

Katz, L. F. and Krueger, A. B. (1992), 'The effect of the minimum wage on the fast-food industry', *ILR Review* **46**(1), 6–21.

Keefer, B. (2016), A work vs. perk model of leadership and organizational culture. Working paper.

Kim, S. K. (1995), 'Efficiency of an information system in an agency model', *Econometrica* **63**(1), 89–102.

Kim, S. K. (1997), 'Limited liability and bonus contracts', *Journal of Economics & Management Strategy* **6**(4), 899–913.

Knetsch, J. L. and Wong, W.-K. (2009), 'The endowment effect and the reference state: Evidence and manipulations', *Journal of Economic Behavior & Organization* **71**(2), 407–413.

Knight, J., Lina, S. and Gunatilaka, R. (2009), 'Subjective well-being and its determinants in rural China', *China Economic Review* **20**(4), 635–649.

Kole, S. and Lehn, K. M. (2000), Workforce integration and the dissipation of value in mergers, the case of USAir's acquisition of Piedmont Aviation, *in* 'Mergers and productivity', University of Chicago Press, pp. 239–286.

Kölle, F., Sliwka, D. and Zhou, N. (2016), 'Heterogeneity, inequity aversion, and group performance', *Social Choice and Welfare* **46**(2), 263–286.

Konow, J. (2003), 'Which is the fairest one of all? A positive analysis of justice theories', *Journal of Economic Literature* **41**(4), 1188–1239.

Koop, G. J. and Johnson, J. G. (2012), 'The use of multiple reference points in risky decision making', *Journal of Behavioral Decision Making* **25**(1), 49–62.

Kőszegi, B. (2014), 'Behavioral contract theory', *Journal of Economic Literature* **52**(4), 1075–1118.

Kőszegi, B. and Rabin, M. (2006), 'A model of reference-dependent preferences', *The Quarterly Journal of Economics* **121**(4), 1133–1165.

Kőszegi, B. and Rabin, M. (2007), 'Reference-dependent risk attitudes', *American Economic Review* **97**(4), 1047–1073.

Kőszegi, B. and Rabin, M. (2009), 'Reference-dependent consumption plans', *American Economic Review* **99**(3), 909–936.

Kragl, J. (2015), 'Group versus individual performance pay in relational employment contracts when workers are envious', *Journal of Economics and Management Strategy* **24**(1), 131–150.

Kragl, J. (2016), 'Relational bonus contracts versus rank-order tournaments with envious workers', *Journal of Institutional and Theoretical Economics* **172**(3), 417–453.

Kragl, J. and Schmid, J. (2009), 'The impact of envy on relational employment contracts', *Journal of Economic Behavior and Organization* **72**(2), 766–779.

Kräkel, M. (2000), 'Relative deprivation in rank-order tournaments', *Labour Economics* **7**(4), 385–407.

Krapp, M. and Sandner, K. (2016), 'Impact of an equal pay norm on the optimal design of incentive contracts', *Journal of Business Economics* **86**(3), 301–338.

Kuhn, P., Kooreman, P., Soetevent, A. and Kapteyn, A. (2011), 'The effects of lottery prizes on winners and their neighbors: Evidence from the Dutch postcode lottery', *American Economic Review* **101**(5), 2226–2247.

Kuhn, P. and Villeval, M. C. (2015), 'Are women more attracted to co-operation than men?', *The Economic Journal* **125**(582), 115–140.

Kulik, C. T. and Ambrose, M. L. (1992), 'Personal and situational determinants of referent choice', *Academy of Management Review* **17**(2), 212–237.

Landais, C., Michaillat, P. and Saez, E. (2018), 'A macroeconomic approach to optimal unemployment insurance: Theory', *American Economic Journal: Economic Policy* **10**(2), 152–81.

Larsen, J. T., Peter McGraw, A., Mellers, B. A. and Cacioppo, J. T. (2004), 'The agony of victory and thrill of defeat: Mixed emotional reactions to disappointing wins and relieving losses', *Psychological Science* **15**(5), 325–330.

Lazear, E. P. and Rosen, S. (1981), 'Rank-order tournaments as optimum labor contracts', *Journal of Political Economy* **89**(5), 841–864.

Levine, D. K. (1998), 'Modeling altruism and spitefulness in experiments', *Review of Economic Dynamics* **1**(3), 593–622.

Li, J. (2009), 'Team production with inequity-averse agents', *Portuguese Economic Journal* **8**(2), 119–136.

LiCalzi, M. and Spaeter, S. (2003), 'Distributions for the first-order approach to principal-agent problems', *Economic Theory* **21**(1), 167–173.

Lien, J. W. and Zheng, J. (2015), 'Deciding when to quit: Reference-dependence over slot machine outcomes', *American Economic Review* **105**(5), 366–70.

Linde, J. and Sonnemans, J. (2012), 'Social comparison and risky choices', *Journal of Risk and Uncertainty* **44**(1), 45–72.

Loewenstein, G. F., Thompson, L. and Bazerman, M. H. (1989), 'Social utility and decision making in interpersonal contexts', *Journal of Personality and Social Psychology* **57**(3), 426–441.

Loewenstein, G., O'Donoghue, T. and Rabin, M. (2003), 'Projection bias in predicting future utility', *The Quarterly Journal of Economics* **118**(4), 1209–1248.

Loomes, G. and Sugden, R. (1986), 'Disappointment and dynamic consistency in choice under uncertainty', *The Review of Economic Studies* **53**(2), 271–282.

Lord, R. G. and Hohenfeld, J. A. (1979), 'Longitudinal field assessment of equity effects on the performance of Major League Baseball players', *Journal of Applied Psychology* **64**(1), 19–26.

Luft, J. (1994), 'Bonus and penalty incentives contract choice by employees', *Journal of Accounting and Economics* **18**(2), 181–206.

Luttmer, E. F. (2005), 'Neighbors as negatives: Relative earnings and well-being', *The Quarterly Journal of Economics* **120**(3), 963–1002.

Macera, R. (2018*a*), 'Intertemporal incentives under loss aversion', *Journal of Economic Theory* **178**, 551–594.

Macera, R. (2018*b*), 'Present or future incentives? On the optimality of fixed wages with moral hazard', *Journal of Economic Behavior & Organization* **147**, 129–144.

Macho-Stadler, I. and Pérez-Castrillo, D. (2018), Moral hazard: Base models and two extensions, *in* L. C. Corchón and M. A. Marini, eds, 'Handbook of Game Theory and Industrial Organization', Vol. 1, Edward Elgar Publishing.

Mahoney, T. A. and Weitzel, W. (1978), 'Secrecy and managerial compensation', *Industrial Relations: A Journal of Economy and Society* **17**(2), 245–251.

Marchegiani, L., Reggiani, T. and Rizzolli, M. (2016), 'Loss averse agents and lenient supervisors in performance appraisal', *Journal of Economic Behavior & Organization* **131**, 183–197.

März, O. (2016), Framing, expectations and reference points. ECARES Working Paper 2016-40.

Mas, A. (2006), 'Pay, reference points, and police performance', *The Quarterly Journal of Economics* **121**(3), 783–821.

Mas, A. (2017), 'Does transparency lead to pay compression?', *Journal of Political Economy* **125**(5), 1683–1721.

Masclet, D. (2002), Peer pressure in work teams: The effects of inequity aversion. Working paper.

Meer, J. and West, J. (2016), 'Effects of the minimum wage on employment dynamics', *Journal of Human Resources* **51**(2), 500–522.

Mellers, B. A., Schwartz, A., Ho, K. and Ritov, I. (1997), 'Decision affect theory: Emotional reactions to the outcomes of risky options', *Psychological Science* **8**(6), 423–429.

Miglietta, S. (2008), Incentive contracts and status-concerned agents. Working paper.

Mitman, K. and Rabinovich, S. (2015), 'Optimal unemployment insurance in an equilibrium business-cycle model', *Journal of Monetary Economics* **71**, 99–118.

Moen, E. R. and Rosén, Å. (2011), 'Incentives in competitive search equilibrium', *The Review of Economic Studies* **78**(2), 733–761.

Moen, E. R. and Rosén, Å. (2013), 'On-the-job search and moral hazard', *Journal of the European Economic Association* **11**(6), 1404–1431.

Mohnen, A., Pokorny, K. and Sliwka, D. (2008), 'Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence', *Journal of Labor Economics* **26**(4), 693–720.

Mookherjee, D. (1984), 'Optimal incentive schemes with many agents', *The Review of Economic Studies* **51**(3), 433–446.

Müller, S. and Rau, H. (2017), Decisions under uncertainty in social contexts. Working paper.

Neilson, W. S. (2006), 'Axiomatic reference-dependence in behavior toward others and toward risk', *Economic Theory* **28**(3), 681–692.

Neilson, W. S. and Stowe, J. (2010), 'Piece-rate contracts for other-regarding workers', *Economic Inquiry* **48**(3), 575–586.

Nesterov, Y. (2004), *Introductory Lectures on Convex Optimization*, Applied Optimization, 87 edn, Kluwer Academic Publishers, Boston, MA.

Neumark, D. and Wascher, W. L. (2007), 'Minimum wages and employment', *Foundations and Trends in Microeconomics* **3**(1–2), 1–182.

O'Donoghue, T. and Sprenger, C. (Forthcoming), Reference-dependent preferences, *in* D. Bernheim, S. DellaVigna and D. Laibson, eds, 'Handbook of Behavioral Economics', Elsevier.

Payne, J. W., Laughhunn, D. J. and Crum, R. (1980), 'Translation of gambles and aspiration level effects in risky choice behavior', *Management Science* **26**(10), 1039–1060.

Pesendorfer, W. (2006), 'Behavioral economics comes of age: A review essay on advances in behavioral economics', *Journal of Economic Literature* **44**(3), 712–721.

Petrongolo, B. and Pissarides, C. A. (2001), 'Looking into the black box: A survey of the matching function', *Journal of Economic Literature* **39**(2), 390–431.

Pierre, N. (2016), Loss averse compensation: Bonus vs. malus. Working paper.

Pierre, N. (2018), The effect of loss aversion on the multiperiod principal-agent problem. Working paper.

Pissarides, C. A. (2000), *Equilibrium Unemployment Theory*, MIT Press.

Pope, D. G. and Schweitzer, M. E. (2011), 'Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes', *American Economic Review* **101**(1), 129–57.

Post, T., Van den Assem, M. J., Baltussen, G. and Thaler, R. H. (2008), 'Deal or no deal? Decision making under risk in a large-payoff game show', *American Economic Review* **98**(1), 38–71.

Prendergast, C. (1999), 'The provision of incentives in firms', *Journal of Economic Literature* **37**(1), 7–63.

Rabin, M. (1998), 'Psychology and economics', *Journal of Economic Literature* **36**(1), 11–46.

Rabin, M. (2000), 'Risk aversion and expected-utility theory: A calibration theorem', *Econometrica* **68**(5), 1281–1292.

Rablen, M. D. (2010), 'Performance targets, effort and risk-taking', *Journal of Economic Psychology* **31**(4), 687–697.

Raihani, N. J. and McAuliffe, K. (2012), 'Does inequity aversion motivate punishment? Cleaner fish as a model system', *Social Justice Research* **25**(2), 213–231.

Rayo, L. and Becker, G. S. (2007), 'Evolutionary efficiency and happiness', *Journal of Political Economy* **115**(2), 302–337.

Rebitzer, J. B. and Taylor, L. J. (1995), 'The consequences of minimum wage laws: Some new theoretical ideas', *Journal of Public Economics* **56**(2), 245–255.

Rey-Biel, P. (2008), 'Inequity aversion and team incentives', *Scandinavian Journal of Economics* **110**(2), 297–320.

Rocheteau, G. (2001), 'Equilibrium unemployment and wage formation with matching frictions and worker moral hazard', *Labour Economics* **8**(1), 75–102.

Rogerson, W. P. (1985), 'The first-order approach to principal-agent problems', *Econometrica* **53**(6), 1357–67.

Samuelson, P. A. (1963), 'Risk and uncertainty: A fallacy of large numbers', *Scientia* **57**(98), 1–6.

Sannikov, Y. (2008), 'A continuous-time version of the principal-agent problem', *The Review of Economic Studies* **75**(3), 957–984.

Schmieder, J. F. and von Wachter, T. (2016), 'The effects of unemployment insurance benefits: New evidence and interpretation', *Annual Review of Economics* **8**, 547–581.

Schöttner, A. (2005), 'Precision in U-type and J-type tournaments', *Schmalenbach Business Review* **57**(2), 167–192.

Schwerter, F. (2016), Social reference points and risk taking. Working paper.

Senn, J. (2015), Myopic loss aversion, the endowment effect for risk and the limits of expectation-based reference-dependent preferences. Working paper.

Shapiro, C. and Stiglitz, J. E. (1984), 'Equilibrium unemployment as a worker discipline device', *American Economic Review* **74**(3), 433–444.

Smith, A. (2012), Lagged beliefs and reference-dependent preferences. Working paper.

Solnick, S. J. and Hemenway, D. (1998), 'Is more always better?: A survey on positional concerns', *Journal of Economic Behavior & Organization* **37**(3), 373–383.

Song, C. (2016), An experiment on reference points and expectations. Working paper.

Spear, S. E. and Wang, C. (2005), 'When to fire a CEO: Optimal termination in dynamic contracts', *Journal of Economic Theory* **120**(2), 239–256.

Spears, D. (2012), 'Low expectations: Reference-dependent preferences and labor supply in Cape Town, South Africa', *Poverty & Public Policy* **4**(1), 1–18.

Sprenger, C. (2015), 'An endowment effect for risk: Experimental tests of stochastic reference points', *Journal of Political Economy* **123**(6), 1456–1499.

Städter, S. (2018), 'The effects of pay regulation when agents are loss averse', *Applied Economics Letters* **25**(21), 1493–1498.

Starmans, J. (2017), Search unemployment in a frictional labor market with dispersion in incentive pay. Working paper.

Stiglitz, J. E. and Weiss, A. (1983), 'Incentive effects of terminations: Applications to the credit and labor markets', *American Economic Review* **73**(5), 912–927.

Teyssier, S. (2007), Optimal group incentives with social preferences and self-selection. Working paper.

Tversky, A. and Kahneman, D. (1991), 'Loss aversion in riskless choice: A reference-dependent model', *The Quarterly Journal of Economics* **106**(4), 1039–1061.

Tversky, A. and Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty* **5**(4), 297–323.

Van den Berg, G. J. and Ridder, G. (1998), 'An empirical equilibrium search model of the labor market', *Econometrica* **66**(5), 1183–1221.

Vendrik, M. C. and Woltjer, G. B. (2007), 'Happiness and loss aversion: Is utility concave or convex in relative income?', *Journal of Public Economics* **91**(7-8), 1423–1448.

Wakker, P. P. (2010), *Prospect Theory: For Risk and Ambiguity*, Cambridge University Press.

Wakker, P. P., Köbberling, V. and Schwieren, C. (2007), 'Prospect-theory's diminishing sensitivity versus economics' intrinsic utility of money: How the introduction of the Euro can be used to disentangle the two empirically', *Theory and Decision* **63**(3), 205–231.

Wang, C. (2011), 'Termination of dynamic contracts in an equilibrium labor market model', *Journal of Economic Theory* **146**(1), 74–110.

Wang, C. (2013), 'Incentives and the cost of firing in an equilibrium labor market model with endogenous layoffs', *International Economic Review* **54**(2), 443–472.

Wang, C. and Yang, Y. (2015), 'Equilibrium matching and termination', *Journal of Monetary Economics* **76**, 208–229.

Williamson, O. E. (1985), *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, Free Press.