Rethinking Hawks and Doves: Interim-Efficient Labor Contracts for Other-Regarding Agents

Thomas Daske*

September 21, 2015

Abstract

I introduce social preferences to a principal-two agents model. Before contracting, the agents’ social types are their private information. I derive interim-efficient labor contracts, for which I find the following: (1) In a perfectly competitive labor market, market equilibrium is ex interim Pareto dominated by a Perfect Bayesian equilibrium in which each firm perpetually hires two agents for a single period, offering them an interim-efficient labor contract; after any one period, all agents are released and randomly reallocated among firms. (2) Relative and team performance pay are interim-efficient and payoff equivalent. (3) There is no direct mechanism that incentivizes each agent to reveal his social type, while implementing unanimous approval of interim-efficient contracting. – These findings rationalize labor turnover for social reasons. They explain the coexistence of competition and collaboration in nearly any organization. They provide an example of mutual, incentive-compatible trust in a multilateral relationship.

Keywords: Contract design; relative and team incentives; altruism; spite; asymmetric information; interim-efficiency; corporate culture; labor turnover; trust.

JEL Classification: D62; D64; D82; M52

*Technische Universität München, Arcisstrasse 21, D-80333 Munich. Email: thomas.daske@tum.de
“Beware of false prophets,
which come to you in sheep’s clothing,
but inwardly they are ravening wolves.”

Matthew 7:15 (King James Version)

“Juliet: O serpent hid with a flowing face!
Did ever dragon keep so fair a cave?
Beautiful tyrant, fiend angelical!
Dove-feathered raven, wolish-ravening lamb.
Despised substance of divinest show!
Just opposite to what thou justly seem’st
A damned saint, an honourable villain.”

William Shakespeare, Romeo and Juliet

1 Introduction

Human interaction frequently involves a delicate question: What are my opponent’s intentions? – Does he care about his own well-being only? Or does he also care how his actions would affect me? And if so, is he friend or is he foe? – It is a truism that people rarely know beforehand what their opponents’ intentions really are. To ask them would hardly help, for foes in secret would assert being friends in heart. And taking shelter from a barrier of incentives that scare foes away might come at the adverse effect of scaring friends away also. As a matter of fact, people’s good or bad intentions are a persistent source of asymmetric information.

An institution strongly marked by its members’ intentions is the (white-collar) workplace. Two of its creatures are notorious: the careerist and the team player. – Do you have any questions? Do not annoy the careerist, for he is always ‘very busy at the moment.’ Rather approach the team player, ‘of course!’ he will help you. – What distinguishes the one from the other is their evaluations of the externalities that their actions impose on others. The team player is willing to spend some of his time to make others accomplish their doing, seeing clearly it will draw him back behind his own time schedule. The careerist has no time for that, though he appreciates the time of others. Closely related, but inverted, is their
perceptions of secret sabotage.\textsuperscript{1} As a rule of thumb: Stay close to team players, beware of careerists.

The positive and negative externalities inherent to voluntary help and secret sabotage are very often even made explicit in the contractual organization of work. They appear as relative or team performance incentives, and, typically, they are provided in parallel. In nearly any organization, coworkers collaborate in some project team while, at the same time, competing for promotion, bonuses, or simply approval. Such organizations are consultancies, political parties, and academia. Competition is as tough, and collaboration as poor, as careerists are numerous within the same occupation. Certainly, all employees prefer working alongside team players rather than careerists. But careerists do not reveal themselves from the outset. It rather seems to be a matter of luck whether employees, after a while of their affiliation, perceive the social environment at work as pleasant and supportive, or whether they wish to leave. Employers face a related problem with regard job entrants. The loyalty of careerists may much more lie with their employer than with their coworkers. On the other hand, the loyalty of team players may much more lie with each other than with their employer, such that team players may find it easier to collaborate against their employer.\textsuperscript{2} In this respect, employers may prefer careerists over team players. It seems to be a matter of luck whether employers turn out to be satisfied with their recruiting decisions, or whether they wish to fire and hire.

Consistent with these impressions from work life, the social environment at work, and especially the perceived support from coworkers, has been identified as a key factor for job satisfaction and actual labor turnover.\textsuperscript{3} From a theoretical point of view, however, this observation is somewhat puzzling, for it conflicts with the notion of labor market equilibrium as well as with the perception that strategic measures would help to fight off adverse selection. With this study, I seek to shed light on the economic relevance of asymmetric information on people’s intentions by rationalizing labor turnover for social reasons. I assume that, before contracting, job applicants are privately informed about their perceptions of the externalities that their actions impose on their coworkers. I show that labor turnover is then inherent to the provision of interim-efficient work incentives.

Among economists, it is by now widely accepted that many people, in one way or another,

\textsuperscript{1}Sabotage may be directed on a coworker’s productivity, for instance by not sharing valuable information, or it may be directed on his reputation, for instance by a misuse of an ‘anytime feedback tool’ as it is deployed by Amazon and "frequently" misused by Amazon’s employees (Kantor and Streitfeld 2015).

\textsuperscript{2}If, for instance, the employer has committed herself to pay the ‘employee of the month’ a bonus, then coworker collaboration may undermine competition for that bonus. A more severe example is white-collar crime, with coworkers covering up for each other.

internalize the externalities that their actions impose on others – that many people exhibit social preferences. I turn here to the concepts of altruism and spite. Altruism and spite capture people’s valuations of the distribution of wealth between others and themselves. They determine people’s willingness to pay for an increase, or decrease, of other people’s wealth by a given amount. In a nutshell, they operate like redistributive intentions. They have been shown to be consistent with many people’s behavior in experiments on games of cooperation and competition.4

I introduce altruistic and spiteful preferences to a principal-two agents model. The agents are risk-neutral and, to some extent, altruistic or spiteful toward their coworkers, but they do not internalize the externalities that their actions impose on the principal. The principal herself seeks material wealth only. The agents’ social types are independent draws from the same, commonly known distribution. Regarding this distribution I make the weakest assumption imaginable: Its variance is positive, its support is allowed to be arbitrary. That is, I assume people are different, and no more than that. To clearly isolate the impact that asymmetric information on social preferences has on the provision of efficient incentives, I abstract from any other form of (information) asymmetry: Except for their social preferences, the agents are identical in all their characteristics; individual effort is observable and verifiable by the principal. I consider a perfectly competitive labor market and assume that every single firm hires exactly two agents. As a benchmark, I characterize labor market equilibrium under full information: Since efforts are observable, each agent is incentivized to produce first-best efficiently and, due to free entry for firms, receives all the market returns from his individual output. I then turn to asymmetric information on the agents’ social types, with each agent’s type being his private information before contracting. I derive symmetric, interim-efficient labor contracts. I find that, for the triplet of a firm and two randomly selected, privately informed agents, the interim-efficient work incentives implement (by means of Perfect Bayesian equilibrium) interim Pareto improvement upon the first-best allocation of labor market equilibrium.

The feasibility of interim Pareto improvement upon the first-best is feasible for two reasons. First, social preferences allow for nonpecuniary compensation. And second, the principal on the one hand and the agents on the other have asymmetric, though individually rational, beliefs about the composition of social types at work. While every privately informed agent has to form a belief about one agent’s type only, the principal has to form a belief about the composition of two types. The interplay of both allows for mutually

4For evidence on altruism see Levine (1998) and Andreoni and Miller (2002). For evidence on spite see Saijo and Nakamura (1995), Levine (1998), Herrmann and Orzen (2008), and Prediger et al. (2014). See Carpenter and Seki (2011) for the important finding that participants’ prosocial behavior exhibited in the laboratory is positively correlated with prosocial behavior in a real-world labor environment.
beneficial ‘trade’ of nonpecuniary compensation.

Interim Pareto improvement upon the first-best through single-period, interim-efficient contracting provides each triplet of a principal and two randomly selected, privately informed agents with the incentive to deviate unanimously from labor market equilibrium. Indeed, there exists a Perfect Bayesian equilibrium in which each firm perpetually hires two randomly selected, privately informed agents for a single period, while utilizing an interim-efficient labor contract. After any one period of contracting, all agents are released and, as randomly matched pairs, reallocated among firms. This finding rationalizes labor turnover for social reasons.

The interim-efficient labor contracts implement a positive or negative externality of every agent’s effort on the other agent’s pay. That is, they implement either team or relative performance pay, and both forms of incentives are payoff equivalent.\footnote{A similar result is found by Dur and Sol (2010), but for a very different reason. In their model, both relative and team performance pay incentivize coworkers to invest in mutual favors, unrelated to production, that are reciprocated with altruism. This altruism helps coworkers to achieve their social optimum. Since coworkers also gain utility from favors, the principal can extract rents by reducing salaries.} When perceiving less altruistic agents as hawkish and more altruistic agents as dovish, the redistributive effects of interim-efficient contracting are exactly those of the hawk-dove game. While more altruistic types ‘cooperate’, less altruistic types ‘defect’.\footnote{Neugebauer et al. (2008) explore experimentally the social preferences that people exhibit in the hawk-dove game. As most participants act selfishly, while some act altruistically, behavior is largely consistent with the concept of altruism and spite. Neither does observed behavior support reciprocity, nor does it support inequity aversion (Fehr and Schmidt 1999) as the underlying preference concept.} In this context, the principal serves as a ‘bank’: she rewards dovishness when dove meets dove, and charges hawkishness when hawk meets hawk. As the principal seeks to maximize profits, she prefers hawkish agents. In contrast, every agent would benefit from meeting a dove, and suffer from meeting a hawk. Preferences for the composition of types are thus asymmetric between the principal on the one hand and the agents on the other. Intuitively, the principal can take advantage of hawkish careerists by giving them the opportunity to meet dovish team players. Since team players ease competition and enhance collaboration, careerists are willing to pay for that opportunity, which is to the principal’s advantage. However, to attract team players comes at a cost. Since careerists would take advantage of team players, team players must be given the opportunity to benefit excessively from meeting one another, which is to the principal’s disadvantage.\footnote{Notice that the principal prefers less altruistic agents not because she has a taste for ‘evil’ persons, but because coworker altruism facilitates collaboration against her.}

That agents and principals differ in their preferences for the composition of types at work contrasts with Lazear (1989), who puts forward the hawk-dove analogy with regard to
sabotage in contests.\footnote{Kandel and Lazear (1992) make a similar point for coworkers' responsiveness to peer pressure in teams.} Hawks, in Lazear (1989), are characterized by having a more effective sabotage technology than doves, and coworkers sabotage one another simply because they can; their intrinsic motivation to do so is pure payoff maximization. Since pure payoff maximization precludes nonpecuniary compensation, hawks cannot be taken advantage of; their presence is to everyone's disadvantage. Lazear shows that types do not self-sort, that hawks will impersonate doves in order to infiltrate dovish firms. He concludes that principals should screen applicants in order to keep hawks away. In the present study, the principal cannot gain from screening the agents: If the principal cannot credibly commit herself to contract after having learned the applicants' types, every hawkish applicant would infer from being screened that his colleague will be a hawk, too. The hawkish applicant would thus reject to contract with a principal who attempts to screen him. I show in particular that, with perfect labor market competition, imperfect commitment of the principal rules out the existence of a direct mechanism that incentives the agents to truthfully reveal their types before contracting, while implementing unanimous approval of an interim-efficient labor contract.\footnote{This finding contributes to the research, pioneered by Bester and Strausz (2000, 2001), which deals with the validity of the revelation principle if the principal has imperfect commitment.} In so far as they are concerned with the agents' social types, (principal-to-agent) job talks are cheap talk. Corporate culture, the social environment at work, is thus exposed to infiltration. Every new colleague might turn corporate culture into the better or the worse, depending on who judges it.

These findings are in stark contrast with the existing literature on corporate culture. Kosfeld and von Siemens (2009, 2011) argue that the provision of efficient incentives implies the separation of different social types in labor market equilibrium, establishing different corporate cultures across firms. In their studies, the workforce consists of conditional cooperators and unconditional noncooperators, where the first value mutual cooperation. The production process allows for gains from cooperation, and conditional cooperators benefit nonpecuniarily from cooperation.\footnote{Types then separate in labor market equilibrium, since conditional cooperators are willing to work for less pay. Effectively, conditional cooperators have access to a better technology; they are able to reach an agreement on mutual cooperation yielding pecuniary Pareto improvement.} – Taking this perspective, altruistic and spiteful agents all can be perceived as conditional cooperators. Even spiteful agents are willing to sign an implicit contract on mutual cooperation when it implements equally distributed material gains. The problem is that such contracts are hard to enforce. An example is voluntary help, the omission of which might appear as secret sabotage: I may possess some piece of information valuable to a coworker's current project. If the success of that project does not have any material effect on me, then sharing my information is a purely social act. This act
is hard to enforce, and it is not simply a matter of reciprocity.\footnote{See Rotemberg (1994), Rob and Zemsky (2002), and Dur and Sol (2010) for how coworkers’ cooperative attitudes might evolve endogenously. Implicit to these theories is the assumption that mutual cooperation is effectively contractible: To support a coworker inevitably enhances his attitude towards cooperation.} If my coworker does not know that I know what I know, then he cannot blame me for not sharing my knowledge with him.\footnote{This view is in line with Bandyet al. (2005). They find that many workers internalize, in favor of their coworkers, the negative externalities they impose on each other when being exposed to relative incentives. However, most of those workers only do that when having the opportunity to monitor each other.}

The findings also contrast with the view that corporate culture is something that could be ‘created’. Rotemberg (1994) and Chillemi (2008) argue that firms should try to foster coworker altruism. While this makes sense if production requires solely teamwork, it hardly matches the multitude of incentives that workers are exposed to in real world labor environments, including relative performance incentives. Pure coworker altruism facilitates collaboration against the principal, and it undermines competition for promotion, bonuses, and approval. To foster coworker altruism might therefore backfire on the principal.

The payoff equivalence of interim-efficient relative or team performance incentives suggests compromise to the long-lasting debate on whether and when relative incentives are superior to team incentives, and vice versa.\footnote{Fleckinger and Roux (2012) provide a survey on the matter.} While Lazear and Rosen (1981) have suggested relative incentives in the form of rank-order tournaments as optimal, others have argued that they may crowd-out helping efforts.\footnote{E.g., Lazear (1989) and Rob and Zemsky (2002). Drago and Garvey (1998) provide empirical support.} Furthermore, relative performance incentives require some degree of individual accountability; if only team performance is observable, pay cannot condition on relative performance. On the other hand, team performance pay entails the incentive to free-ride on the efforts of coworkers. And yet, both forms of incentives are pervasively provided in parallel. This coexistence could simply be attributed to the multidimensionality of white-collar work, with different dimensions of tasks being differently constrained with regard to the complementarity or observability of coworkers’ efforts.\footnote{See, e.g., Fleckinger (2012).} However, the parallel provision of relative and team incentives is especially intriguing when taking employees’ social concerns into account. Relative incentives cannot be optimal if all coworkers exhibit ‘prosocial’ preferences (e.g., Chillemi 2008, and Bartling 2011), and team incentives cannot be optimal if they all exhibit antisocial’ preferences (e.g., Itoh 2004). That is, if hawkish careerists and dovish team players separated in labor market equilibrium, a parallel use of relative and team incentives would yield inefficient outcomes.\footnote{Switching off relative incentives in favor of team incentives, or vice versa, seems to be simple: For instance, contests for promotion can be avoided when paying all those not promoted the money equivalent of promotion. Moral hazard in teams can often be overcome through relative incentives (Gershkov et al. 2009).}
hand, if labor market equilibrium involves the pooling of careerists and team players, ef-

cient labor contracts for job entrants would have to discriminate between those different

social types.\footnote{See, e.g., Fershtman et al. (2006) and Brunner and Sandner (2012) who determine the efficient (linear)

contracts for asymmetric agents whose altruistic and/or spiteful types are presumed common knowledge.} However, in most organizations, job entrants with identical vitas are offered identical, often standardized contracts. But even if labor contracts were to discriminate be-
tween different social types within the same firm (and hierarchy level), the notion of pooling

equilibrium conflicts with the empirical observation of labor turnover for social reasons.

Pioneered by Frank (1984a, 1984b), increasing effort has been spent on aligning agency

theory with people’s social preferences. However, the relevance and nature of asymmetric

information on social preferences has been ignored. The prevailing perception has been that

asymmetric information on social preferences could be treated like asymmetric information

on material entities, like ability, or private valuations of some good at stake – that it is simply

a matter of adverse selection. It is the baseline message of these pages that this perception

is mistaken. When met with appropriate incentives, the effects of asymmetric information

on people’s social preferences are mutually beneficial rather than adverse.

The paper proceeds as follows. Section 2 illustrates the economic relevance of asymmetric

information on people’s social preferences with the help of a simple zero-sum game. Section

3 derives interim-efficient labor contracts. Section 4 shows that these contracts implement

relative or team performance pay, and that both forms of compensation are payoff equivalent.

Section 5 rationalizes labor turnover for social reasons. Section 6 shows that job talks are

cheap talk and relates the findings to the notion of trust. Section 7 concludes.

2 Hawks and Doves – and Principal

A simple example helps to clarify the basic insight. Two risk neutral agents are each endowed

with one monetary unit. A risk neutral principal, $P$, plans to extract a rent from the agents’

wealth.

The following is common knowledge among the agents and the principal. Every agent $i$

exhibits a preference regarding the distribution of material wealth between himself and the

other agent. He maximizes utility $u_i = \pi_i + \delta_i \pi_{-i}$ where $\pi_i$ is $i$’s wealth, $\pi_{-i}$ is the other

agent’s wealth, and $\delta_i$ determines $i$’s social type. For simplicity, assume in this Section that

agent $i$ is either spiteful, with $\delta_i = -\frac{1}{2}$, or altruistic, with $\delta_i = \frac{1}{2}$. His social type is his

private information. Types are i.i.d. across agents and every agent is equally likely spiteful

or altruistic. The principal herself seeks material wealth only, and no agent cares about her.

We can think of her as (the representative of) a large, faceless organization.
The principal invites the agents to play a zero-sum game she calls hawk-dove-principal. The rules are as follows: The agents play the modified game of hawk-dove as given on the left hand side of Figure 1. The principal is not involved strategically, she rather serves as a ‘bank’. If both agents cooperate, P transfers $x > 0$ to each of them and makes herself a loss of $2x$ (that she can afford by assumption). If both agents choose to betray each other, P collects $4x$ from each of them and gains $8x$ accordingly. If one agent cooperates but is betrayed by the other, the cooperator loses $2x$ while the betrayer and P each gain $x$. Assuming the agents cannot expend more than their endowments, $x \leq \frac{1}{4}$. The game will be played if either agent agrees upon playing it; otherwise, the game is not played and endowments remain untouched.

Hawk-dove-principal, when played, implements a redistribution of wealth the direction of which is determined by the agents’ actions. These actions are determined by the agents’ social types. The table in the upper right corner of Figure 1 shows how the potential game outcomes affect an agent who is spiteful; see the figure description for details. To betray is a dominant strategy for the spiteful agent, so we may call him a hawk. The table in

Figure 1: On the left: The hawk-dove game that the principal invites the agents to play, with $x > 0$. At the upper right: A spiteful type behaves hawkish; he maximizes utility $u_{\text{Hawk}} = \pi_{\text{Hawk}} - \frac{1}{2}\pi_{-i}$ by unconditionally betraying (B) the other agent $-i$. At the lower right: An altruist behaves dovish; he maximizes utility $u_{\text{Dove}} = \pi_{\text{Dove}} + \frac{1}{2}\pi_{-i}$ by unconditionally cooperating (C) with the other agent $-i$. 

\[
\begin{array}{ccc}
\pi_{-i} & u_{\text{Hawk}} & u_{-i} \\
\hline
\text{Betray} & \text{Cooperate} & \\
B & -4x \backslash -4x & x \backslash -2x \\
C & -2x \backslash x & x \backslash x \\
\end{array}
\]

\[
\begin{array}{ccc}
\pi_{-i} & u_{\text{Dove}} & \\
\hline
\text{Betray} & \text{Cooperate} & \\
B & -6x \backslash ... & 0 \backslash ... \\
C & -\frac{3}{2}x \backslash ... & \frac{3}{2}x \backslash ... \\
\end{array}
\]
Figure 2: The monetary returns that a hawk \(H\), a dove \(D\), and the principal \(P\) would obtain ex post from each of the feasible compositions of types \((\cdot, \cdot)\), and the likelihoods \((Pr)\) that types and principal rationally attribute to each of these compositions to occur.

The lower right corner of Figure 1 shows how the game affects an altruist. For an altruist, it is a dominant strategy to cooperate, so we may call him a dove. In particular, the game incentivizes the agents to reveal their social types (ex post) by providing them with mutually exclusive dominant strategies.

Since information is asymmetric ex interim, hawk-dove-principal is effectively a gamble over the composition of types at play. Each type would benefit from meeting a dove but would suffer from meeting a hawk. The principal would benefit from meeting at least one hawk, but she would suffer from meeting two doves. Since playing the game requires the agents to unanimously agree upon play, and since the principal would certainly not invite the agents to play it when expecting a loss, hawk-dove-principal comes into play only if it yields Pareto improvement ex interim. Notice that a redistribution of wealth yielding Pareto improvement ex post is generally not feasible.\(^{18}\)

Figure 2 depicts the agents’ and the principal’s beliefs about the different compositions of types to occur as well as the payoffs they each would get respectively. Beliefs about

\(^{18}\)Consider any zero-sum game, \(\pi_P + \pi_i + \pi_{-i} = 0\), that provides \(P\) with a positive rent, \(\pi_P > 0\). Ex post Pareto improvement requires \(0 \leq u_i, u_{-i}\). Therefore, \(0 < \pi_P = -\pi_i - \pi_{-i} = -\pi_i - (u_{-i} - \delta_{-i}\pi_i) = - (1 - \delta_{-i})\pi_i - u_{-i} \leq - (1 - \delta_{-i})\pi_i\). This implies \(\pi_i < 0\), since \(\delta_{-i} = \pm \frac{1}{2}\). By exchanging the roles of \(i\) and \(-i\), we get \(\pi_{-i} < 0\). But then, \(0 \leq u_i + u_{-i} = (1 + \delta_{-i})\pi_i + (1 + \delta_i)\pi_{-i} < 0\). A contradiction.
compositions differ between agents and principal. For a dove ($D$) it is equally likely to meet a hawk ($H$) or another dove. She attributes likelihood zero to composition $(H, H)$. Similarly, it is equally likely for a hawk to meet a dove or another hawk, and he attributes likelihood zero to composition $(D, D)$. The principal, however, attributes positive likelihoods to all feasible compositions. Given that types are i.i.d., and equally likely across agents, she expects each of the compositions $(H, H)$, $(H, D)$, $(D, H)$, and $(D, D)$ to occur with likelihood $\frac{1}{4}$. By symmetry of hawk-dove-principal, she will be equally affected by $(H, D)$ and $(D, H)$, so she attributes likelihood $\frac{1}{2}$ to meeting one hawk and one dove. Beliefs about the composition of types at play are thus asymmetric between the principal on the one hand and the agents on the other, even though all their beliefs that are individually rational. As we shall see, this asymmetry in individually rational beliefs about the composition of types at play is one of the two driving forces behind interim Pareto improvement. The other one is the agents’ social preferences which allow material compensation to be substituted by mental compensation.

Suppose every agent, whether hawk or dove, expects the other agent (whether hawk or dove) to agree to play. Then both hawks and doves expect utility zero from play. This is immediate from Figure 1 where the levels of utility that a hawk, or dove, would derive from meeting hawk or dove just add up to zero. Every agent is thus indifferent between agreeing to play and rejecting play, and is assumed to agree. Agreeing to play while consistently believing that any hawk will choose to betray and any dove will choose to cooperate therefore constitutes a Perfect Bayesian equilibrium for the agents, regardless of their types. Assuming that both agents do agree to play, the principal expects a positive rent of $2x$, which is immediate from Figure 2. So, indeed, we face interim Pareto improvement.\footnote{In Section 6, I discuss the implications of this finding for the revelation principle. The Appendix includes a generalization of the finding to agents who exhibit constant relative risk aversion.} How come?

Looking at Figure 2, we see that a dove gains $x$ from meeting another dove, but that she loses $2x$ from meeting a hawk. She thus expects a wealth of $-\frac{1}{2}x$. For this material loss she receives mental compensation: Her counterpart, hawk or dove, gains $x$, which she values as $\frac{1}{2}x$, since $\delta_i = \frac{1}{2}$ for a dove. A dove’s interim expected utility from play (on top of her endowment of one) is thus zero. The principal, however, would make a substantial loss of $2x$ when meeting two doves, and she receives only little compensation, $x$, when meeting one hawk and one dove. And yet, she is indifferent between meeting one or two doves: For her, meeting two doves is only half as likely as meeting hawk and dove. That is, if she can manage to benefit from meeting two hawks, she is strictly better off ex interim.

A hawk benefits from meeting a dove in two ways, materially (he gains $x$) and mentally (the dove ends up with $-2x$, which the hawk values as $x$). The principal can exploit both forms of compensation, material and mental, by charging hawkishness when hawk meets
hawk. She demands the maximum amount that leaves a hawk just indifferent, which is $4x$, and thus gains $8x$. Unfortunately for her, two hawks are quite unlikely, so she expects an overall rent of $2x$.\(^{20}\)

Obviously, the determinants of the game have been aligned to the distribution of altruistic and spiteful types. Interim Pareto improvement requires the institutional environment to match the distribution of types. The necessary and sufficient ingredients for rendering interim Pareto improvement possible are the feasibility of mental compensation and asymmetric, though individually rational, beliefs about the composition of types at play.

Now, how does hawk-dove-principal relate to work incentives and labor contracts? First, with different social types in the population of agents, and with agents being privately informed about their types, the principal can implement hawk-dove-principal by means of both relative and team performance pay. She can employ two agents, and link their wealth by imposing a positive or negative externality of every agent’s effort on the other agent’s pay. This makes an altruist act dovish and a spiteful type act hawkish. Ex post, every agent will have revealed his social type through the effort he exerted. Ex interim, the principal and the agents will have asymmetric, though individually rational, beliefs about the composition of types at work. And second, labor contracts incentivize the agents to produce first what is to be distributed then. In the example above, the agents’ endowments can be endogenized, substituted by the returns from production in the principal’s firm. The theme of the next two Sections is that hawk-dove-principal, when modified like this, yields labor contracts that are interim-efficient.

3 Interim-Efficient Labor Contracts

In this and the next Section, I show that single-period, interim-efficient contracting between a principal and two randomly selected agents who are privately informed about their social types allows for interim Pareto improvement upon equilibrium in a perfectly competitive labor market. My strategy is as follows. For a perfectly competitive labor market, with each firm hiring exactly two agents, I characterize labor market equilibrium. The production technology is such that, in equilibrium, each agent produces first-best efficiently and earns the market return of his individual output, while firms make zero profits. I take this allocation as a (first-best efficient) status quo and show that a single firm can realize a positive ex ante expected profit when offering two randomly selected, privately informed agents a single-period, interim-efficient labor contract. Reservation utilities are taken as the utilities that

\(^{20}\)For the sake of completeness, the principal can provide herself with a maximum expected rent of $2x = \frac{1}{2}$, since $x \leq \frac{1}{4}$ due to the agents’ endowments of one unit per agent.
the agents would realize in labor market equilibrium. That is, I derive interim-efficient labor contracts that Pareto improve ex interim upon the first-best efficient status quo. I first identify necessary conditions which such labor contracts must satisfy. These conditions are closely related to the redistributive effects of hawk-dove-principal from Section 2. I then derive sufficient conditions that do allow for interim Pareto improvement upon the first-best.

3.1 The Analytical Framework

In a perfectly competitive labor market, each active firm hires exactly two agents in order to produce a homogeneous good. Each firm is represented by a risk-neutral, profit maximizing principal \( P \); the agents hired by \( P \) are interchangeably referred to as \( i \) and \( -i \). Agents are risk-neutral. Physically, all agents are equally productive. Their efforts are perfectly substitutable, and they can be observed and justified by \( P \). For his individual output of \( x_i \) units of the good, every agent \( i \) faces individual costs of effort \( C(x_i) \), where \( C : [0, \infty) \rightarrow [0, \infty) \) satisfies \( C(0) = C_x(0) = 0; C_x, C_{xx} > 0; \) and \( \lim_{x \to \infty} C_x(x) = \infty. \) Total output, \( x_i + x_{-i} \), is sold at an exogenously given market price of \( 1 \) per unit. To incentivize production, \( P \) offers \( i \) and \( -i \) a contract \( W = [w_i(x_i, x_{-i}), w_{-i}(x_{-i}, x_i)] \), with \( w_i(x_i, x_{-i}) \) and \( w_{-i}(x_{-i}, x_i) \) denoting the wages of agents \( i \) and \( -i \) contingent on their individual efforts. I assume limited liability for the agents: \( w_i(x_i, x_{-i}) \geq 0 \) for all \( x_i, x_{-i} \in [0, \infty) \). Limited liability does not apply to principals.

For given efforts, each active firm \( P \) gains a profit of \( \pi_P = x_i + x_{-i} - w_i(x_i, x_{-i}) - w_{-i}(x_{-i}, x_i) \). Every agent \( i \) realizes a material wealth of \( \pi_i = w_i(x_i, x_{-i}) - C(x_i) \) from production. He exhibits an altruistic or spiteful preference with regard to the ex post distribution of wealth between himself and his coworker \( -i \). From his own wealth \( \pi_i \), and a wealth of \( \pi_{-i} \) for his coworker \( -i \), agent \( i \) gains utility

\[
    u_i = \pi_i + \delta_i \pi_{-i}. \tag{1}
\]

The weight \( \delta_i \in [\delta_{\text{min}}, \delta_{\text{max}}] \), where \(-1 < \delta_{\text{min}} < \delta_{\text{max}} < 1 \), denotes \( i \)'s degree of altruism, or \( i \)'s type. Every agent's type is exogenously given and unaffected by his coworker's type.

It is common knowledge among firms and agents that the agents' social types are independent draws from the same distribution, given by the c.d.f. \( F : [\delta_{\text{min}}, \delta_{\text{max}}] \to [0, 1] \). \( F \) is continuously differentiable, and its density \( f \) satisfies \( f(\delta_{\text{min}}) = 0 = f(\delta_{\text{max}}) \), and \( f(\delta) > 0 \) at each \( \delta \in (\delta_{\text{min}}, \delta_{\text{max}}) \). Before contracting, every agent is privately informed about the

---

21 Lemma 1 below does not require any further assumptions on the functional form of \( W \).

22 All the Lemmas and Propositions in this paper are also valid for type distributions containing mass points. Notice also that the location of \( [\delta_{\text{min}}, \delta_{\text{max}}] \) within \((-1, 1)\) can be taken arbitrary: all results do hold.
realization of his own social type.

The following Lemma characterizes labor market equilibrium. Notice that the result does not condition on the information that other parties may or may not have on an agent’s social type.\textsuperscript{23}

**Lemma 1** Equilibrium in the perfectly competitive labor market is characterized as follows: Each agent is provided with the incentive to exert the first-best efficient effort $x^* = C_x^{-1}(1)$; each agent $i$, with type $\delta_i$, earns $x^*$ and realizes utility $u_{\delta_i} = (1 + \delta_i) R$, where $R = x^* - C(x^*)$; firms make zero profits.

In the following, I refer to the allocation of wealth characterized by Lemma 1 as the *first-best efficient status quo*.

In order to provide herself with a positive ex ante expected profit, a single principal $P$ plans to deviate from labor market equilibrium and to hire two randomly selected agents $i$ and $-i$ for a single period. The composition of social types is the outcome of two independent random draws according to the exogenous distribution $F$ of social types in the population of agents. The principal’s and the agents’ common prior about the composition of types at work is thus $(\delta_i, \delta_{-i}) \sim F \times F$. However, before contracting, the agents are privately informed about their social types. Formally, each agent learns privately the realization of his own type before contracting; agent $i$’s updated belief about the composition of types is thus $\delta_{-i} \sim F$; similarly, agent $-i$’s updated belief is $\delta_i \sim F$. The principal sticks with her prior $(\delta_i, \delta_{-i}) \sim F \times F$.

The timing of decision making is as follows: First, the principal offers the agents a single-period contract $W = [w_i(x_i, x_{-i}), w_{-i}(x_{-i}, x_i)]$, where $w_i(x_i, x_{-i}), w_{-i}(x_{-i}, x_i) \geq 0$ denote the wages of agents $i$ and $-i$ contingent on their individual efforts. Second, the agents decide simultaneously on whether to accept the offer, or not; contracting requires unanimous approval. If $W$ does not find unanimous approval, then all parties realize their status quo utilities and profits. If $W$ does find unanimous approval, then the agents simultaneously decide on their effort choices $x_i, x_{-i}$. Finally, after contracting for a single period, wage payments $w_i, w_{-i}$ are made and both agents are released (with all parties realizing their status quo utilities and profits in all subsequent periods). $P$ plans to offer the agents a contract $W$ that, by means of Perfect Bayesian equilibrium, implements interim Pareto improvement upon the first-best efficient status quo. The contract $W$ is supposed to attract each feasible type $\delta \in [\delta_{\min}, \delta_{\max}]$.\textsuperscript{24} That is, in Perfect Bayesian equilibrium, $W$ is supposed for populations of agents which are entirely spiteful or entirely altruistic.

\textsuperscript{23}The proofs of all Lemmas and Propositions are put in the Appendix.

\textsuperscript{24}I do not address the question whether $P$ can achieve superior outcomes by attracting only some types,
to yield any type an interim expected utility greater than his reservation utility from the first-best efficient status quo, \( u_i = (1 + \delta_i) R \), provided that each type believes that each other type will accept \( W \).

By assumption, the contract \( W = [w(x_i, x_{-i}), w(x_{-i}, x_i)] \) is symmetric, and the wage incentive scheme \( w : [0, \infty)^2 \rightarrow [0, \infty) \) is twice partially continuously differentiable in \( x_i \) and \( x_{-i} \). In order to keep the analysis tractable, I restrict attention to dominant strategy implementation: \( W \) provides every type \( \delta \in [\delta_{\min}, \delta_{\max}] \) with a dominant strategy effort level \( x(\delta) \), where \( u \leq x(\delta) \leq v \) for some constants \( u, v \) satisfying \( 0 \leq u < v \). The effort function \( x : [\delta_{\min}, \delta_{\max}] \rightarrow [u, v] \) maps every type \( \delta \) onto this type’s individually optimal effort level \( x(\delta) \). In addition, \( W \) is supposed to imply that the effort function \( x \) is continuously differentiable and strictly monotone: either \( x_\delta(\delta) > 0 \), or \( x_\delta(\delta) < 0 \), for all \( \delta \in [\delta_{\min}, \delta_{\max}] \). As contracting is only for a single period, this property ensures that \( W \) provides different social types with mutually exclusive dominant effort strategies, such that each agent reveals his social type through the effort he exerts.\(^{25}\)

Provided that \( W \) does attract all feasible types, and that the agents reveal their social types through the efforts they exert, every agent’s rational belief about his coworker’s social type is given by the exogenous, commonly known distribution of types, \( F : [\delta_{\min}, \delta_{\max}] \rightarrow [0, 1] \). The principal’s rational belief about the composition of social types at work, \( F_P : [\delta_{\min}, \delta_{\max}]^2 \rightarrow [0, 1] \), is induced by the fact that types are independent draws from the same distribution \( F \); it satisfies \( F_P(\delta, \tau) = F(\delta) F(\tau) \) for all \( (\delta, \tau) \in [\delta_{\min}, \delta_{\max}]^2 \). The principal’s and the agents’ beliefs about the composition of types at work are mutually consistent if and only if their expected utilities and profits under \( W \) are indeed higher than their utilities and profits in the first-best efficient status quo: If each privately informed agent \( i \) of type \( \delta_i \) believes that his coworker’s type is \( \tau \sim F \), then his interim expected utility from \( W \) satisfies \( \mathbb{E}[u_i | \delta_i] \geq u_{\delta_i} \); and if the principal believes that the composition of social types at work is \( (\delta, \tau) \sim F \times F \), then her ex ante expected profit from \( W \) satisfies \( \mathbb{E}_{(\delta, \tau)}[\pi_P] > 0 \).\(^{26}\) If the latter conditions are satisfied, unanimous approval of \( W \) is the outcome of a Perfect Bayesian equilibrium, implementing interim Pareto improvement upon the first-best.

If it satisfies all the above conditions, the contract \( W \) is said to \textit{implement interim Pareto improvement upon the first-best efficient status quo}. Figure 3 summarizes the timing of belief formation and decision making.

\(^{25}\)I restrict attention to dominant strategy implementation in order to keep the analysis simple.

\(^{26}\)The operators \( \mathbb{E}[^{\cdot}] \) and \( \mathbb{E}_{(\delta, \tau)}[^{\cdot}] \) indicate that expectations are taken with respect to \( F \) and \( F \times F \), respectively.
agents privately learn their social types
beliefs about composition of types are formed
single period contract offered by principal
contract accepted by the agents
dominant strategy efforts exerted and observed
payments made, agents released
time

\[ \delta_i, \delta_{-i} \sim F \]
\[ i: \delta_i \sim F \]
\[ -i: \delta_{-i} \sim F \]
\[ P: (\delta_i, \delta_{-i}) \sim F \times F \]

Figure 3: The timeline of belief formation and decision making in single-period, interim-efficient contracting between a principal \( P \) and two randomly selected, privately informed agents \( i \) and \(-i\). By assumption, the contract \( W \) offered by the principal is such that the principal and the agents consistenly believe that each agent \( i \), regardless of his type \( \delta_i \), is willing to accept \( W \) and to reveal his type through exerting his dominant strategy effort \( x_i = x(\delta_i) \). Dominant strategy efforts are mutually exclusive for different social types.

### 3.2 Necessary Conditions

In the following, the operators \( \mathbb{E} [ \cdot ] \) and \( \text{Var} [ \cdot ] \) indicate that expectations are taken with respect to the type distribution \( F \). The assumptions on \( F \) imply that \( \text{Var} [ \delta ] > 0 \). The operator \( \mathbb{E}_{(\delta, \tau)} [ \cdot ] \) indicates that expectations are taken over the joint distribution \( F_{(\delta, \tau)} \) of two independent draws \( \delta, \tau \in [\delta_{\min}, \delta_{\max}] \), which is determined by \( F_{(\delta, \tau)} = F(\delta) F(\tau) \) for all \( (\delta, \tau) \in [\delta_{\min}, \delta_{\max}]^2 \). For any function \( g : \mathbb{R}^2 \rightarrow \mathbb{R} \), I denote by \( \mathbb{E} [ g(\delta, \tau) | \delta ] \) the conditional expected value of the random variable \( g(\delta, \tau) \) for a given value of \( \delta \). Similarly for \( \mathbb{E} [ g(\delta, \tau) | \tau ] \).

Suppose first that there are contracts \( W \) that do allow for interim Pareto improvement upon the first-best efficient status quo. Then \( P \) chooses \( W \) in order to maximize

\[
\mathbb{E}_{(\delta, \tau)} [ \pi_P ] = \mathbb{E}_{(\delta, \tau)} [ x(\delta) - w(x(\delta), x(\tau)) + x(\tau) - w(x(\tau), x(\delta))] \\
= 2\mathbb{E} [ x(\delta)] - 2\mathbb{E} [ \mathbb{E} [ w(x(\delta), x(\tau)) | \delta ]],
\]

where \( x : [\delta_{\min}, \delta_{\max}] \rightarrow [x, \overline{x}] \) denotes the presumed dominant strategy effort levels of arbitrary types \( \delta, \tau \in [\delta_{\min}, \delta_{\max}] \). Interm Pareto improvement upon the first-best requires that each type \( \delta_i \) derives an interim expected utility from implementation of \( W \) which is not inferior to his reservation utility, \( u_{\delta_i} = (1 + \delta_i) R \). Let \( \pi(x(\delta_i), x(\tau)) = w(x(\delta_i), x(\tau)) - C(x(\delta_i)) \) denote the ex post wealth of type \( \delta_i \) if this type meets some type \( \tau \). Then, by Lemma 1, type \( \delta_i \)’s interim expected utility from implementation of \( W \)
must satisfy
\[
\mathbb{E} [u_i | \delta_i] = \mathbb{E} [\pi (x (\delta_i), x (\tau)) | \delta_i] + \delta_i \mathbb{E} [\pi (x (\tau), x (\delta_i)) | \delta_i] \\
= (1 + \delta_i) R + u (\delta_i),
\]
for some real number \( u (\delta_i) \geq 0 \). For every type \( \delta \in [\delta_{\min}, \delta_{\max}] \), the mapping
\[
u : \begin{cases}
[\delta_{\min}, \delta_{\max}] &\rightarrow [0, \infty) \\
\delta &\rightarrow u (\delta)
\end{cases}
\]
describes a type \( \delta \)'s interim expected utility gain from implementation of \( W \). I refer to the function \( u (\delta) \) as the gain function. To attract all types in \([\delta_{\min}, \delta_{\max}]\) requires that \( u \geq 0 \). Given \( W \) and \( F \), the gain function is unique. A sound understanding of its properties is key.

**Lemma 2** Suppose there is a contract \( W = [w(x_i, x_{-i}), w(x_{-i}, x_i)] \), with \( w : [0, \infty)^2 \rightarrow [0, \infty) \) twice partially continuously differentiable, that provides each type \( \delta \in [\delta_{\min}, \delta_{\max}] \) with a strictly dominant effort strategy \( x (\delta) \in [x, \overline{x}] \), where \( x : [\delta_{\min}, \delta_{\max}] \rightarrow [x, \overline{x}] \) is continuously differentiable and strictly monotone. Suppose further that, for the triplet of a principal and two randomly selected, privately informed agents, \( W \) implements interim Pareto improvement upon the first-best efficient status quo. Then the following holds necessarily:

(i) The gain function \( u : [\delta_{\min}, \delta_{\max}] \rightarrow [0, \infty) \) is twice differentiable and strictly convex. It satisfies \( \mathbb{E} [u_\delta (\delta)] < 0 \) and \( \mathbb{E} [u (\delta)] = \mathbb{E} [(1 + \delta) u_\delta (\delta)]. \)

(ii) The principal expects a profit of \( \mathbb{E} [u_i | \pi_P] = 2 \mathbb{E} [u_\delta (\delta)] + 2 (\mathbb{E} [x - C (x)] - R). \)

(iii) Type \( \delta_i \) expects an own wealth of \( \mathbb{E} [\pi_i | \delta_i] = R + u (\delta_i) - \delta_i u_\delta (\delta_i). \)

(iv) Type \( \delta_i \) expects a wealth of \( \mathbb{E} [\pi_{-i} | \delta_i] = R + u_\delta (\delta_i) \) for his coworker \(-i\).

The gain function fully determines an agent’s own expected wealth, \( \mathbb{E} [\pi_i | \delta_i] \), as well as the externality he imposes on his coworker, \( \mathbb{E} [\pi_{-i} | \delta_i] \). Obviously, \( \mathbb{E} [u_i | \delta_i] = \mathbb{E} [\pi_i | \delta_i] + \delta_i \mathbb{E} [\pi_{-i} | \delta_i] = (1 + \delta_i) R + u (\delta_i), \) as demanded. A stylized illustration of these conditional expectations as functions of an agent’s type is given in Figure 4. Own expected wealth is inverse U-shaped, since \( \frac{d}{d\delta_i} \mathbb{E} [\pi_i | \delta_i] = -\delta_i u_{\delta\delta} (\delta_i) \), and \( u_{\delta\delta} > 0 \). It is largest for the pure payoff maximizer, \( \delta_i = 0 \); his own expected wealth must not be smaller than in the status quo, \( \mathbb{E} [\pi_i | 0] \geq R \), for he does not internalize the externality he imposes on others. The more an agent internalizes the externality he imposes on his coworker (that is, the larger \( |\delta_i| \)), the more rent can \( P \) extract by substituting material compensation, \( \mathbb{E} [\pi_i | \delta_i] \), with mental compensation, \( \delta_i \mathbb{E} [\pi_{-i} | \delta_i] \). By Lemma 2(iv), expected mental compensation across
agents can be written as \( \mathbb{E} [\delta_i | \delta_i] = \mathbb{E} [\delta] R + \mathbb{E} [\delta u_\delta (\delta)] \). The conditions of Lemma 2(i) imply \( \mathbb{E} [\delta u_\delta (\delta)] = \mathbb{E} [u (\delta)] + |\mathbb{E} [u_\delta (\delta)]| \). Since \( u \geq 0 \) is strictly convex, we have \( \mathbb{E} [u (\delta)] > 0 \) and thus \( \mathbb{E} [\delta u_\delta (\delta)] > 0 \). Expected mental compensation entails the interim expected utility gain across agents, \( \mathbb{E} [u (\delta)] \), and a second term, \( |\mathbb{E} [u_\delta (\delta)]| \), which enters directly \( P \)'s ex ante expected profit as given by Lemma 2(ii). By Lemma 2(iii), expected individual wealth across agents is then \( \mathbb{E} [\mathbb{E} [\pi_i | \delta_i]] = R - |\mathbb{E} [u_\delta (\delta)]| \). Therefore, \( |\mathbb{E} [u_\delta (\delta)]| \) can be said to measure the strength of redistribution from the agents to the principal. What \( P \) extracts from the agents’ wealth is given back to the agents mentally: as in hawk-dove-principal from Section 2, mental compensation substitutes for material compensation. Furthermore, since \( |\mathbb{E} [u_\delta (\delta)]| = \mathbb{E} [\delta u_\delta (\delta)] - \mathbb{E} [u (\delta)] < \mathbb{E} [\delta u_\delta (\delta)] \), the principal cannot extract fully the money equivalent of expected mental compensation. In addition, due to the strict convexity of \( u \), the principal can only make one type indifferent between accepting and rejecting \( W \); all the other types are ex interim strictly better off when \( W \) is implemented.

An agent’s expected externality on his coworker’s wealth, \( \mathbb{E} [\pi_{-i} | \delta_i] = R + u_\delta (\delta_i) \), strictly increases in the agent’s social type. Since \( \mathbb{E} [u_\delta (\delta)] < 0 \), the least altruistic type, \( \delta_{\min} \), imposes a negative externality on his coworker, \( u_\delta (\delta_{\min}) < 0 \). On the other hand, since \( \mathbb{E} [u (\delta)] = \mathbb{E} [(1 + \delta) u_\delta (\delta)] \), the most altruistic type, \( \delta_{\max} \), imposes a positive externality on his coworker, \( u_\delta (\delta_{\max}) > 0 \). For the agents, the redistributive effects of interim-efficient contracting are therefore similar to those of the basic hawk-dove game.

The conditions of Lemma 2(i) are invariant under a multiplication of \( u \) with a positive scalar. This property is important when it comes to the agents’ limited liability. As suggested by parts (iii) and (iv) of Lemma 2, the magnitudes of \( u \) and \( u_\delta \) do impact the agents’ ex post wage payments. Limited liability requires that \( u \), and thus the strength of redistribution between the principal and the agents, \( |\mathbb{E} [u_\delta (\delta)]| \), has a sufficiently small upper bound.

By Lemma 2(ii), the principal’s ex ante expected profit is separable in the strength of redistribution and a measure for the efficiency of production, \( \mathbb{E} [x - C (x)] - R \). However, production is ex interim necessarily less than first-best efficient, \( \mathbb{E} [x - C (x)] < R \). With bounded strength of redistribution, \( P \)'s objective must be to implement dominant strategy efforts \( x : [\delta_{\min}, \delta_{\max}] \rightarrow [\underline{x}, \overline{x}] \) sufficiently close to the first-best efficient, \( x^* = C_{\overline{x}}^{-1} (1) \).

---

27Suppose the opposite: \( u_\delta (\delta_{\max}) < 0 \). Then \( u_\delta < 0 \), since \( u_{\delta, \delta} > 0 \). Therefore, \( \mathbb{E} [(1 + \delta) u_\delta (\delta)] < 0 \), since \( 0 < 1 + \delta \) for all \( \delta \in [\delta_{\min}, \delta_{\max}] \). However, the strict convexity of \( u \geq 0 \) implies \( 0 < \mathbb{E} [u (\delta)] \). Together, \( 0 < \mathbb{E} [u (\delta)] = \mathbb{E} [(1 + \delta) u_\delta (\delta)] \leq 0 \); a contradiction.

28Otherwise, if \( \mathbb{E} [x - C (x)] = R = x^* - C (x^*) \), all the agents are required to exert the first-best efficient effort level \( x^* \), effectively resolving all information asymmetry.
Figure 4: A stylized illustration of an agent’s expectations about his own wealth, \( \mathbb{E}[\pi_i|\delta_i] \), and his coworker’s wealth, \( \mathbb{E}[\pi_{-i}|\delta_i] \), both as functions of the agent’s type \( \delta_i \). For the strictly convex gain function \( u \geq 0 \), these expectations read \( \mathbb{E}[\pi_i|\delta_i] = R + u(\delta_i) - \delta_i u_\delta(\delta_i) \) and \( \mathbb{E}[\pi_{-i}|\delta_i] = R + u_\delta(\delta_i) \).

### 3.3 Sufficient Conditions

Presume the existence of a gain function \( u : [\delta_{\min}, \delta_{\max}] \to [0, \infty) \) and an effort function \( x : [\delta_{\min}, \delta_{\max}] \to [x, \bar{x}] \) satisfying jointly all the assumptions and conditions of Lemma 2. The conditional expectations given by Lemma 2 can be disentangled by imposing some structure on the wage function \( w \). Let \( w(x_i, x_{-i}) = v(x_i) + z(x_i) - z(x_{-i}) \) for sufficiently smooth functions \( v, z : [0, \infty) \to \mathbb{R} \). Given the ex post wealth of \( \pi(x(\delta), x(\tau)) = w(x(\delta), x(\tau)) - C(x(\delta)) \) for an agent of type \( \delta \) meeting some type \( \tau \), the identities in parts (iii) and (iv) of Lemma 2 now read

\[
\begin{align*}
  u(\delta) - \delta u_\delta(\delta) + R &= v(x(\delta)) + z(x(\delta)) - C(x(\delta)) - \mathbb{E}[z(x(\tau))] , \\
  u_\delta(\delta) + R &= \mathbb{E}[v(x(\tau))] + \mathbb{E}[z(x(\tau))] - \mathbb{E}[C(x(\tau))] - z(x(\delta)).
\end{align*}
\]

The principal’s ex ante expected profit simplifies to

\[
\mathbb{E}(\delta, \tau)[\pi_P] = 2\mathbb{E}[x(\tau)] - 2\mathbb{E}[v(x(\tau))].
\]
Combining (7) with Lemma 2(ii) yields \( \mathbb{E}[v(x(\tau)) - C(x(\tau))] = \mathbb{E}[u_\delta(\delta)] + R \). Substituting the latter into (6) yields
\[
z(x(\delta)) = \mathbb{E}[z(x(\delta))] + \mathbb{E}[u_\delta(\delta)] - u_\delta(\delta).
\] (8)

Adding now (5) and (8) implies
\[
v(x(\delta)) = R - \mathbb{E}[u_\delta(\delta)] + C(x(\delta)) + u(\delta) + (1 - \delta) u_\delta(\delta).
\] (9)

Denote by \( h: [\bar{x}, \bar{\tau}] \to [\delta_{\min}, \delta_{\max}] \) the inverse of the strictly monotone effort function \( x: [\delta_{\min}, \delta_{\max}] \to [\bar{x}, \bar{\tau}] \); that is, \( \delta = h(x(\delta)) = h(x) \) for any \( \delta \in [\delta_{\min}, \delta_{\max}] \). Substituting \( \delta = h(x) \) into (9), and ‘normalizing’ \( \mathbb{E}[z(x)] = -\mathbb{E}[u_\delta(\delta)] \) in (8), yields a candidate for a contract \( W \) implementing the effort function \( x \) and the gain function \( u \): Consider the wage incentive scheme \( w(x_i, x_{-i}) = v(x_i) + z(x_i) - z(x_{-i}) \), where
\[
v(x) = R - \mathbb{E}[u_\delta(\delta)] + C(x) + [1 - h(x)] u_\delta(h(x)) + u(h(x)),
\] (10)
\[
z(x) = -u_\delta(h(x)).
\] (11)

The following Lemma is the central device for the construction of interim-efficient labor contracts.

**Lemma 3** Suppose \( u: [\delta_{\min}, \delta_{\max}] \to [0, \infty) \) satisfies all the conditions of Lemma 2(i). Suppose the function \( x: [\delta_{\min}, \delta_{\max}] \to [\bar{x}, \bar{\tau}] \) is continuously differentiable and strictly monotone, and it satisfies \( \mathbb{E}[x(\delta) - C(x(\delta))] > R + \mathbb{E}[u_\delta(\delta)] \). Let \( h: [\bar{x}, \bar{\tau}] \to [\delta_{\min}, \delta_{\max}] \) denote the inverse of \( x \). Then the contract \( W = [w(x_i, x_{-i}), w(x_{-i}, x_i)] \), with
\[
w(x, y) = R - \mathbb{E}[u_\delta(\delta)] + C(x) + u(h(x)) - h(x) u_\delta(h(x)) + u_\delta(h(y)),
\] (12)
provides each type \( \delta \in [\delta_{\min}, \delta_{\max}] \) with the strictly dominant effort strategy \( x(\delta) \). For the triplet of a principal and two randomly selected, privately informed agents, \( W \) implements interim Pareto improvement upon the first-best efficient status quo; that is, \( \mathbb{E}(\delta, \tau)[\pi_P] > 0 \) and \( \mathbb{E}[u_i|\delta_i] = (1 + \delta_i) R + u(\delta_i) \) for all \( \delta_i \in [\delta_{\min}, \delta_{\max}] \). If \( x \) increases (decreases) in \( \delta \), then \( W \) imposes a positive (negative) externality of every agent’s effort on the other agent’s pay.

With Lemma 3, the specification of an interim-efficient labor contract simplifies to two tasks: the identification of a gain function \( u: [\delta_{\min}, \delta_{\max}] \to [0, \infty) \) satisfying the conditions of Lemma 2(i), and the identification of a continuously differentiable, strictly monotone
function \( x : [\delta_{\min}, \delta_{\max}] \to [x, \overline{x}] \) satisfying \( \mathbb{E}[x(\delta) - C(x(\delta))] > R + \mathbb{E}[u_\delta(\delta)] \). I address the gain function first.

**Lemma 4** Let \( \alpha > 0 \) and \( \delta_* = -1 + \sqrt{\text{Var}[\delta] + (1 + \mathbb{E}[\delta])^2} \). Then \( u : [\delta_{\min}, \delta_{\max}] \to \mathbb{R} \), with \( u(\delta) = \alpha(\delta_* - \delta)^2 \), satisfies the conditions of Lemma 2(i). When embedding \( u \) into the setting of Lemma 3, the contract \( W \) yields the principal an ex ante expected profit of \( \mathbb{E}(\delta, \tau)[\pi_P] = 4\alpha(\delta_* - \mathbb{E}[\delta]) + 2(\mathbb{E}[x - C(x)] - R) \), where \( \delta_{\min} < \mathbb{E}[\delta] < \delta_* < \delta_{\max} \).

Beside the quadratic gain function, which is unique except for the choice of \( \alpha > 0 \), there may be many others. However, with quadratic \( u \), the information requirement is fairly weak: Support, mean, and variance of the distribution of social types need to be common knowledge among the principal and the agents; the distribution itself, or its higher moments, need not be known. In all what follows, I stick with the quadratic gain function, for the weaker the information requirement, the stronger the theory.

Lemma 4 also highlights the role of having different types in the population of agents. Necessarily, \( \text{Var}[\delta] > 0 \). Otherwise, \( \delta_* = \mathbb{E}[\delta] \), in which case \( \mathbb{E}(\delta, \tau)[\pi_P] < 0 \). In particular, a positive variance is already sufficient to render interim Pareto improvement upon the first-best feasible; the only requirement for this is the choice of an effort function \( x \) satisfying \( \mathbb{E}[x - C(x)] > R + \mathbb{E}[u_\delta(\delta)] = R - 2(\delta_* - \mathbb{E}[\delta]) \). While the choice of the gain function is constrained by the availability of information on the distribution of social types, the choice of the effort function is now literally a matter of taste. Any strictly monotone function \( x : [\delta_{\min}, \delta_{\max}] \to [x, \overline{x}] \) satisfying \( \mathbb{E}[x - C(x)] > R - \varepsilon \) for some sufficiently small \( \varepsilon > 0 \) can be chosen. With the gain function fixed, any choice of \( x \) implements the very same interim utilities for the agents. This has two fundamental implications. First, it does not matter whether \( x \) is increasing or decreasing: The interim income effects are the same in either case, even though, by Lemma 3, the labor contract provides the agents with team incentives in the one case and with relative incentives in the other. And second, there is not ‘the’ optimal labor contract: For each \( \varepsilon > 0 \), there are even infinitely many functions \( x : [\delta_{\min}, \delta_{\max}] \to [x, \overline{x}] \) satisfying \( \mathbb{E}[x - C(x)] > R - \varepsilon \), all of which are implementing the very same interim income effects (which are determined by the gain function alone), while they all are shaping the interim-efficient labor contract differently.

4 **Relative and Team Incentives: A Payoff Equivalence**

In this Section, I provide two examples of effort functions satisfying the conditions of Lemma 3. The resulting interim-efficient labor contracts implement relative performance pay in one case and team performance pay in the other. By Lemma 3, the ‘sign’ of the externality
In particular, will not work at all. Obviously, as is shown there, an agent

Though Proposition 5 is a corollary of Lemmas 3 and 4, I prove it explicitly in the Appendix.

ciently. All the other types will exert inefficiently low efforts; the most altruistic type,

That is, the contract incentivizes the least altruistic type, \( \delta_{\text{min}} \), to produce first-best efficiently. All the other types will exert inefficiently low efforts; the most altruistic type, \( \delta_{\text{max}} \), will not work at all. Obviously, \( x \) is continuously differentiable and strictly decreasing in \( \delta \). In particular, \( \lim_{\kappa \to 0} x(\delta) = x^* \) for each \( \delta \in (\delta_{\text{min}}, \delta_{\text{max}}) \). Thus, for any fixed \( \alpha > 0 \), we have \( \mathbb{E}[x - C(x)] > R - 2\alpha (\delta^* - \mathbb{E}[\delta]) \) when letting \( \kappa \) sufficiently small. The inverse of \( x \) is given by \( h : [0, x^*] \to [\delta_{\text{min}}, \delta_{\text{max}}] \), with \( h(x) = \delta_{\text{max}} - \delta \left( x/x^* \right)^{1/2} \). Plugging \( u \) and \( h \) into (12) yields the following form of relative performance pay (RPP).\(^{29}\)

**Proposition 5** Let \( \alpha > 0 \). For \( \kappa \in (0, 1) \), let \( V(x) = 1_{[0,x^*]}(x) [x/x^*]^{1/2} \). Then the contract \( W_{\text{RPP}} = [w_{\text{RPP}}(x, x_i), w_{\text{RPP}}(x_i, x)] \), with

\[
w_{\text{RPP}}(x, y) = S(\alpha) + C(x) + \alpha \delta \left[ 2\delta_{\text{max}} - \delta V(x) \right] V(x) - 2\alpha \delta V(y),
\]

provides each type \( \delta \in [\delta_{\text{min}}, \delta_{\text{max}}] \) with the strictly dominant effort strategy \( x(\delta) \) as given by (14). For the triplet of a principal and two randomly selected, privately informed agents, \( W_{\text{RPP}} \) implements interim Pareto improvement upon the first-best efficient status quo if \( \kappa \) is taken sufficiently small. \( W_{\text{RPP}} \) imposes a negative externality of every agent’s effort on the other agent’s pay.

Notice that \( W_{\text{RPP}} \) involves direct compensation for the individual costs of effort, and that the agents’ limited liability constraint is satisfied when letting \( \alpha > 0 \) sufficiently small. Though Proposition 5 is a corollary of Lemmas 3 and 4, I prove it explicitly in the Appendix. As is shown there, an agent \( i \) of type \( \delta_i \) who ends up with a coworker of type \( \tau \) derives an

\(^{29}\) **Notation:** Let \( A \) and \( B \) two sets, with \( A \subset B \). Then \( 1_A : B \to \{0, 1\} \) denotes the indicator function of \( A \); that is, \( 1_A(x) = 1 \) if \( x \in A \), and \( 1_A(x) = 0 \) if \( x \in B \backslash A \).
ex post utility level of

\[ u_i = (1 + \delta_i) R + u(\delta_i) + \alpha \left[ 2 (\tau - \mathbb{E}[\tau]) - \delta_i \left( \tau^2 - \mathbb{E}[\tau^2] \right) \right]. \]  \hfill (16)

Obviously, \( \partial u_i / \partial \tau = 2\alpha (1 - \delta_i \tau) > 0 \) for all \( \delta_i, \tau \in (-1, 1) \). Every agent would thus benefit from meeting a more altruistic type but suffer from meeting a less altruistic type. If the principal ends up with two types \( \delta \) and \( \tau \), she realizes an ex post profit of

\[ \pi_P = x(\delta) - C(x(\delta)) - R + x(\tau) - C(x(\tau)) - R \\
+ 4\alpha (\delta_* - \mathbb{E}[\delta]) + \alpha \left[ (1 - \delta)^2 + (1 - \tau)^2 - 2\mathbb{E}[(1 - \delta)^2] \right], \]  \hfill (17)

where \( x(\delta) \) and \( x(\tau) \) denote the strictly dominant effort strategies of types \( \delta \) and \( \tau \), given by (14). Obviously, \( \partial \pi_P / \partial \delta < 0 \) for all \( \delta \in (-1, 1) \). The principal would thus benefit from less altruistic types and suffer from more altruistic types. Together, the redistributive effects of \( W_{RPP} \) are those of hawk-dove-principal from Section 2. Competition is hardest between least altruistic types, which benefits \( P \). Most altruistic agents would not work at all and yet receive the salary \( S(\alpha) \), for which \( S(\alpha) \to R \) as \( \alpha \to 0 \); effectively, most altruistic agents would collaborate against the principal. Interestingly, since the type space was taken arbitrary, relative performance pay is interim-efficient even if all types are altruistic, \( \delta_{\text{min}} > 0 \).

Figure 5 gives a stylized illustration of relative performance pay according to Proposition 5. It depicts how an agent \( i \)'s individual effort under \( W_{RPP} \) affects his own wealth, given by the effort internality \( I(x_i) \), and how his effort affects his coworker's wealth, given by the effort externality \( E(x_i) \). According to (15), \( I(x_i) = S(\alpha) + \alpha \mathbb{E}[2\delta_{\text{max}} - \delta V(x_i)] V(x_i), \) and \( E(x_i) = -2\alpha \mathbb{E}[\delta V(x_i)] \), where \( V(x) = 1_{[0,x^*]}(x)[x/x^*]^{\frac{1}{\gamma}} \). For \( \delta_i \in \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\} \), the dashed curves \( u_{\delta_i}(x_i) = I(x_i) + \delta_i E(x_i) \) show how a type \( \delta_i \)'s effort choice directly affects his interim expected utility.\(^{30}\)

For an example of team performance pay, denote by \( \bar{x} \) the unique solution of \( x - C(x) = -2 (\delta_* - \mathbb{E}[\delta]) \). Since \( \delta_* > \mathbb{E}[\delta] \), we have \( \bar{x} \in (x^*, \infty) \). For \( \kappa \in (0, 1) \), consider the effort function \( x: [\delta_{\text{min}}, \delta_{\text{max}}] \to [x^*, \bar{x}] \) with

\[ x(\delta) = \bar{x} - (\bar{x} - x^*) \left( \frac{\delta_{\text{max}} - \delta}{\delta} \right)^{\kappa}. \]  \hfill (18)

Again, the contract incentives the least altruistic type to produce first-best efficiently. All the other types will now exert inefficiently high efforts. Obviously, \( x \) is differentiable and strictly decreasing in \( \delta \), and \( \lim_{\kappa \to 0} x(\delta) = x^* \) for each \( \delta \in (\delta_{\text{min}}, \delta_{\text{max}}) \). Its inverse is

\(^{30}\)Notice that \( \mathbb{E}[u_i(x_i) | \delta_i] = u_{\delta_i}(x_i) + \mathbb{E}[E(x(\tau)) + \delta_i I(x(\tau))] \).
Figure 5: A stylized representation of how an agent’s effort under $W_{RPP}$ affects his own wealth, given by the effort internality $I(x_i)$, and how it affects his coworker’s wealth, given by the effort externality $E(x_i)$. For $\delta_i \in \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$, the dashed curves $u_i(x_i) = I(x_i) + \delta_i E(x_i)$ show how a type $\delta_i$’s effort choice directly affects his expected utility; maximum utility is achieved at the dominant strategy effort level $x(\delta_i)$, given by (14).

$h : [0, x^*] \to [\delta_{\min}, \delta_{\max}]$, with $h(x) = \delta_{\max} - \delta \left[ (\bar{x} - x) / (\bar{x} - x^*) \right]^{\frac{1}{2}}$. Plugging $u$ and $h$ into (12) yields the following form of team performance pay (TPP).

**Proposition 6** Let $\alpha > 0$. For $\kappa \in (0, 1)$, let $V(x) = 1_{[x^*, \bar{x}]}(x) \left[ (\bar{x} - x) / (\bar{x} - x^*) \right]^{\frac{1}{2}}$. Then the contract $W_{TPP} = \left[ w_{TPP}(x_i, x_{-i}), w_{TPP}(x_{-i}, x_i) \right]$ with

$$w_{TPP}(x, y) = S(\alpha) + C(x) + \alpha \delta \left[ 2\delta_{\max} - \delta V(x) \right] V(x) - 2\alpha \delta V(y)$$

(19)

provides each type $\delta \in [\delta_{\min}, \delta_{\max}]$ with the strictly dominant effort strategy $x(\delta)$ as given by (18). For the triplet of a principal and two randomly selected, privately informed agents, $W_{TPP}$ implements interim Pareto improvement upon the first-best efficient status quo if $\kappa$ is taken sufficiently small. $W_{TPP}$ imposes a positive externality of every agent’s effort on the other agent’s pay.

The proof of Proposition 6 is put in the Appendix. As shown there, the ex post utilities

\[\text{as with relative performance pay, } W_{TPP} \text{ involves direct compensation for the individual costs of effort, and the agents’ limited liability constraint is satisfied when letting } \alpha > 0 \text{ sufficiently small.}\]
and profits associated with $W_{TPP}$ are exactly those of $W_{RPP}$, and given by (16) and (17). That is, $W_{RPP}$ and $W_{TPP}$ are even ex post payoff equivalent. Consequently, also the redistributive effects of $W_{TPP}$ are those of hawk-dove-principal. At first sight, it might come as a surprise that the principal suffers from more altruistic types when providing the agents with interim-efficient team incentives: Under complete information, altruists would form the most efficient teams when provided with team incentives. Under asymmetric information on social preferences, moral hazard in teams can be overcome by letting an agent’s marginal return from his individual effort so high that the least altruistic type, $\delta_{\text{min}}$, produces first-best efficiently. The more altruistic types then exert inefficiently high efforts, which is to the principal’s disadvantage. Since the type space was taken arbitrary, team performance pay is interim-efficient even if all types are spiteful, $\delta_{\text{max}} < 0$.

Figure 6 gives a stylized illustration of team performance pay according to Proposition 6. Similar to Figure 5, it depicts how an agent $i$’s individual effort under $W_{TPP}$ affects his own wealth, given by the effort internality $I(x_i)$, how his effort affects his coworker’s wealth, given by the effort externality $E(x_i)$, and, for $\delta_i \in \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$, how $i$’s effort choice directly affects his interim expected utility if his type is $\delta_i$. According to (19), $I(x_i) = S(\alpha) + \alpha \tilde{\delta} \left[2\delta_{\text{max}} - 3V(x_i)\right] V(x_i)$ and $E(x_i) = -2\alpha \tilde{\delta} V(x_i)$, where now $V(x) = 1_{[x^*, x]} \left(\frac{(x - x^*)}{(x - x^*)}\right)^{\frac{\tilde{\delta}}{2}}$.

Generally, the interim-efficient labor contracts of Propositions 5 and 6 each specify a salary, an individual performance component, and an externality component implying either relative or team incentives. Due to limited liability, externalities are muted (since $\alpha > 0$ must be taken sufficiently small), implying wage compression. The following Proposition is immediate from Lemma 1 and Propositions 5 and 6.

**Proposition 7** Suppose there is perfect competition in the labor market. Then there is a Perfect Bayesian equilibrium in which a principal and two randomly selected, privately informed agents deviate unanimously from labor market equilibrium by contracting according to a single-period, interim-efficient labor contract. For the contracting parties, the interim-efficient labor contract implements interim Pareto improvement upon labor market equilibrium. It implements either relative or team performance pay, both of which are ex interim (and potentially ex post) payoff equivalent.

---

32 A caveat to this conclusion is my simplifying assumption of substitutable and observable individual efforts. Moral hazard in teams arises naturally when only team performance is observable (Holmstrom 1982). The question how hawk-dove-principal could be implemented in the presence of observability constraints is beyond the scope of this study.

33 Many other authors have rationalized wage compression on the grounds of agents’ social concerns; e.g., Frank (1984a, 1984b), Bartling and von Siemens (2010), Engmaier and Wambach (2010), and Stark and Hyll (2011). Lazear (1989) argues that wage compression alleviates sabotage in contests.
Figure 6: A stylized representation of how an agent’s effort under WTP affects his own wealth, given by the effort internality $I(x_i)$, and how it affects his coworker’s wealth, given by the effort externality $E(x_i)$. For $\delta_i \in \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$, the dashed curves $u_{\delta_i}(x_i) = I(x_i) + \delta_i E(x_i)$ show how a type $\delta_i$’s effort choice directly affects his utility; maximum utility is achieved at the dominant strategy effort level $x(\delta_i)$, given by (18).

The payoff equivalence result suggests an explanation for the parallel use of relative and team incentives in nearly any white-collar workplace. White collar work often involves the execution of multiple tasks. If work involves the execution of $N$ independent tasks (all satisfying the model assumptions of Section 3.1), then any $n \in \{0, ..., N\}$ of them can be incentivized with relative performance pay and the remaining $N - n$ with team performance pay. In particular, piece rate compensation is never interim-efficient, even though individual efforts were assumed to be fully contractible. It seems to be a promising direction for future research to understand how (if at all) these findings generalize to tasks which are constrained with respect to the observability or complementarity of individual efforts.

5 Labor Turnover for Social Reasons

Having identified labor contracts that implement interim Pareto improvement upon the first-best efficient status, it is straightforward to extend the finding of Proposition 7 and to provide a rationale for the empirical observation of labor turnover for social reasons.
In the perfectly competitive labor market outlined in Section 3.1, a continuum of firms competes for a continuum of infinitely-lived agents. Firms and agents subsequently contract for a single period \( t \in \mathbb{N} \), and each firm that is active in \( t \) hires exactly two agents. While it is common knowledge among agents and firms that the agent’s social types are distributed according to the c.d.f. \( F \) as specified in Section 3.1., each agent \( i \) knows the realization of his own type \( \delta_i \), which is determined once and for all before period \( t = 1 \). In the following, the key behavioral assumption is that an agent’s social preferences do only relate to his actual coworker per period. Agents do not internalize the effects that a change of their affiliation would have on their current coworker, and they choose their affiliation regardless of the effects on those who stand behind the options not taken.

Contracting between a firm, represented by a principal \( P \), and two agents \( i \) and \(-i\) requires unanimous agreement among the triplet \([P, i, -i] \). If, in some period \( t \), a principal does not reach unanimous agreement with two agents, then she does not operate in \( t \) and realizes zero profit; in this case, the agents are randomly (at zero cost) assigned to some other firm. After having contracted in period \( t \), the triplet \([P, i, -i] \) has the option of renegotiation in order to contract again in period \( t + 1 \), requiring again unanimous agreement.\(^{34}\)

When changing his affiliation, an agent’s type is assumed to be his private information. In particular, once having changed his affiliation, the agent will never encounter his former employer or coworker again (alternatively, this event occurs with likelihood zero).

By assumption, there exists a rematching technology with the following property: If all contracting triplets \([P, i, -i] \) of period \( t \) decide not to contract again in period \( t + 1 \), and if all principals offer one and the same contract \( W = [w_i (x_i, x_{-i}), w_{-i} (x_{-i}, x_i)] = [w (x_i, x_{-i}), w (x_{-i}, x_i)] \) in period \( t + 1 \), then all the agents are randomly matched in pairs and, as pairs, allocated among firms. The rematching technology ensures that each pair of agents consists of two independent random draws according to the c.d.f \( F \) of social types in the population of agents, such that \((\delta_i, \delta_{-i}) \sim F \times F\) for each pair of agents. Therefore, with an uncountably infinite number of agents and a countably infinite number of periods, the rematching technology ensures that agents reencounter former employers or coworkers with likelihood zero. Intuitively, in competitive labor markets, a working life is not enough to encounter a substantial share of all the employers and potential coworkers who are around.

In each period \( t \), agent \( i \) seeks to maximize \( U_t = \sum_{s=0}^{\infty} \rho_t^s E [u_i^{t+s}] \), with \( u_i^{t+s} = \pi_i^{t+s} + \delta_i \pi_{-i}^{t+s} \), where \( \pi_i^{t+s} \) and \( \pi_{-i}^{t+s} \) denote the wealth of coworkers \( i \) and \(-i\), and where \( \rho_t \in (0, 1) \) denotes some individual discount rate.\(^{35}\) A principal \( P \) seeks to maximize the present value

\(^{34}\)In labor market equilibrium, each contracting triplet \([P, i, -i] \) would unanimously agree upon the same contract over and over again.

\(^{35}\)For the moment, the symbol \( E \) is simply used to indicate that expectations are taken with respect to the beliefs that the economic agent has formed about all relevant uncertainties.
of her future profits, $\Pi'_P = \sum_{s=0}^{\infty} \rho^s_P \mathbb{E}[\pi^{t+s}_P]$, where $\mathbb{E}[\pi^{t+s}_P]$ denotes $P$’s ex ante expected profit in period $t + s$, and where $\rho_P \in (0, 1)$ denotes some discount rate. With free entry for firms, $\Pi'_P = 0$ for all principals $P$ in each period $t$. All this is common knowledge among principals and agents.

Consider the contract $W^* = [w^*(x_i, x_{-i}), w^*(x_{-i}, x_i)]$, with

$$w^*(x, y) = w_{RPP}(x, y) + S,$$

where $w_{RPP}(x, y)$ is the wage incentive scheme from Proposition 5, and where $S$ equals half of a principal’s ex ante expected profit under $W_{RPP}$ from Proposition 5. With the strictly dominant effort strategies $x : [\delta_{\min}, \delta_{\max}] \to [0, x^*]$ implemented by $W_{RPP}$, as given by (14), Lemma 4 implies $S = 2\alpha (\delta_{\min} - \mathbb{E}[\delta]) + \mathbb{E}[x - C(x)] - R$. Consistent with perfect labor market competition, the principal expects zero profits from interim-efficient contracting according to $W^*$.

**Proposition 8** In a perfectly competitive labor market, the following can be established as the outcome of a Perfect Bayesian equilibrium: In each period, each active firm hires a pair of two randomly matched agents and provides them with the single-period, interim-efficient labor contract $W^*$ from (20); alternatively, the contracts offered across firms are ex interim payoff-equivalent to $W^*$. After any one period of contracting, all agents are released and, as pairs of randomly matched agents, reallocated between firms.

Notice that labor market equilibrium as characterized by Lemma 1 is also the outcome of a subgame perfect equilibrium in this sequential game: If agent $-i$ follows the strategy to reject $W$, then agent $i$ is indifferent between accepting and rejecting $W$, such that unanimous rejection of $W$ in each period is equilibrium behavior as well.

The Perfect Bayesian equilibrium of Proposition 8 implies that staffing in each firm and period is a gamble over the composition of social types at work, effectively implementing hawk-dove-principal from Section 2. Intuitively, each agent suffers from ending up with a hawkish careerist, in which case he gives it another try in a different firm. A principal suffers from ending up with dovish team players, who effectively collaborate against her, in which case she gives it another try by firing and hiring. This finding rationalizes labor turnover for social reasons.

Certainly, labor turnover requires a minimum level of labor market competition. Workers need some freedom of choice in order to change their affiliation. However, it seems intuitive that the findings are robust to sufficiently small search costs.

---

36The parameter $\alpha > 0$ is taken as the maximum value that still satisfies the limited liability constraint: $w^*(x_i, x_{-i}) \geq 0$ for all $x_i, x_{-i} \geq 0$. 
6 Job Talks: A Matter of Mutual Trust

Interim-efficient contracting effectively implements hawk-dove-principal from Section 2. The agents, regardless of their types, benefit from most altruistic coworkers, but suffer from least altruistic coworkers, and vice versa for principals. This raises the question whether the principal can utilize bilateral job talks in order to guarantee herself non-negative ex post profit.\footnote{37 Agent-to-agent communication is beyond the scope of this paper. I assume that the principal can ensure that job applicants cannot communicate before contracting.}

In the market environment of Section 5, fix a period $t$ and consider a triplet $[P, i, -i]$, with a principal $P$ and agents $i$ and $-i$ whose types are their private information. $P$ has the option to either implement an interim-efficient contract $W^*$ with agents $i$ and $-i$, or to be inactive in period $t$. If implementation of $W^*$ does not find unanimous approval among $[P, i, -i]$, then $i$ and $-i$ are each immediately (before production in $t$ takes place) and randomly rematched with other agents at different firms, where, for simplicity, the same contract $W^*$ (or an ex interim payoff equivalent contract) is implemented without job talks. Rematching is assumed to impose an arbitrarily small utility loss on the agents, such that agents prefer to contract ‘early’ in each period. For simplicity, these costs are set to zero. Before decisions upon approving or rejecting $W^*$ are made, the agents’ types are their private information. Again, agent $i$’s belief about $-i$’s type is $\delta_{-i} \sim F$; agent $-i$ believes $\delta_i \sim F$; and $P$ believes $(\delta_i, \delta_{-i}) \sim F \times F$.

The principal deploys a communication mechanism that specifies message spaces $M_i, M_{-i}$ and a strategic game in which the agents simultaneously choose their messages to the principal. Communication is direct: Each privately informed agent reports some social type, $M_i = [\delta_{\min}, \delta_{\max}] = M_{-i}$, and chooses a message according to a reporting strategy $\mu_i(\delta_i), \mu_{-i}(\delta_{-i}) : [\delta_{\min}, \delta_{\max}] \to \mathcal{M}$, where $\mathcal{M}$ denotes the set of probability distributions over the type space, $[\delta_{\min}, \delta_{\max}]$; that is, each type reports some social type according to a (potentially degenerate) mixed strategy over the type space $[\delta_{\min}, \delta_{\max}]$.

After having received the agents’ messages, $P$ decides on whether or not to agree to contract according to $W^*$. Based on the received messages $m = (m_i, m_{-i})$, she forms a belief about the composition of types, $p(m)$, and chooses $y \in \{0, 1\}$, where $y = 0$ ($y = 1$) indicates rejection (approval). Since $P$ seeks to guarantee herself a non-negative profit, while allowing for positive profits when both agents are sufficiently hawkish, the mechanism can be assumed to incentivize each type to agree to contract. With imperfect commitment, $P$’s equilibrium strategy must then satisfy: $y(m) \in \arg \max y \int_{[\delta_{\min}, \delta_{\max}]} p((\delta, \tau)|m) d(\delta, \tau)$.\footnote{38 Furthermore, in Perfect Bayesian equilibrium, the principal’s belief must be consistent with Bayesian updating for all pairs of messages that occur in equilibrium with positive probability.}
The following Proposition shows that the classical revelation principle\textsuperscript{39}, which also applies to situations in which the asymmetry of information is associated with the agents’ social preferences (see Myerson 1979), does not generally apply if the principal cannot fully commit herself to a specific allocation function. This insight was generated first by Bester and Strausz (2000).\textsuperscript{40} To the best of my knowledge, the existing studies that reject the revelation principle on the grounds of imperfect commitment of the principal build upon the ‘direct mechanism’-aspect: Interim superior outcomes can potentially be implemented when expanding message spaces beyond type spaces. The following result builds on direct mechanisms but rejects the revelation principle in its ‘truthful revelation’-aspect. Recall that $W^*$ from (20) implements interim Pareto improvement upon an ex post first-best efficient allocation.\textsuperscript{41}

**Proposition 9** With imperfect commitment of the principal, there is no direct mechanism that incentivizes each agent to truthfully reveal his social type, while implementing unanimous approval of the interim-efficient labor contract $W^*$ from (20).

In the setting studied here, imperfect commitment of the principal implies that job talks are cheap talk. The principal fears most altruistic agents because coworker altruism facilitates collaboration against her. If she would simply ask the agents about their social types, then dovish team players would impersonate hawkish careerists. From the principal’s point of view, they would come along as “wolfs in sheep’s clothing”, as “dove-feathered ravens”. That the principal and the agents agree nevertheless to an interim efficient labor contract, implying that they all expose themselves to the threat of ending up with a disliked type, appears like mutual, incentive-compatible trust.

In a survey on the economics and biology of trust, Fehr (2009) points out that the growing academic literature on the matter is still lacking a conclusive theoretical understanding of what trust actually is. The present study suggests the following notion:

*In environments of asymmetric information on people’s social (altruistic or spiteful) preferences, trust is the (mutual) act of voluntary participation in a Bayesian game that implements interim Pareto improvement upon ex post efficient allocations.*

\textsuperscript{39}The following formulation is borrowed from Bester and Strausz (2000, p.167). **Revelation principle:** Suppose that $\{M, v\}$ supports a Bayesian equilibrium in which the principle has the ex ante expected payoff $V^*$ and agent $i$ of type $t_i$ has the interim expected payoff $U^*_i(t_i)$. Then there exists a direct mechanism $\{T, \tilde{v}\}$ which supports a Bayesian equilibrium with payoffs $\tilde{V}$ and $\tilde{U}_i(t_i)$ such that $\tilde{V} = V^*$ and $\tilde{U}_i(t_i) = U^*_i(t_i)$ for all $i$ and $t_i$. Moreover, in the Bayesian equilibrium under the direct mechanism each agent $i$ reports his type $t_i$ truthfully, i.e. $\mu_i(t_i|t_i) = 1$ for all $i$ and $t_i$.

\textsuperscript{40}Evans and Reiche (2008) present a similar result. See Bester and Strausz (2001) for the single agent case. Within the scope of this study, the consideration of general communication devices (Bester and Strausz 2007), which allow for noisy communication, is far beyond reach.

\textsuperscript{41}Proposition 9 can also be established in the framework of hawk-dove-principal from Section 2; the proof then parallels the one for interim-efficient contracting.
This study shows that the so defined *trust environments* are not an empty set. In particular, this notion of trust is independent of the specific distribution of social types; the necessary and sufficient assumption about the distribution of social types to render interim Pareto improvement upon ex post efficient allocations possible is that its variance be positive, while its support and density can be taken arbitrary. For a coherent notion of trust, the feasibility of interim Pareto improvement upon ex post efficient allocations seems to be important, since otherwise adverse selection is open for attack through screening, sorting, and signaling, potentially resolving the asymmetry of information and, hence, the necessity of trust. According to the above notion, trust is a form of risk-taking. This ultimately raises the question whether the findings of this study are robust to risk aversion. In the Appendix, I reestablish the game of hawk-dove-principal from Section 2 for risk-averse agents. I show that interim efficient Pareto improvement upon ex post efficient allocations can be implemented for moderate degrees of constant relative risk aversion.\(^{42}\) In so far, all the findings of this study are likely to be robust to moderate risk aversion. Finally, it seems worthwhile to investigate whether there are forms of social preferences other than altruism and spite that do allow for interim Pareto improvement upon ex post efficient allocations.

7 Concluding Remarks

Modern work life is strongly marked by competition and collaboration and, thus, by employees’ willingness to compete and to cooperate. It has brought about workplace creatures such as ‘careerists’ and ‘team players’. Competition is as tough, and collaboration as poor, as careerists are numerous within the same occupation. However, employees and employers rarely know beforehand whether entrants on a job will emerge as hawkish careerists or dovish team players. Similarly, entrants on a job rarely know beforehand whether their future colleagues will turn out to be supportive. These impressions are in line with the empirical observation that the perceived support from coworkers is a key factor for job satisfaction and actual labor turnover.

With this study I have rationalized labor turnover for social reasons on the grounds of asymmetric information on employees’ social preferences. Single-period, interim-efficient contracting allows for interim Pareto improvement upon states that are ex post efficient, and therefore provides all firms and agents with the incentive to deviate unanimously from equilibrium in a perfectly competitive labor market. With interim-efficient contracting, staffing is effectively a gamble, for the principal as well as for the agents, over the composition of social types.\(^{42}\) The here proposed notion of trust is thus consistent with the empirical finding that “trust decisions are not tightly connected to a person’s risk attitudes” (Houser et al. 2010).

\(^{42}\)The here proposed notion of trust is thus consistent with the empirical finding that “trust decisions are not tightly connected to a person’s risk attitudes” (Houser et al. 2010).
social types at work. This finding indicates that corporate culture has a random component associated with asymmetric information on employees’ social preferences. Every new colleague might turn the social environment at work into the better or the worse. Based on these findings, I have proposed an incentive theoretical notion of trust.

Generally, when met with interim-efficient incentives, the effects of asymmetric information on people’s social preferences are ex interim mutually beneficial rather than adverse. Interim-efficient incentives may in fact lead people to deviate unanimously from ex post efficient allocations. Social preferences seem to involve their own economics of information.
References


A Appendix

Proof of Lemma 1. Every agent’s marginal utility from own wealth, 1, exceeds his marginal (dis)utility from his coworker’s wealth, \( \delta_i \in [\delta_{\text{min}}, \delta_{\text{max}}] \subset (-1,1) \), and \( P \) is a pure payoff maximizer. Thus, for given returns from total output, any allocation of these returns among the agents and \( P \) must be Pareto efficient in equilibrium. To identify the first-best efficient total output, consider a contract equivalent to self-employment: every agent \( i \) receives the returns from his individual output, and \( P \) makes zero profits. Then \( i \) maximizes

\[ u_i = x_i - C(x_i) + \delta_i [x_{-i} - C(x_{-i})] \]

by exerting the effort \( x^* = C_x^{-1}(1) \), which yields him a utility level of \( u_i = (1 + \delta_i)R + v_i \), where \( R = x^* - C(x^*) \). Now suppose there is a contract \( W = [w_i(x_i, x_{-i}), w_{-i}(x_{-i}, x_i)] \) that Pareto improves ex post upon self-employment, with \( P \) realizing a positive profit. Let \( x_i \) and \( x_{-i} \) denote the (not necessarily unique) equilibrium effort choices of \( i \) and \( -i \) under \( W \), and let \( w_i \) and \( w_{-i} \) denote the respective wage payments of agents \( i \) and \( -i \). Pareto improvement requires that there exist \( v_i, v_{-i} \geq 0 \) such that both

\[
\begin{align*}
    u_i &= [w_i - C(x_i)] + \delta_i [w_{-i} - C(x_{-i})] = (1 + \delta_i)R + v_i, \\
    u_{-i} &= [w_{-i} - C(x_{-i})] + \delta_{-i} [w_i - C(x_i)] = (1 + \delta_{-i})R + v_{-i}.
\end{align*}
\]

Let \( \mathbf{w} = (w_i, w_{-i})^T \), \( \mathbf{x} = (x_i, x_{-i})^T \), \( \mathbf{C} = (C(x_i), C(x_{-i}))^T \), \( \mathbf{v} = (v_i, v_{-i})^T \), \( \mathbf{1} = (1,1)^T \), and

\[
\Delta = \begin{pmatrix} 1 & \delta_i \\ \delta_{-i} & 1 \end{pmatrix}.
\]

By (21) and (22), \( \Delta (\mathbf{w} - \mathbf{C}) = R \mathbf{1} + \mathbf{v} \), where \( \Delta \) is invertible. Thus, \( \mathbf{w} = \mathbf{C} + R \mathbf{1} + \Delta^{-1} \mathbf{v} \), and \( P \)'s profit \( \pi_P = \mathbf{1}^T \mathbf{x} - \mathbf{1}^T \mathbf{w} \) must satisfy

\[
0 < \mathbf{1}^T \mathbf{x} - \mathbf{1}^T \mathbf{w} = -\Delta^{-1} \mathbf{v} + \mathbf{1}^T (\mathbf{x} - \mathbf{C} - R \mathbf{1})
\]

\[
= -\mathbf{1}^T \Delta^{-1} \mathbf{v} + \sum_{k=i,-i} \left[ [x_k - C(x_k)] - [x^* - C(x^*)] \right].
\]

Since \( x - C(x) \) is maximal at \( x^* \), the sum \( \mathbf{1}^T (\mathbf{x} - \mathbf{C} - R \mathbf{1}) \) is not positive. The remainder \( -\mathbf{1}^T \Delta^{-1} \mathbf{v} \) is not positive either, since \( v_i, v_{-i} \geq 0; \delta_i, \delta_{-i} \in (-1,1) \); and

\[
\mathbf{1}^T \Delta^{-1} \mathbf{v} = \frac{1}{1 - \delta_i \delta_{-i}} \begin{pmatrix} 1 & \delta_i \\ -\delta_{-i} & 1 \end{pmatrix} \begin{pmatrix} v_i \\ v_{-i} \end{pmatrix} = \frac{1}{1 - \delta_i \delta_{-i}} [(1 - \delta_{-i}) v_i + (1 - \delta_i) v_{-i}] \geq 0.
\]

\( \Delta \) is invertible. Thus, \( \mathbf{w} = \mathbf{C} + R \mathbf{1} + \Delta^{-1} \mathbf{v} \), and \( P \)'s profit \( \pi_P = \mathbf{1}^T \mathbf{x} - \mathbf{1}^T \mathbf{w} \) must satisfy

\[
\begin{align*}
    0 &< \mathbf{1}^T \mathbf{x} - \mathbf{1}^T \mathbf{w} \\
    &= -\Delta^{-1} \mathbf{v} + \mathbf{1}^T (\mathbf{x} - \mathbf{C} - R \mathbf{1}) \\
    &= -\mathbf{1}^T \Delta^{-1} \mathbf{v} + \sum_{k=i,-i} \left[ [x_k - C(x_k)] - [x^* - C(x^*)] \right].
\end{align*}
\]

Since \( x - C(x) \) is maximal at \( x^* \), the sum \( \mathbf{1}^T (\mathbf{x} - \mathbf{C} - R \mathbf{1}) \) is not positive. The remainder \( -\mathbf{1}^T \Delta^{-1} \mathbf{v} \) is not positive either, since \( v_i, v_{-i} \geq 0; \delta_i, \delta_{-i} \in (-1,1) \); and

\[
\begin{align*}
    \mathbf{1}^T \Delta^{-1} \mathbf{v} &= \frac{1}{1 - \delta_i \delta_{-i}} \begin{pmatrix} 1 & \delta_i \\ -\delta_{-i} & 1 \end{pmatrix} \begin{pmatrix} v_i \\ v_{-i} \end{pmatrix} \\
    &= \frac{1}{1 - \delta_i \delta_{-i}} [(1 - \delta_{-i}) v_i + (1 - \delta_i) v_{-i}] \geq 0.
\end{align*}
\]
Thus, $0 < \pi_P \leq 0$; a contradiction. This proves that the Pareto frontier of contracting is characterized by having each agent exert the first-best efficient effort level $x^*$. Due to perfect labor market competition, each contract $W$ that is applied by some firm in labor market equilibrium must be an element of the Pareto frontier. Hence, in labor market equilibrium, production in each firm is first-best efficient; that is, efforts in each firm are $x_i = x^* = x_{-i}$.

With free entry for firms, equilibrium requires $\pi_P = 0$ for each principal $P$.

Now suppose there are at least two agents, $i$ and $i'$, who earn less than the returns from their individual outputs, $x^*$, in labor market equilibrium. Then there is a principal $P$ hiring agents $i$ and $-i$, with $i$ earning $x^* - \varepsilon$, and $-i$ earning $x^* + \varepsilon$, for some $\varepsilon > 0$. According to (1), agent $i$'s utility from production is thus $u_i = (1 + \delta_i) (R - (1 - \delta_i) \varepsilon < u_{\delta_i}$. Similarly, there is another principal $P'$ hiring agents $i'$ and $-i'$, with $i'$ earning $x^* - \varepsilon'$, and $-i'$ earning $x^* + \varepsilon'$, for some $\varepsilon' > 0$. According to (1), agent $i'$ realizes utility $u_{i'} = (1 + \delta_{i'}) (R - (1 - \delta_{i'}) \varepsilon' < u_{\delta_{i'}}$. But then, a principal $P''$ can enter the market and realize a positive profit by offering both agents $i$ and $i'$ the following contract: For some $\eta > 0$ sufficiently small, $P''$ pays $i$ the wage $w^i(x_i) = x_i - \eta$ if $x_i > \eta$, and $w^i(x_i) = 0$ otherwise. (Notice that nonnegative wage payments are required due to the agents’ limited liability.) Similarly, $P''$ pays $i'$ the wage $w^{i'}(x_{i'}) = x_{i'} - \eta$ if $x_{i'} > \eta$, and $w^{i'}(x_{i'}) = 0$ otherwise. Then both agents exert the effort $x^*$. Agent $i$ realizes utility $u_i = (1 + \delta_i) (R - \eta)$, for which $(1 + \delta_i) (R - \eta) > (1 + \delta_i) (R - (1 - \delta_i) \varepsilon)$ if $\eta > 0$ is chosen sufficiently small. Similarly, $u_{i'} = (1 + \delta_{i'}) (R - \eta) > (1 + \delta_{i'}) (R - (1 - \delta_{i'}) \varepsilon')$ for sufficiently small $\eta$. Notice that a choice of $\eta$ satisfying $\eta \leq \frac{1 - \delta_{\max}}{1 + \delta_{\max}} \min \{\varepsilon, \varepsilon'\}$ ensures that both agents $i$ and $i'$ are better off under the contract offered by $P''$, regardless of the exact values of $\delta_i, \delta_{i'} \in [\delta_{\min}, \delta_{\max}]$; that is, the contract offered by $P''$ need not condition on the agents’ types, such that it does not matter whether the agents’ types are their private information before contracting or whether they are common knowledge. In addition, nonnegative wages require $\eta < x^*$. Thus, $i$ and $i'$ accept the offer, and $P''$ realizes a profit of $\pi_{P''} = 2\eta > 0$, which contradicts the assumption of labor market equilibrium. Therefore, in labor market equilibrium, each agent earns $x^*$. 

**Proof of Lemma 2.** *Ad (i):* By assumption, $W$ implements a strictly dominant effort strategy $x(\tau)$ for each type $\tau \in [\delta_{\min}, \delta_{\max}]$. From implementation of $W$ with a randomly selected type $\tau \sim F$, an agent $i$ of type $\delta_i$ expects utility

$$
\mathbb{E} [u_i | \delta_i] = \mathbb{E} [w(x(\delta_i), x(\tau)) | \delta_i] - C(x(\delta_i)) + \delta_i \mathbb{E} [w(x(\tau), x(\delta_i)) - C(x(\tau)) | \delta_i]. \quad (26)
$$

Since $w(x_i, x_{-i})$ is twice partially continuously differentiable in $x_i$, $\mathbb{E} [u_i | \delta_i]$ is twice continuously differentiable in $x_i = x(\delta_i)$; in particular, Leibniz’ rule for differentiation under the integral sign is applicable. Since $w(x_i, x_{-i})$ provides each type $\delta_i \in [\delta_{\min}, \delta_{\max}]$ with a strictly
dominant effort strategy \( x_i = x(\delta_i) \), the \( \text{FOC} \frac{d}{dx_i} \mathbb{E}[u_i | \delta_i] = 0 \) and the \( \text{SOC} \frac{d^2}{dx_i^2} \mathbb{E}[u_i | \delta_i] < 0 \) must be satisfied for type \( \delta_i \) at \( x_i = x(\delta_i) \). That is,

\[
C_x(x(\delta_i)) = \mathbb{E}[w_1(x(\delta_i), x(\tau)) + \delta_i w_2(x(\tau), x(\delta_i)) | \delta_i],
\]

\[
C_{xx}(x(\delta_i)) > \mathbb{E}[w_{11}(x(\delta_i), x(\tau)) + \delta_i w_{22}(x(\tau), x(\delta_i)) | \delta_i].
\]  

(27) (28)

By assumption, \( W \) is such that \( x_i(\delta_i) \) is continuously differentiable and strictly monotone in \( \delta_i \); that is, \( \frac{dx}{d\delta_i} \neq 0 \). Differentiating the \( \text{FOC} \) (27) with respect to \( \delta_i \) yields

\[
C_{xx}(x(\delta_i)) \frac{dx}{d\delta_i} = \mathbb{E}[w_2(x(\tau), x(\delta_i)) | \delta_i]
\]

\[
+ \frac{dx}{d\delta_i} \mathbb{E}[w_{11}(x(\delta_i), x(\tau)) + \delta_i w_{22}(x(\tau), x(\delta_i)) | \delta_i].
\]  

(29)

The \( \text{SOC} \) (28) implies that \( \frac{d}{d\delta_i} \) and \( \mathbb{E}[w_2(x(\tau), x(\delta_i)) | \delta_i] \) have the same sign. By assumption, \( \mathbb{E}[u_i | \delta_i] = (1 + \delta_i) R + u_0(\delta_i) \), where \( \mathbb{E}[u_i | \delta_i] \) is differentiable in \( \delta_i \). Hence, also \( u_0(\delta_i) \) is differentiable in \( \delta_i \). Differentiating (26) with respect to \( \delta_i = \delta \) yields

\[
R + u_\delta(\delta) = \left. \frac{d}{d\delta_i} \mathbb{E}[u_i | \delta_i] \right|_{\delta_i = \delta}
\]

\[
= \left. \frac{d}{dx} \mathbb{E}[u_i | \delta_i] \right|_{\delta_i = \delta} \cdot \left. \frac{dx}{d\delta_i} \right|_{\delta_i = \delta} + \mathbb{E}[w(x(\tau), x(\delta)) - C(x(\tau)) | \delta]
\]

\[
= \mathbb{E}[w(x(\tau), x(\delta)) - C(x(\tau)) | \delta].
\]  

(30)

Since \( w(x_i, x_{-i}) \) is partially differentiable in \( x_{-i} \), Leibniz’ rule implies that \( u_\delta \) is differentiable in \( \delta \). Differentiating (30) with respect to \( \delta \) yields

\[
u_{\delta\delta}(\delta) = \left. \frac{d^2}{d\delta_i^2} \mathbb{E}[u_i | \delta_i] \right|_{\delta_i = \delta}
\]

\[
= \mathbb{E}[w_2(x(\tau), x(\delta)) - C(x(\tau)) | \delta] \cdot \left. \frac{dx}{d\delta_i} \right|_{\delta_i = \delta}.
\]  

(31)

Since \( \frac{dx}{d\delta_i} \) and \( \mathbb{E}[w_2(x(\tau), x(\delta_i)) | \delta_i] \) have the same sign, and \( \frac{dx}{d\delta_i} \neq 0 \) by assumption, (31)
implies \( u_{\delta\delta}(\delta) > 0 \). Furthermore, taking expectations over \( \delta \sim F \) in (30) yields

\[
R + \mathbb{E}[u_{\delta}(\delta)] = \mathbb{E}[\mathbb{E}[w(x(\tau), x(\delta)) - C(x(\tau)) | \delta]] \\
= \mathbb{E}[\mathbb{E}[w(x(\tau), x(\delta)) - x(\delta) + x(\delta) - C(x(\tau)) | \delta]] \\
= -\frac{1}{2}\mathbb{E}_{(\delta, \tau)}[\pi_P] + \mathbb{E}[x - C(x)]
\]

(32)

Thus, \( \frac{1}{2}\mathbb{E}_{(\delta, \tau)}[\pi_P] = -\mathbb{E}[u_{\delta}(\delta)] + \mathbb{E}[x - C(x)] - R \). On the one hand, \( \mathbb{E}[x - C(x)] < R = C_x^{-1}(1) - C(C_x^{-1}(1)) \); otherwise, each type would produce first-best efficiently almost surely, eliminating any effects of asymmetric information on the agents' social preferences. On the other hand, \( \mathbb{E}_{(\delta, \tau)}[\pi_P] > 0 \) by assumption. Thus, \( \mathbb{E}[u_{\delta}(\delta)] < 0 \). Finally, identity (30) yields

\[
(1 + \delta) R + u(\delta) = \mathbb{E}[u_i|\delta] \\
= \mathbb{E}[w(x(\delta), x(\tau)) - C(x(\delta)) | \delta] + \delta \mathbb{E}[w(x(\tau), x(\delta_i)) - C(x(\tau)) | \delta] \\
= \mathbb{E}[w(x(\delta), x(\tau)) - C(x(\delta)) | \delta] + \delta R + \delta u_{\delta}(\delta).
\]

(33)

Thus, \( \mathbb{E}[w(x(\delta_i), x(\tau)) - C(x(\delta_i)) | \delta_i] = \mathbb{E}[\pi_i|\delta_i] = R + u(\delta_i) - \delta_i u_{\delta}(\delta_i) \). Taking expectations over \( \delta \sim F \) yields \( \mathbb{E}[\mathbb{E}[w(x(\tau), x(\delta)) - C(x(\tau)) | \delta]] = R + \mathbb{E}[u(\delta)] - \mathbb{E}[u_{\delta}(\delta)] \). Substituting the latter into the first line of (32) yields \( \mathbb{E}[u(\delta)] = \mathbb{E}[(1 + \delta) u_{\delta}(\delta)] \). This completes the proof of Lemma 2(ii).


Proof of Lemma 3. From implementation of \( W \) with a randomly selected type \( \tau \sim F \), an agent \( i \) of type \( \delta_i \) expects utility

\[
\mathbb{E}[u_i|\delta_i] = \mathbb{E}[\pi(x_i, x_{-i})|\delta_i] + \delta_i \mathbb{E}[\pi(x_{-i}, x_i)|\delta_i] \\
= (1 + \delta_i)(R - \mathbb{E}[u_{\delta}(\delta_i)]) + u(h(x_i)) + (\delta_i - h(x_i)) u_{\delta}(h(x_i)) \\
+ \mathbb{E}[u_{\delta}(h(x_{-i}))] + \delta_i \mathbb{E}[u(h(x_{-i})) - h(x_{-i}) u_{\delta}(h(x_{-i}))]
\]

(34)

Agent \( i \)'s marginal utility of his individual effort is thus

\[
\frac{d}{dx_i}\mathbb{E}[u_i|\delta_i] = [\delta_i - h(x_i)] h_x(x_i) u_{\delta\delta}(h(x_i)).
\]

(35)

Since \( h:[x, \bar{x}] \rightarrow [\delta_{\min}, \delta_{\max}] \) is a one-to-one correspondence (and thus strictly monotone), and since \( u_{\delta\delta}(h) > 0 \) for all \( h(x_i) \in [\delta_{\min}, \delta_{\max}] \), agent \( i \) has a unique optimum effort level \( x_i^* \), which satisfies \( \delta_i - h(x_i) = 0 \); thus, \( x_i^* = h^{-1}(\delta_i) = x(\delta_i) \). By Lemma 2(ii), \( P \)'s ex ante expected profit is \( \mathbb{E}_{(\delta, \tau)}[\pi_P] = -2\mathbb{E}[u_{\delta}(\delta)] + 2(\mathbb{E}[x(\delta) - C(x(\delta))] - R) \), which is positive.
by assumption. Rewriting \((34)\) with \(x_i = x_i^* = x(\delta_i)\) and \(h(x(\delta_i)) = \delta_i\) yields

\[
\mathbb{E}[u_i | \delta_i] = (1 + \delta_i) (R - \mathbb{E}[u_\delta(\delta)]) + u(\delta_i) + \mathbb{E}[u_\delta(\delta)] + \delta_i \mathbb{E}[u(\delta) - \delta u_\delta(\delta)]
\]

(36)

By assumption, \(\mathbb{E}[u(\delta)] = \mathbb{E}[(1 + \delta) u_\delta(\delta)]\). Hence, \(\mathbb{E}[u_i | \delta_i] = (1 + \delta_i) R + u(\delta_i)\). Finally, since \(\frac{d}{dx} u_\delta(h(x)) = u_{\delta \delta}(h(x)) h_x(x)\) and \(u_{\delta \delta} > 0\), agent \(i\)'s effort \(x_{-i}\) imposes a positive (negative) externality on agent \(i\)'s pay if and only if \(h_x > 0\) (\(h_x < 0\)), thus, if and only if \(\frac{dx}{d\delta} = \frac{dh^{-1}(\delta)}{d\delta} > 0\) \((\frac{dx}{d\delta} < 0)\). 

**Proof of Lemma 4.** Notice first that the conditions of Lemma 2(ii) are invariant to a multiplication of \(u\) with a positive scalar. Let \(\alpha = 1\) in the following. Obviously, \(u_{\delta \delta} > 0\), and \(u(\delta) \geq 0\) for all \(\delta\). Since \(u_\delta(\delta) = -2(\delta_s - \delta)\), we have \(\mathbb{E}[u(\delta)] = \mathbb{E}[(1 + \delta) u_\delta(\delta)]\) if and only if \(\mathbb{E}[(\delta_s - \delta)^2] = -2 \mathbb{E}[(1 + \delta)(\delta_s - \delta)]\), thus, if and only if \((1 + \delta_s)^2 = \mathbb{E}[(1 + \delta)^2]\). Thus, \(\mathbb{E}[u] = \mathbb{E}[(1 + \delta) u_\delta]\) is satisfied for \(\delta_s = -1 + [\mathbb{E}[(1 + \delta)^2]]^{\frac{1}{2}}\). Since \(\text{Var} [\delta] = \mathbb{E}[[\delta^2] - \mathbb{E}^2[\delta]]\), we have \(\delta_s = -1 + [\text{Var} [\delta] + (1 + \mathbb{E}[\delta])^2]^{\frac{1}{2}}\). Since \(\text{Var} [\delta] > 0\) by assumption, we have \(\delta_s > \mathbb{E}[\delta]\). Thus, \(\mathbb{E}[u_\delta(\delta)] = 2(\mathbb{E}[\delta] - \delta_s) < 0\). Therefore, the conditions of Lemma 2(ii) are satisfied. With Lemma 2(ii), \(\mathbb{E}_{\bar{\pi}_P}[\pi_P] = 4(\delta_s - \mathbb{E}[\delta]) + 2(\mathbb{E}[x - C(x)] - R)\). Obviously, \(\delta_{\min} < \mathbb{E}[\delta]\). Finally, \(-1 + [\mathbb{E}[(1 + \delta)^2]]^{\frac{1}{2}} = \delta_s < \delta_{\max}\) if and only if \(\mathbb{E}[(1 + \delta)^2] < (1 + \delta_{\max})^2\); and the latter holds, since \([\delta_{\min}, \delta_{\max}] \subset (-1, 1)\) and \(\text{Var} [\delta] > 0\). Therefore, \(\delta_{\min} < \mathbb{E}[\delta] < \delta_s < \delta_{\max}\). 

**Proof of Proposition 5.** Agent \(i\) of type \(\delta_i\) chooses his effort \(x_i^*\) in order to maximize \(\left[2\delta_{\max} - 2\delta_i - \delta V(x_i)\right] V(x_i)\), where \(V(x) = 1_{[0,x^*]}(x) [x/x^*]^\frac{\alpha}{2}\). Thus, \(x_i^* = x(\delta_i)\), where \(x(\delta_i) = x^* \left[\left(\delta_{\max} - \delta_i\right)/\delta\right]^\alpha\). For all \(\delta \in [\delta_{\min}, \delta_{\max}]\),

\[
V(x(\delta)) = \frac{\delta_{\max} - \delta_i}{\delta}.
\]

Equation (37)

Suppose \(W_{RPP}\) is implemented with types \(\delta, \tau \in [\delta_{\min}, \delta_{\max}]\). Then agent \(i\) of type \(\delta_i = \delta\) realizes an ex post utility level of

\[
u_i = \left[w_{RPP}(x(\delta), x(\tau)) - C(x(\delta))\right] + \delta \left[w_{RPP}(x(\tau), x(\delta)) - C(x(\tau))\right]
\]

\[
= (1 + \delta) R + \alpha \left[\delta_s^2 + \delta^2 - 2\delta \delta_s + 2\delta \delta_s - 2\mathbb{E} [\delta] + \delta \delta_s^2 - 2\delta \mathbb{E} [\delta] + 2\tau - \delta \tau^2\right].
\]

Equation (38)

Rewriting with \(u(\delta) = \alpha (\delta_s - \delta)^2\) yields

\[
u_i = (1 + \delta) R + u(\delta) + \alpha \left[2(\tau - \mathbb{E} [\tau]) + \delta \left(2\delta_s + \delta_s^2 - 2\mathbb{E} [\tau] - \tau^2\right)\right].
\]

Equation (39)
By definition, \( \delta_* = -1 + \left[ \mathbb{E} \left[ (1 + \tau)^2 \right] \right]^{\frac{1}{2}} \), thus, \( \delta_*^2 = -2\delta_* + 2\mathbb{E} [\tau] + \mathbb{E} [\tau^2] \). Therefore,

\[
u_i = (1 + \delta) R + u(\delta) + \alpha \left[ 2(\tau - \mathbb{E}[\tau]) - \delta \left( \tau^2 - \mathbb{E}[\tau^2] \right) \right]. \tag{40}
\]

Taking expectations w.r.t. \( \tau \) yields \( \mathbb{E}[\nu_i] = (1 + \delta) R + u(\delta) \). Implementing \( W_{RPP} \) with types \( \delta, \tau \in [\delta_{\min}, \delta_{\max}] \) yields \( P \) an ex post profit of

\[
\pi_P = x(\delta) + x(\tau) - w_{RPP}(x(\delta), x(\tau)) - w_{RPP}(x(\tau), x(\delta)).
\]

Rewriting with (37), and \( S(\alpha) \) from (13), yields

\[
\pi_P = x(\delta) - C(x(\delta)) - R + x(\tau) - C(x(\tau)) - R
\]
\[
+ \alpha \left[ -4\mathbb{E}[\delta] - 2 + (1 - \delta)^2 + (1 - \tau)^2 - 2\delta_*^2 \right]. \tag{41}
\]

Since \( \delta_*^2 = -1 - 2\delta_* + \mathbb{E} \left[ (1 + \tau)^2 \right] \) by definition, (41) yields

\[
\pi_P = x(\delta) - C(x(\delta)) - R + x(\tau) - C(x(\tau)) - R
\]
\[
+ 4\alpha (\delta_* - \mathbb{E}[\delta]) + \alpha \left[ (1 - \delta)^2 + (1 - \tau)^2 - 2\mathbb{E} [(1 - \delta)^2] \right]. \tag{42}
\]

Therefore, \( \mathbb{E}_{(\delta, \tau)}[\pi_P] = 2\mathbb{E}[x(\delta) - C(x(\delta))] - R + 4\alpha (\delta_* - \mathbb{E}[\delta]) \). Since \( \lim_{\kappa \to 0} x(\delta) = x^* \) for any \( \delta \in [\delta_{\min}, \delta_{\max}] \), and \( \delta_* > \mathbb{E}[\delta] \), we have \( \lim_{\kappa \to 0} \mathbb{E}_{(\delta, \tau)}[\pi_P] > 0 \). \( \blacksquare \)

**Proof of Proposition 6.** Agent \( i \) of type \( \delta_i \) chooses his effort \( x_i^* \) in order to maximize \( [2\delta_{\max} - 2\delta_i - 3V(x_i)] V(x_i) \), where \( V(x) = 1_{[x^*, \infty]}(x) \left( \frac{\bar{V} - x}{\bar{V} - x^*} \right)^{\frac{1}{1 - \kappa}} \). Thus, \( x_i^* = x(\delta_i) \), where \( x(\delta_i) = \bar{V} - (\bar{V} - x^*) \left[ (\delta_{\max} - \delta_i) / \delta \right]^\kappa \). Notice that \( \bar{V} \left( x(\delta) \right) = (\delta_{\max} - \delta_i) / \delta \), as in (37). Comparison of (15) and (19) reveals that the distributive effects of \( W_{TPP} \) are exactly those of \( W_{RPP} \). Ex post utilities and profits under \( W_{TPP} \) are thus the same as under \( W_{RPP} \). \( \blacksquare \)

**Proof of Proposition 8.** In the following, I refer to the matching process stated in Proposition 8 as permanent rematching. By assumption, the rematching technology ensures that if all firms and agents do engage in permanent rematching, while always contracting according to \( W^* \), then each agent \( i \) believes in each period to end up with a coworker of type \( \delta_{-i} \sim F \), while each (active) principal believes in each period to end up with types \( (\delta_i, \delta_{-i}) \sim F \times F \).

First assume that all firms and agents do engage in permanent rematching, while always contracting according to \( W^* \). Then, by Lemma 4, each firm expects zero profits in all periods, as required by perfect competition. Therefore, if each firm believes that all agents and all
the other firms do engage in permanent rematching, while always contracting according to \( W^* \), then no firm has an incentive to deviate from engaging in permanent rematching while always offering \( W^* \).\(^{43}\)

For some fixed period \( t \), consider a triplet \([P, i, -i]\) contracting according to \( W^* \). By assumption, the agents’ types are their private information before engaging in production. Suppose agent \( i \) believes that \( P \) and \(-i\) engage in permanent rematching and will thus reject to contract in the same constellation in the subsequent period \( t + 1 \). Then \( W^* \) provides agent \( i \) with the dominant effort strategy \( x(\delta_i) \), as given by \((14)\); in addition, consistent with his belief, agent \( i \) expects coworker \(-i\) to reveal his social type by exerting the effort \( x(\delta_{-i}) \). Given his belief \( \delta_{-i} \sim F \), and according to \((??)\) and Lemmas 3 and 4, agent \( i \) expects a utility level of \( \mathbb{E}[u_i^t] = (1 + \delta_i) (R + S) + \alpha (\delta_* - \delta_i)^2 \). For \( \kappa \) sufficiently small, we have \( S > 0 \); thus, \( \mathbb{E}[u_i^t] > (1 + \delta_i) R \). By symmetry, the same reasoning holds for agent \(-i\); thus, \( \mathbb{E}[u_{-i}^t] > (1 + \delta_{-i}) R \). Hence, in each period, \( W^* \) implements interim Pareto improvement upon the first-best efficient status.

The feasibility of interim Pareto improvement upon the first-best in each period implies that, after having contracted according to \( W^* \) in some period \( t \), the triplet \([P, i, -i]\) does not reach a unanimous agreement to contract again in period \( t + 1 \): Consistent with their beliefs that contracting is for one period only, the agents have revealed their social types through the efforts they have exerted in \( t \). That is, both agents’ types are now common knowledge among \([P, i, -i]\). By Lemma 1, any ex post efficient contract \( W \) implements first-best efficient efforts \( x^* \), and any allocation of the respective market return, \( 2x^* = w_i + w_{-i} \), among \( i \) and \(-i\) is Pareto efficient. (Due to perfect competition, the principal realizes zero profits.) Obviously, at least one of the agents, say \( i \), does not earn more than his coworker: \( w_i = x^* - \varepsilon \) and \( w_{-i} = x^* + \varepsilon \) for some \( \varepsilon \geq 0 \). Since \( x^* - \varepsilon + \delta_i (x^* + \varepsilon) \leq (1 + \delta_i) x^* \) for all \( \varepsilon \geq 0 \) and \( \delta_i \in (-1, 1) \), agent \( i \)'s utility from \( W \) is bounded above: \( u_i^{t+1} \leq (1 + \delta_i) R < \mathbb{E}[u_i^{t+1}] \).

Therefore, at least one agent is ex interim strictly better off when engaging in rematching in period \( t + 1 \), and therefore rejects \( W \). For the same reason, a contract offered by some principal \( P \) in period \( t + 1 \) only finds unanimous approval of two randomly selected agents if it is interim-efficient.

Hence, given that all principals and agents believe that all principals and agents do engage in permanent rematching, while always contracting according to \( W^* \), it is incentive compatible for each principal and agent to do exactly that. These beliefs are therefore mutually consistent.

\(^{43}\)Necessarily, \( \kappa > 0 \) must be bounded below, \( \kappa \geq \kappa^* \) for some \( \kappa^* > 0 \). Otherwise, according to \((14)\) and Lemma 4, each principal would have the incentive to let \( \kappa \) arbitrarily small in order to marginally increase production efficiency. This incentive would generally rule out the existence of any equilibrium.
Due to the payoff equivalence of $W_{RPP}$ of Proposition 5 and $W_{TPP}$ of Proposition 6, the arguments are equally valid if some, or all, principals make use of team performance pay according to $\tilde{W} = [\tilde{w}(x_i, x_{-i}), \tilde{w}(x_{-i}, x_i)]$, with $\tilde{w}(x, y) = w_{TPP}(x, y) + S$. ■

**Proof of Proposition 9.** Suppose there is a direct mechanism that incentivizes truthful revelation, while guaranteeing $P$ a non-negative ex post profit If it is common knowledge that the direct mechanism incentivizes truthful revelation, while guaranteeing $P$ a non-negative ex post profit (due to $P$'s imperfect commitment).

By (16), contracting according to $W^*$ from (20) with a coworker of type $\tau$ yields agent $i$ of type $\delta_i$ an ex post utility level of

\[ u_i(\tau) = (1 + \delta_i)(R + S) + u(\delta_i) + \alpha \left[ 2(\tau - \mathbb{E}[\tau]) - \delta_i \left( \tau^2 - \mathbb{E}[\tau^2] \right) \right]. \tag{43} \]

Denote by $u_{\delta_i} = \mathbb{E}[u_i(\tau)] = (1 + \delta_i)(R + S) + u(\delta_i)$, with $u(\delta_i)$ from Lemma 4, agent $i$'s interim expected utility from immediate rematching in case contracting does not find unanimous approval among $[P, i, -i]$. The term $u_{\delta_i}$ gives a type $\delta_i$'s reservation (expected) utility when entering the job talk stage. Since $\partial u_i/\partial \tau = 2\alpha (1 - \delta_i \tau) > 0$ and $\mathbb{E}[u_i(\tau)] = u_{\delta_i}$, there is a unique cutoff value $t(\delta_i) \in (\delta_{\min}, \delta_{\max})$ for each $\delta_i \in (\delta_{\min}, \delta_{\max})$ such that $u_i(\tau) < u_{\delta_i}$ if $\tau < t(\delta_i)$, and $u_i(\tau) > u_{\delta_i}$ if $\tau > t(\delta_i)$. The value $t(\delta_i)$ determines a type $\delta_i$'s reservation type. In particular, $\delta_{\min} < t(\delta_i) < \delta_{\max}$ for each $\delta_i \in [\delta_{\min}, \delta_{\max})$.

By (17) and $w^* = w_{RPP} + S$, implementation of $W^*$ with types $\delta$ and $\tau$ yields $P$ an ex post profit of

\[
\pi_P(\delta, \tau) = x(\delta) - C(x(\delta)) - \mathbb{E}[x - C(x)] + \alpha \left[ -2(\delta - \mathbb{E}[\tau]) + (\delta^2 - \mathbb{E}[\tau^2]) \right] + x(\tau) - C(x(\tau)) - \mathbb{E}[x - C(x)] + \alpha \left[ -2(\tau - \mathbb{E}[\tau]) + (\tau^2 - \mathbb{E}[\tau^2]) \right]: \tag{44}
\]

The principal's reservation profit is $\pi_P = 0$. Notice that $\pi_P(\delta, \tau)$ is symmetric and additively separable in the composition of types. Two cases are to be discussed separately: (i) $\pi_P(\delta_{\min}, \delta_{\max}) < 0$; and (ii) $\pi_P(\delta_{\min}, \delta_{\max}) \geq 0$.

Consider (i). Since $\partial\pi_P/\partial\delta < 0$ and $\mathbb{E}(\delta, \tau)[\pi_P] = 0$, there exists for each $\delta \in [\delta_{\min}, \delta_{\max}]$ a unique cutoff value $t_P(\delta) \in (\delta_{\min}, \delta_{\max})$ such that $\pi_P(\delta, t_P(\delta)) = 0$, while $\pi_P(\delta, t_P(\delta)) > 0$ if $\tau < t_P(\delta)$, and $\pi_P(\delta, t_P(\delta)) < 0$ if $\delta > t_P(\delta)$. The value $t_P(\delta_i)$ determines $P$’s reservation type for agent $-i$ if $P$ knows that agent $i$ is of type $\delta_i$. That is, if $P$ knows that $i$’s type is $\delta_i$, then $P$ agrees to contract with $i$ and $-i$ if and only if $\delta_{-i} \leq t_P(\delta_i)$. Since $\partial\pi_P/\partial\delta < 0$, we have $\partial t_P(\delta)/\partial\delta < 0$. Since $\pi_P(\delta_{\min}, \delta_{\max}) < 0$ by assumption, we have $t_P(\delta_{\min}) < \delta_{\max}$.

Suppose there is a direct mechanism that incentivizes truthful revelation, while guaranteeing $P$ a non-negative ex post profit If it is common knowledge that the direct mechanism
incentivizes truthful revelation, while guaranteeing \( P \) a non-negative ex post profit, then agent \(-i\) of type \( \delta_i \) anticipates that \( P \) would only agree to contract according to \( W^* \) if agent \(-i\) is of type \( \delta_{-i} \leq t_P(\delta_i) < \delta_{\text{max}} \). Provided that agent \(-i\) and \( P \) do agree to contract, type \( \delta_i \)'s updated belief about \(-i\)'s type is given by the c.d.f. \( F_{\delta_{-i} \leq t_P(\delta_i)} : [\delta_{\text{min}}, t_P(\delta_i)] \to [0, 1] \), with

\[
F_{\delta_{-i} \leq t_P(\delta_i)}(\delta_{-i}) = \frac{\int_{\delta_{\text{min}}}^{\delta_{-i}} f(t(\tau)) \, dt}{\int_{\delta_{\text{min}}}^{t_P(\delta_i)} f(t(\tau)) \, dt},
\]

where \( f \) is the density of the exogenously given c.d.f. \( F \) of types in the population of agents. Since \( \partial u_i / \partial t > 0 \), and since his reservation type satisfies \( t(\delta_i) < \delta_{\text{max}} \), agent \( i \)'s expected utility from contracting with \( P \) and \(-i\) is thus strictly smaller than his reservation utility from interim-efficient contracting with some randomly selected type in another firm: \( \mathbb{E}_{\delta_{-i} \leq t_P(\delta_i)}[u_i(t(\tau))] < u_{\text{br}} \), where the operator \( \mathbb{E}_{\delta_{-i} \leq t_P(\delta_i)} \) indicates that expectations are taken with respect to the c.d.f. (45). Agent \( i \), regardless of his type \( \delta_{-i} \in [\delta_{\text{min}}, \delta_{\text{max}}] \) thus rejects to contract with \( P \) and \(-i\).

Now consider (ii). Since \( \pi_P(\delta_{\text{min}}, \delta_{\text{max}}) \geq 0 \), \( \partial \pi_P / \partial \delta < 0 \), and \( \mathbb{E}_{(\delta, \tau)}[\pi_P] = 0 \), there exist unique \( \delta_0'_{\text{min}} \) with \( \delta_{\text{min}} < \delta_0'_{\text{min}} < \delta_{\text{max}} \) such that the following holds: \( \pi_P(\delta_0'_{\text{min}}, \delta_{\text{max}}) = 0 \), and for each \( \delta \in (\delta_0'_{\text{min}}, \delta_{\text{max}}) \) there exists a unique cutoff value \( t_P(\delta) \in (\delta'_{\text{min}}, \delta'_{\text{max}}) \) such that \( \pi_P(\delta, t_P(\delta)) = 0 \), while \( \pi_P(\delta, t_P(\delta)) > 0 \) if \( \tau < t_P(\delta) \), and \( \pi_P(\delta, t_P(\delta)) < 0 \) if \( \delta > t_P(\delta) \). Since \( \partial t_P(\delta) / \partial \delta < 0 \), \( \pi_P(\delta'_{\text{min}}, \delta_{\text{max}}) = 0 \), and \( \pi_P(\delta'_{\text{min}}, \delta'_{\text{max}}) = \pi_P(\delta'_{\text{max}}, \delta'_{\text{min}}) \), we have \( t_P(\delta_{\text{min}}) = \delta_{\text{max}} \) and \( t_P(\delta_{\text{max}}) = \delta_0'_{\text{min}} \). Again, \( t_P(\delta_i) \) determines \( P \)'s reservation type for agent \(-i\) if \( P \) knows that agent \( i \) is of type \( \delta_i \). Furthermore, if \( \delta_i \leq \delta_0'_{\text{min}} \), then \( P \) accepts any type \( \delta_{-i} \in [\delta_{\text{min}}, \delta_{\text{max}}] \), such that \( t_P(\delta) = \delta_{\text{max}} \) for all \( \delta \in [\delta_{\text{min}}, \delta_{\text{max}}] \). If it is common knowledge that the direct mechanism incentivizes truthful revelation, while guaranteeing \( P \) a non-negative ex post profit, then agent \( i \) of type \( \delta_i \) anticipates that \( P \) would only agree to contract according to \( W^* \) if agent \(-i\) is of type \( \delta_{-i} \leq t_P(\delta_i) \). First suppose agent \( i \) is of type \( \delta_i \in [\delta_0'_{\text{min}}, \delta_{\text{max}}] \). Then \( i \) rejects to contract exactly for the reason outlined in case (i). Now suppose agent \( i \) is of type \( \delta_i \in [\delta_{\text{min}}, \delta_0'_{\text{min}}] \). Then \( i \) anticipates that each agent \(-i\) of type \( \delta_{-i} \in [\delta_{\text{min}}, \delta_{\text{max}}] \) will reject to contract, according to (i). Since \( t(\delta_i) < \delta_{\text{max}} \) and \( \partial u_i / \partial \tau > 0 \), each type \( \delta_i \in [\delta_{\text{min}}, \delta_0'_{\text{min}}] \) will reject to contract, too. This completes the proof. ■

The game of hawk-dove-principal if the agents exhibit CRRA.

In the setting of Section 2, with all else equal, take each agent \( i \)'s utility as

\[
u_i = \frac{\pi_i^{1-\rho}}{1-\rho} + \delta_i \frac{\pi_i^{1-\rho}}{1-\rho},
\]

with \( \rho \geq 0 \) denoting all agents’ unanimous degree of CRRA, and with \( \delta_i \in \{-\frac{1}{2}, \frac{1}{2}\} \). For
Figure 7: The strategic game that the principal invites the agents to play, where $0 < x \leq 1/4$, and $\beta = 1 - \rho$ for the degree of CRRA $\rho \in [0, 1) \cup (1, \infty)$.

<table>
<thead>
<tr>
<th>$\pi_i$</th>
<th>$\beta$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>$(1 - 4x)^{1/\beta} - 1 \ (1 - 4x)^{1/\beta} - 1$</td>
<td>$(1 + x)^{1/\beta} - 1 \ (1 - 2x)^{1/\beta} - 1$</td>
</tr>
<tr>
<td>$C$</td>
<td>$(1 - 2x)^{1/\beta} - 1 \ (1 + x)^{1/\beta} - 1$</td>
<td>$(1 + x)^{1/\beta} - 1 \ (1 + x)^{1/\beta} - 1$</td>
</tr>
</tbody>
</table>

briefness and simplicity, let $\rho \neq 1$, and $\beta = 1 - \rho \in (-\infty, 0) \cup (0, 1]$ in the following. Thus, $\beta u_i = \pi_i^\beta + \delta_i \pi_{-i}^\beta$.

The principal $P$ invites the agents to play the strategic game on the left hand side of Figure 7, where $0 < x \leq 1/4$ due to the agents’ limited liability. If agents $i$ and $-i$ receive ex post payments of $z_i^{1/\beta} - 1$ and $z_{-i}^{1/\beta} - 1$ (positive or negative) from $P$, then, given their individual endowments of one, agent $i$ realizes an ex post utility of

$$\beta u_i = \left(1 + z_i^{1/\beta} - 1\right) + \delta_i \left(1 + z_{-i}^{1/\beta} - 1\right) = z_i + \delta_i z_{-i}. \quad (47)$$

The agents’ ex post utilities are therefore the very same (numerically) for any value of $\beta$, in particular the same as in the case of risk neutrality, $\beta = 1$. Hence, for any $\beta \in (-\infty, 0) \cup (0, 1]$, the agents’ ex post utilities are the same as already depicted on the right hand side of Figure 1. Again, each agent reveals his type through dominant strategy implementation. Their reservation utilities are $u_{Dove} = 1 + \delta_{Dove}1 = 3/2$ and $u_{Hawk} = 1 + \delta_{Hawk}1 = 1/2$. Beliefs about the composition of types are the same as in Figure 2. It is straightforward to see that all types are indifferent ex interim between accepting and rejecting play; again, they are assumed to accept. What is left to show is that $P$ can choose $x > 0$ in response to $\beta$ in such a way that she realizes a positive profit ex ante. Her ex ante profit satisfies

$$E_{(\delta, x)} [\pi_P] = 2 - \frac{1}{2} (1 - 4x)^{1/\beta} - (1 + x)^{1/\beta} - \frac{1}{2} (1 - 2x)^{1/\beta}. \quad (48)$$
Let $g(x, \beta) = \mathbb{E}_{(\delta, \tau)}[\pi_P]$. Obviously, $g(0, \beta) = 0$ for all $\beta$. Furthermore,

$$
\frac{d}{dx} g(x, \beta) \bigg|_{x=0} = 2 \cdot \frac{1}{\beta} (1 - 4x)^{1/\beta - 1} - \frac{1}{\beta} (1 + x)^{1/\beta - 1} + \frac{1}{\beta} (1 - 2x)^{1/\beta - 1} \bigg|_{x=0} = \frac{2}{\beta}.
$$

We thus have $\mathbb{E}_{(\delta, \tau)}[\pi_P] > 0$ for each $\beta \in (0, 1]$ if $x$ is sufficiently small, contingent on $\beta$. One can show that $\mathbb{E}_{(\delta, \tau)}[\pi_P] < 0$ for all $\beta \in (-\infty, 0)$ and all $x > 0$. This proves that interim Pareto improvement upon ex post efficient allocations can be implemented for moderate degrees of CRRA, $\rho \in [0, 1)$. ■