European University Institute

MAX WEBER PROGRAMME FOR POSTDOCTORAL STUDIES

EUI WORKING PAPERS

MWP 2017/04
Max Weber Programme

Contracting with Type-Dependent Naïveté

Matteo Foschi

European University Institute
Max Weber Programme

Contracting with Type-Dependent Naïveté

Matteo Foschi

EUI Working Paper **MWP** 2017/04

ISSN 1830-7728

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract

I analyse the optimal contracting behaviour of an employer who faces workers with di erent, incorrect beliefs about their productivity (naïve workers). Researchers in contract theory have analysed cases where the principal has full in-formation on agents' true productivity, and cases where the principal has priors on productivity levels. I contribute to this discussion by introducing the novel assump-tion that workers' naïveté depends on their actual productivity level. In particular, I focus on the use the employer makes of this information when designing contracts under asymmetric information. The results highlight a new trade-o the employer faces between exploiting strongly naïve workers and designing e cient contracts for the most widespread type of worker, according to her posteriors.

Keywords

Self-Awareness, Naïveté, Contract, Screening, Non-Common Priors, Mechanism De-sign, Multidimensional Types.

JEL Classification: D42 D82 D84 D86 J41

This version: April 13, 2017

I would like to thank Subir Bose, David Myatt, Herakles Polemarchakis and Chris Wallace for useful comments that improved the paper. I would also like to thank the seminar attendants at the RES 2016 Conference, Ce2 workshop 2016, York, Leicester and the European University Institute. I gratefully acknowledge financial support from the Royal Economic Society.

Matteo Foschi
Max Weber Fellow, 1016-2017
European University Institute, matteo.foschi@eui.eu

1. Introduction

When facing a new task, individuals form expectations about their ability to carry it out, and about the amount of effort required to do so. Typically, they are assumed to always hold unbiased beliefs about their abilities. Often, however, this is not the case. From the workman estimating the time to build a wall to the athlete who forms expectations about the amount of effort needed to achieve a specific goal, the final result is not always the one expected. In economics we often assume that such ‘errors’ result from specific realisations of random variables either side of the unbiased expectation. However, estimations can be distorted by one having wrong perception of the situation, or by firm beliefs that turn out to be inconsistent with reality. Economics deals with these kinds of situations with the concept of *naïveté* (Strotz, 1956), that is, the inability of an individual to form unbiased expectations about an unknown event. In other words, a systematic over- or underestimation of the realisation of a random variable.

Extensive and highly influential work has been carried out on how naïveté affects contract design and agents’ welfare, in several different frameworks (the most relevant papers are Eliaz and Spiegler, 2006, 2008). In the literature, naïveté is often assumed to be an independent feature of agents. Nevertheless, extensive experimental evidence has shown that one’s inability to evaluate skills in a particular domain is often correlated with a lack of the very same skills one is trying to estimate.[2] Hence, for example, the inability of a worker to estimate his own abilities correctly may be a sign of his unsuitability for the job; a student’s misperception of the amount of hours needed to study for a test may be correlated to her lack of the skills required to pass the test; a consultant biased opinion on an issue may be a signal of his poor expertise on the matter at hand.

Following these results, in this paper, I investigate situations where the level of naïveté of an individual depends on his own innate ability. In particular, I study workers who have systematically wrong (naïve) beliefs about their own productivity. Unlike the existing literature I make the novel assumption that workers’ naïveté depends on their ability (their *type*). On the other hand, the employer, who is perfectly unbiased, designs contracts to hire the workers.[3] Besides a surplus extraction motive, when maximising profits the principal is interested in using this information in order to design more efficient contracts. Here, a new trade-off

[2]Important contributions are Svenson (1981); Dunning and Kruger (1999); Dittrich, Güth, and Marciejovsky (2005); Banner, Dunning, Ehrlinger, Kerri, and Kruger (2008); Moore and Healy (2008). These are analysed in section 2.

[3]The assumption about the employer having unbiased belief can be thought of as her having more experience, or better knowing the suitability of the worker population for the specific job she is hiring for. Ultimately, dropping this assumption simply changes the interpretation of the model, but not its results.

emerges on the use of this information. Does the principal use it to exploit agents' naïveté even further or simply to increase the overall economic efficiency of contracts?

The model features two periods. A principal hires agents in period 1 to carry out a task in period 2. Before facing the task they are assigned to, agents have limited information about their true type and they are assumed to form biased (wrong) beliefs about it in period 1 — i.e. they are naïve. Furthermore, each agent's beliefs depend on his actual type — that is, naïveté is type-dependent. While their naïveté prevents agents from updating their beliefs the principal is fully aware of the correlation between type and beliefs. The *use* the principal makes of this information is the key insight of the paper. In equilibrium, she designs contracts that, first, screen among differently naïve agents in period 1 and, second, screen among different types of agent in period 2. Given her formed posterior from period 1 screening, the principal faces a trade-off: to design efficient contracts either for the most naïve types, in which case she *exploits* them, taking advantage of their wrong beliefs, or for the types she deems most probable given her posteriors, therefore maximising efficiency regardless of naïveté levels.

The model can be applied to job market contracts, job interviews, procurement (where bids can be assumed to convey a level of 'confidence' the seller has in his ability to provide the service) or any economic situation where the level of confidence of an agent is assumed to transmit some information about the agent.[4]

Finally, although the evidence shows the existence of a correlation between beliefs and ability, the evidence on the direction of this correlation is mixed and not unanimous. For this reason, in the paper, I consider all interesting directions (and magnitudes) of correlation.

The paper is organised as follows. In section 2 I present the related literature. In section 3 I explain the model and the assumptions. In section 4 I study the case of perfect correlation between naïveté and agents' types. I then relax this assumption and study the general case in section 5. I conclude the paper in section 6. All proofs of lemmas, results and propositions are relegated to the Appendix.

2. Related Literature

Extensive experimental evidence motivates the main assumption behind this work. A first set of papers (among others: Svenson, 1981; Chi, Glaser, and Rees, 1982; Dunning and Kruger, 1999; Dunning, Ehrlinger, Johnson, and Kruger, 2003;

[4]For example, think about the junior academic market, where in interviews candidates are often asked: "how much do you expect to be offered given the advertised wage range?". The question can be thought of as an attempt at learning the level of confidence of the canditate. This is because the interviewer may believe that there is a correlation between the candidate's confidence and quality.

Banner, Dunning, Ehrlinger, Kerri, and Kruger, 2008) show that the skills needed to evaluate competence in a specific domain are exactly the same as those required to engender this competence. Hence, individuals without such skills should find it relatively hard to estimate their own competence correctly. Building on these findings, a second set of papers (Dittrich, Güth, and Marciejovsky, 2005; Banks, Lawson, and Logvin, 2007; Moore and Healy, 2008; Ferraro, 2010) present further experimental evidence of a positive correlation between competence and self-awareness. Finally, a third set of papers (Lichtenstein, Fischoff, and Phillips, 1982; Loewenstein, O'Donoghue, and Rabin, 2003; Conlin, O'Donoghue, and Vogelsang, 2007) focus on the concept of 'over-confidence' and projection-bias providing strong experimental evidence on bias individuals' expectations.[5]

The contributions listed above highlight two main facts: (i) individuals are not perfectly capable of estimating their own skills; and (ii) often their estimation of their capabilities depends on the same skills they are trying to evaluate. To date, the economics literature has only dealt with these facts separately.

Harris and Raviv (1979) first studied a model with agents who lack knowledge about their type. They assume, however, that agents form unbiased beliefs, corresponding with the principal's ones.[6] Self-awareness and naïveté were first introduced by Strotz (1956) and were later applied in contract theory. Among others, the papers that most relate to this one are Eliaz and Spiegler (2006) and Eliaz and Spiegler (2008).[7] In both papers, time-inconsistent agents differ in their levels of naïveté, with some of them being perfectly self-aware (*sophisticated*). In Eliaz and Spiegler (2006), the employer has full information about consumers' preferences. The optimal menu provides a commitment device for sophisticated agents, who would like to play according to their present preferences, as opposed to their future preferences. Relatively naïve agents, instead, are exploited because of their inability to correctly estimate their actual type. In Eliaz and Spiegler (2008), the

[5]Less related, Bagues and Perez-Villadoniga (2012) show that these findings extend to the estimation of other people's skills as well. They use evidence from a field experiment to show that recruiters prefer to hire applicants with capabilities similar to their own. One of the proposed explanations is that evaluators' accuracy is higher when evaluating those dimensions in which their knowledge is greater.

[6]Their results show that the principal finds it optimal to charge a fixed fee to the agent in exchange for her profits (hence the name 'firm-selling' equilibrium). Full efficiency is achieved in this case.

[7]Other important contributions: O'Donoghue and Rabin (2001), Asheim (2007) and Heidhues and Köszegi (2010) (among others) study the interaction between naïveté and self-control, modelled as present-biased preferences. Amador, Werning, and Angeletos (2006) analyse the trade-off faced by a multi-self agent who is aware of his time-inconsistency problems, but is not aware of his true preferences until later periods. Gilpatric (2008) studies the problem of moral hazard in the presence of naïvé agents with time-inconsistent preferences. Less related is Von Thadden and Zhao (2012) who study a classical principal agent model where agents do not know their action space until a later stage.

authors extend the model to one where the employer has priors over consumers' preference-changes. Hence, two screening processes take place, exactly as in this paper. The first screening separates differently self-aware agents; the second separates with respect to their preferences. In both papers, however, agents' beliefs and types are assumed to be independent.

Type-dependent beliefs were first introduced in models of "sequential screening" (Courty and Li, 2000). There, agents hold unbiased beliefs about their type, but the precision of these beliefs depends on the agents' type itself. In line with the results of this paper, Courty and Li (2000) find that optimal contract design depends on the "informativeness" of agents' initial knowledge rather than on the principal's priors. Unlike the present work, however, they assume non-naïve agents — i.e. agents with unbiased expectations — leaving no space for exploitation. On the contrary, the optimal mechanism features "refund contracts", which grant agents the option to claim a refund after they learn their true willingness to pay. In recent years, Courty and Li's (2000) model has been extended and applied. Among others, Kovác and Krähmer (2013) study sequential delegation, Deb and Said (2015) study the case of a principal with limited commitment power, Evans and Reiche (2015) relax the commitment assumption completely and Grubb (2015) applies the model to the cellular phone service market.

My model builds on these contributions to study situations where types (or preferences) affect the bias in agents' beliefs. My results bridge the findings of screening models with diversely naïve agents (as in Eliaz and Spiegler, 2006, 2008) and of sequential screening model (as in Courty and Li, 2000), providing a new perspective on the connections between these two literatures.

3. The Model

An employer (the principal, she) seeks to hire a worker (the agent, he) from a population. Workers are hired in period 1 and asked to complete an individual task in period 2. The outcome of the task depends on the level of effort $e \in [0,1]$ a worker exerts then. I assume that the level of effort exerted by the worker is perfectly observable.[8]

To hire workers, in period 1, the employer offers a set of contracts $w(e) : [0,1] \to \mathbb{R}$ that each worker can either accept or reject. When a worker accepts a contract in period 1, and exerts effort e in exchange for wage $w(e)$ in period 2, the employer enjoys profits $\Pi = y(e) - w(e)$, where $y(e)$ is increasing and concave in e.

[8]Extending the model to a moral hazard framework where e is partially, or not at all, observable, is left for future research.

When a worker accepts a contract, he enjoys utility $U^j = w(e) - \theta_j e$, where θ_j is the cost of effort and represents a worker's *productivity type*. Finally, if a worker rejects a contract, both he and the employer obtain zero utility/profits.

The population of workers is composed of a portion λ of *productive* types, who have $\theta_j = \theta_P$, and a portion $(1 - \lambda)$ of *unproductive* types, who have $\theta_j = \theta_U > \theta_P$.

The first main assumption of the paper is that in period 1 neither the employer nor the workers are aware of the workers' productivity types. While the employer forms unbiased expectations workers have biased heterogeneous beliefs about themselves, that is, they are *naïve*. Given this, the employer's expectation about a worker's utility is given by $E(\theta) = \lambda U^P + (1 - \lambda)U^U$. A worker's belief about his own utility, instead, depends on his *belief type*. A worker can be *optimistic* or *pessimistic* about his true productivity. In the first case, the agent believes himself to be a productive type with probability $\phi > \lambda$, that is $\Pr\{\theta_j = \theta_P\} = \phi$. In the second case, he believes himself to be a productive type with probability $\delta < \lambda$, that is $\Pr\{\theta_j = \theta_P\} = \delta$. A belief type i expects his productivity to be $E_i(\theta) = i\theta_P + (1-i)\theta_U$, $i = \{\phi, \delta\}$. Notice that an agent is considered optimistic (pessimistic) with respect to the average of the population and not with respect to his actual productivity.[9]

The second main assumption of the paper, and the one that constitutes the main departure from the literature, states that a worker's beliefs and productivity are not independent. Here, I assume that the distribution of belief types is conditional on a worker's true productivity. In particular, there is a proportion p_P (p_U) of pessimistic types among productive (unproductive) workers. Let $\mathcal{B}$ denote the belief of a worker. Then the employer has priors: $p_P = \Pr\{\mathcal{B} = \delta | \theta = \theta_P\}$ and $p_U = \Pr\{\mathcal{B} = \delta | \theta = \theta_U\}$. This allows her to update her priors on a worker's productivity when she knows his belief type.

Workers update their prior only when they face the task. In period 2, they learn their true productivity *before* choosing the level of effort to exert.

Given the assumptions above, the employer faces two different connected screening problems. In period 1 she wants to separate workers according to their belief type. This allows her to update her priors in period 2 and separate workers on the basis of their productivity type. Notice that the employer and the agents always have different beliefs throughout the game. In period 1, the employer forms unbiased expectations, while workers rely on their naïve beliefs. In period 2, the employer updates her priors given the separation of period 1, while workers learn their true productivity and behave as fully informed agents. This implies that the maximisation problem the employer solves is subject to period 1 constraints, which

[9]This is because the productivity-type space is composed of only two types. Hence, if the optimistic or pessimistic nature of a worker were based on his true productivity, all productive (unproductive) workers would be pessimistic (optimistic) simply because $E(\theta) > \theta_P$ ($E(\theta) < \theta_U$).

depend on workers' belief types, and period 2 constraints which depend on workers' true productivity types.

Before stating the problem formally, I define $(w_i^j, e_i^j) \equiv (w_i(e_i^j), e_i^j)$ as the wage and effort level that a worker of belief type i and productivity type j chooses in period 2. Notice that workers' utility only depends on the level of effort they choose (or believe they will choose) in period 2, and that once they sign a contract they are constrained to carry out the task — i.e. there is no individual rationality constraint in period 2. Therefore, I can restrict my attention, without loss of generality, to four effort levels, and the corresponding wages set by the employer: $e_\delta^U, e_\delta^P, e_\phi^U, e_\phi^P$.

Given this, the employer solves:

$$\max_{\{w_i^j\}_{i=\delta,\phi,j=P,U}} \quad E(\Pi) \tag{1}$$

$$\begin{aligned}
\text{s.t.} \quad E_\delta(U^j(w_\delta(e))) &\geq 0, && (IR_\delta)\\
E_\phi(U^j(w_\phi(e))) &\geq 0, && (IR_\phi)\\
E_\delta(U^j(w_\delta(e))) &\geq E_\delta(U^j(w_\phi(e))), && (IC_\delta)\\
E_\phi(U^j(w_\phi(e))) &\geq E_\phi(U^j(w_\delta(e))), && (IC_\phi)\\
U^P(w_\delta^P, e_\delta^P) &\geq U^P(w_\delta^U, e_\delta^U), && (IC_{P,\delta})\\
U^U(w_\delta^U, e_\delta^U) &\geq U^U(w_\delta^P, e_\delta^P), && (IC_{U,\delta})\\
U^P(w_\phi^P, e_\phi^P) &\geq U^P(w_\phi^U, e_\phi^U), && (IC_{P,\phi})\\
U^U(w_\phi^U, e_\phi^U) &\geq U^U(w_\phi^P, e_\phi^P). && (IC_{U,\phi})
\end{aligned}$$

She maximizes her expected profits with respect to two different contracts: $w_\delta = \{(w_\delta^P, e_\delta^P), (w_\delta^U, e_\delta^U)\}$ and $w_\phi = \{(w_\phi^P, e_\phi^P), (w_\phi^U, e_\phi^U)\}$. These contracts induce separation among belief types in period 1 and among productivity types in period 2. In order to achieve this, the contracts have to satisfy eight different constraints.

The first two are period 1 individual rationality constraints that ensure that each belief type is willing to accept the contract designed for him as opposed to his outside option. The second two are period 1 incentive compatibility constraints that induce separation among belief types. Notice that since these four constraints relate to period 1, they are expressed in expected utility terms, and the expectations are weighted by workers' beliefs.

The last two pairs of constraints are 'contract specific' (two for each contract w_i) period 2 incentive compatibility constraints. They ensure that belief type i, once he has self-selected in period 1 and learned his true productivity in period 2, chooses the wage/effort pair designed for him. Hence, they are expressed as the actual utility the worker obtains.

Notice that the principal does not have to satisfy any period 2 individual rationality constraint since it is assumed that workers cannot 'drop out' of the contract once it has been signed in period 1.

In the next sections, I solve the problem for the optimal set of contracts offered by the employer. I do this under different assumptions about the level of information obtained by knowing a worker's belief type. I start with the case of a perfect correlation between the two type dimensions, i.e. when beliefs are perfectly informative about workers' productivity. In this case, separation in period 1 perfectly reveals the agent's productivity.

Before doing that, however, in order to better express the results let me define the concepts of *exploitation* and *efficiency* in line with the terminology in the existing literature.

Definition 1 (Exploitation)**.** *A worker of belief type i and productivity type j is exploited if*

$$w_i^j - \theta_j e_i^j < 0.$$

That is, he is exploited if he accepts a contract $w_i(e)$ that a fully informed agent of his same productivity type would not accept.

The concept of exploitation was first introduced in Eliaz and Spiegler (2006).[10] It generally applies to a situation where a principal takes advantage of an agent's naïveté in order to extract a surplus from him beyond the limits of the IR. In the context of this paper, a worker may be exploited not because he does not know his true ability (although he does not), but rather because he has systematically wrong beliefs about it.

Second, I define efficient levels of effort as the values of e that equate marginal product to workers' productivity.[11]

Definition 2 (Efficient Effort)**.** *A worker of belief type i and productivity type j exerts* efficient *effort if*

$$e_i^j : y'(e_i^j) = \theta_j.$$

That is, if at e_i^j the marginal product of effort equals the worker's productivity.

Given this, a contract $w_i(e)$ may induce either productive or unproductive workers (or both) to exert efficient levels of effort. Hence, the (classical) definition of efficiency at the top and at the bottom:

Definition 3 (Top vs. Bottom Efficiency)**.** *A contract $w_i(e)$ features* efficiency at the top *if it induces productive workers to exert the efficient level of effort, i.e.*

[10] Some differences arise, however, in the source of exploitation, as will be described later on.

[11] These are the equilibrium effort levels exerted when agents are uninformed about their true type, but are not naïve (Harris and Raviv, 1979; Laffont and Martimort, 2002).

$e_i^P : y'(e_i^P) = \theta_P$. *It features* efficiency at the bottom *if it induces unproductive workers to exert the efficient level of effort, i.e.* $e_i^U : y'(e_i^U) = \theta_U$.

4. Perfectly Informative Beliefs

In this section, I study the case where knowing a worker's belief perfectly reveals his productivity type. This scenario requires that p_P and p_U belong to $\{0, 1\}$ and $p_P + p_U = 1$. That is, either all productive workers are optimistic and all unproductive workers are pessimistic, or vice versa. If all workers were pessimistic (or optimistic) regardless of their productivity, then no information could be learned by knowing their beliefs.

The first result shows that separation in period 1 is not affected by the extent, or direction, of belief-productivity correlation. This is because, at this stage, the employer has only a prior on workers' belief types and cannot exploit the correlation between beliefs and productivity. The next lemma identifies the binding constraints for period 1.

Lemma 1 (Period 1 Screening)**.** *Regardless of the correlation between naïveté and productivity, the period 1 constraints are such that:*

(i) (IR_δ) *binds while* (IR_ϕ) *is slack.*
(ii) (IC_ϕ) *binds while* (IC_δ) *is slack.*

Lemma 1 presents findings similar to a classical screening model. In this case, the optimistic type plays the role of the 'high type' while the pessimistic type the role of the 'low type'. To see this, notice that what determines period 1's type ranking — high vs. low — is not the workers' actual productivity, but rather their subjective expectations about it. Therefore, optimistic (pessimistic) workers play the role of the high (low) type in the population. As in classical screening problems, optimistic workers' IR is slack as is the IC of pessimistic types.

Notice also that the employer's only purpose in inducing separation in period 1 is to be able to form posteriors on workers' productivity, since she gains no direct profits from this separation.

Given the above, and substituting for the expected profits and utilities, the problem that the employer solves is reduced to:

$$\max_{\{w_j^i\}_{j=\delta,\phi,i=P,L}} \lambda \left[p_P(y(e_\delta^P) - w_\delta^P) + (1-p_P)(y(e_\phi^P) - w_\phi^P)\right] + \\ +(1-\lambda)\left[p_U(y(e_\delta^U) - w_\delta^U) + (1-p_U)(y(e_\phi^U) - w_\phi^U)\right] \tag{2}$$

$$\text{s.t.}\quad \delta(w_\delta^P - \theta_P e_\delta^P) + (1-\delta)(w_\delta^U - \theta_U e_\delta^U) = 0 \qquad (IR_\delta)$$

$$\phi(w_\phi^P - \theta_P e_\phi^P) + (1-\phi)(w_\phi^U - \theta_U e_\phi^U) = \phi(w_\delta^P - \theta_P e_\delta^P) + (1-\phi)(w_\delta^U - \theta_U e_\delta^U) \qquad (IC_\phi)$$

$$w_\delta^P - w_\delta^U \geq \theta_P(e_\delta^P - e_\delta^U) \qquad (IC_{P,\delta})$$

$$w_\delta^P - w_\delta^U \leq \theta_U(e_\delta^P - e_\delta^U) \qquad (IC_{U,\delta})$$

$$w_\phi^P - w_\phi^U \geq \theta_P(e_\phi^P - e_\phi^U) \qquad (IC_{P,\phi})$$

$$w_\phi^P - w_\phi^U \leq \theta_U(e_\phi^P - e_\phi^U). \qquad (IC_{U,\phi})$$

As expected, period 2 separation depends on the direction (and extent) of the correlation between beliefs and productivity. The reason for this is that the period 2 ICs are contract-specific. Hence, their relevance depends on the posterior the employer forms on a specific belief type's productivity.

I now present the results under two different scenarios of perfect correlation. In the first scenario, I study a *positive correlation* between beliefs and productivity, that is, the case of optimistic-productive and pessimistic-unproductive workers. In the second scenario, I study the opposite case of *negative correlation*. These scenarios both have specific features, which are relevant only in the special case of this section, and more general features which will be revisited in section 5, where correlation is assumed to be imperfect.

4.1. **Perfect Positive Correlation.** In this section I study the case of $p_P = 0$ and $p_U = 1$. Hence, the only types of agent present in the labour force are productive optimistic, (P, ϕ), and unproductive pessimistic, (U, δ). While the employer understands this and behaves accordingly, workers still believe they are part of a population with four different types of workers. That is, they fail to understand that their beliefs are a perfect indicator of their actual productivity. This creates an opportunity for the employer to take advantage of workers' naïveté and exploit them.

Take the contract for pessimistic workers, for example. The employer sets the contracts in period 1, when workers are unaware of their true productivity. While she knows, however, that every pessimistic worker is unproductive, the latter believe themselves to be productive with a positive probability. This creates two 'channels' of exploitation for the employer to use.

First, when facing a contract that extracts a full surplus from unproductive types, a pessimistic unproductive worker expects to obtain positive utility from it. To see this, consider any contract for pessimistic types that only offers $w_\delta(e_\delta^U) = \theta_U e_\delta^U$. In period 1, an unproductive pessimistic worker evaluates this contract with $E_\delta(U_i(w_\delta(e))) = \theta_U e_\delta^U - E_\delta(\theta) e_\delta^U > 0$. Given this, the employer can decrease the wage even further, increasing profits, until the (IR_δ) binds. A this point, the worker in period 1 expects to obtain zero utility, while in period 2 he faces the truth and obtains negative utility. He is exploited through the *first channel*.

The *second channel* of exploitation takes place through an 'imaginary offer' (Eliaz and Spiegler, 2006).

Definition 4 (Imaginary Offer)**.** *An* imaginary offer *is a pair* (w_i^j, e_i^j) *never chosen by any worker in equilibrium, but that workers believe they will choose with positive probability because of their naïveté.*

The imaginary offer satisfies incentive compatibility and it is used by the employer to increase the expected utility of a worker from contract $w_i(e)$, while not increasing the actual utility he obtains. In this section, the imaginary offer in the contract for the pessimistic worker is set to yield a positive surplus to a productive type. If it is added to contract $w_\delta(e)$, in fact, the pessimistic worker assigns a positive probability to the event of choosing it (and of being a productive type). This increases his expected utility from contract $w_\delta(e)$ and allows the employer to decrease the utility given by the 'actual offer' even further, increasing exploitation.

To avoid any confusion, notice that imaginary offers do not affect profits directly but only through workers' naïveté. In other words, the employer knows that these offers are never chosen and hence they are assigned no positive probability in the expectations of profits. Since workers, however, believe they may choose these offers in period 2 with some positive probability, the equilibrium values of the actual offers depend on the imaginary offers. Hence, their effect on profits is indirect.

Given the two channels, I define two possible levels of exploitation.

Definition 5 (Mild vs. Strong Exploitation)**.** *Exploitation is* mild *if the employer does not take advantage of the second channel — i.e. she does not design an imaginary offer. It is* strong *if she does so.*

In the case of a positive perfect correlation between type dimensions, it is straightforward to understand that the period 2 binding constraints are $(IC_{P,\phi})$ and $(IC_{U,\delta})$. This is because they are intended for the only types that actually exist in the population. Nevertheless, in order to take advantage of the two channels of exploitation, $(IC_{U,\phi})$ and $(IC_{P,\delta})$ should still hold.[12]

Hence, when belief and productivity are perfectly positively correlated, the employer solves:

$$\max_{\{w_i^j\}_{i=\delta,\phi,j=P,L}} \lambda(y(e_\phi^P) - w_\phi^P) + (1-\lambda)(y(e_\delta^U) - w_\delta^U) \tag{3}$$

[12]If they did not, agents would not assign a positive probability to choosing the imaginary offer in period 2.

$$\text{s.t.}\quad \delta(w_\delta^P - \theta_P e_\delta^P) + (1-\delta)(w_\delta^U - \theta_U e_\delta^U) = 0 \qquad (IR_\delta)$$

$$\phi(w_\phi^P - \theta_P e_\phi^P) + (1-\phi)(w_\phi^U - \theta_U e_\phi^U) = \phi(w_\delta^P - \theta_P e_\delta^P) + (1-\phi)(w_\delta^U - \theta_U e_\delta^U) \qquad (IC_\phi)$$

$$w_\delta^P - w_\delta^U = \theta_U(e_\delta^P - e_\delta^U) \qquad (IC_{U,\delta})$$

$$w_\phi^P - w_\phi^U = \theta_P(e_\phi^P - e_\phi^U) \qquad (IC_{P,\phi})$$

$$w_\phi^P - w_\phi^U \leq \theta_U(e_\phi^P - e_\phi^U) \qquad (IC_{U,\phi})$$

$$w_\delta^P - w_\delta^U \geq \theta_P(e_\delta^P - e_\delta^U). \qquad (IC_{P,\delta})$$

First of all, notice that period 2 incentive compatibility for contract $w_i(e)$ is possible as long as $(e_i^P - e_i^U) \geq 0$. Once again, recall that this incentive compatibility is only imaginary since there are no workers with different productivity but thesame belief type.

Solving the binding constraints for $w_\delta^P, w_\delta^U, w_\phi^P, w_\phi^U$,

$$\begin{aligned}
w_\delta^U &= E_\delta(\theta)e_\delta^P + \theta_U(e_\delta^U - e_\delta^P),\\
w_\delta^P &= E_\delta(\theta)e_\delta^P,\\
w_\phi^U &= (E_\delta(\theta) - E_\phi(\theta))e_\delta^P + E_\phi(\theta)e_\phi^U,\\
w_\phi^P &= (E_\delta(\theta) - E_\phi(\theta))e_\delta^P + E_\phi(\theta)e_\phi^U - \theta_P(e_\phi^U - e_\phi^P),
\end{aligned}$$

and substituting the relevant solutions in the maximisation, I obtain:

$$\max_{\{e_j^i\}_{j=\delta,\phi,i=P,L}} \lambda\left(y(e_\phi^P) - (E_\delta(\theta) - E_\phi(\theta))e_\delta^P - E_\phi(\theta)e_\phi^U - \theta_P(e_\phi^P - e_\phi^U)\right) + \\ +(1-\lambda)\left(y(e_\delta^U) - E_\delta(\theta)e_\delta^P - \theta_U(e_\delta^U - e_\delta^P)\right). \qquad (4)$$

From this new problem, the actually chosen levels of effort e_ϕ^P and e_δ^U are easily calculated to be $y'(e_\phi^P) = \theta_P$ and $y'(e_\delta^U) = \theta_U$. Hence, each worker hired exerts the efficient level of effort for his productivity type. This result is common to the case of negatively correlated beliefs and productivity, and it is generalised in Proposition 1 in the next section.

The values of the imaginary offers can also be derived from the maximisation problem. Starting from the effort level for unproductive optimistic workers, it is easy to see that the effect of e_ϕ^U on profits is negative. Hence, in equilibrium $e_\phi^U = 0$. The intuition behind this is that the (w_ϕ^P, e_ϕ^P) offer is already inducing an efficient level of effort. Since the agent believes himself to be an unproductive type with some positive probability, the imaginary action has to require a *low* level of effort. In this way, the worker feels 'safe' that if she turns out to be unproductive, she can always enjoy a small surplus without exerting too much effort; in fact, no effort at all: $e_\phi^U = 0$.[13]

[13]To see that $U_\phi^U > 0$ notice that $e_\delta^P > 0$ as described below. Hence, $w_\phi^U > 0$ even if $e_\phi^U = 0$.

The intuition behind the optimal value for e^P_δ, and its derivation, instead, are not so straightforward. On the one hand, a lower e^P_δ for a given w^P_δ increases $E_\delta(U_i(w_\delta(e)))$. This relaxes the IR of the pessimistic unproductive worker, and allows the employer to decrease even further the wage paid to this type, increasing exploitation and profits. On the other hand, this also increases $E_\phi(U_i(w_\delta(e)))$, violating the (IC) of the optimistic productive worker. This forces the employer to increase $E_\phi(U_i(w_\phi(e)))$ by the same amount, decreasing her profits. Which of these two opposite effects prevails depends on the effect of e^P_δ on (4). If it is positive, e^P_δ is set to the highest possible value, 1. If the effect is negative, e^P_δ is set to the lowest possible value. Finally, however, notice that for incentive compatibility to be possible — i.e. for $(IC_{U,\delta})$ and $IC_{P,\delta})$ to hold, e^P_δ cannot go below the value of e^U_δ. Hence, whether the effect of e^P_δ on profits is positive or negative does not determine the 'direction' of the imaginary offer but rather whether the offer exists or not. In other words, it determines whether the pessimistic worker expects to be screened or pooled in period 2.

The effect of a decrease in e^P_δ on profits depends on the ratio of workers' beliefs, i.e. the overall level of naïveté of the worker population. In particular, the greater the naïveté of the optimistic productive worker, the more he believes he is unproductive. Hence, the lower the positive effect on $E_\phi(U_i(w_\delta(e)))$ of a decrease in e^P_δ, the stronger is the second channel of exploitation of him. This allows the principal to let (IC_ϕ) bind again via (w^U_ϕ, e^U_ϕ), which does not directly affect her profits. Similarly, if ϕ is high and the productive type is self-aware — i.e. $E_\phi(\theta)$ is close to θ_P — (w^U_ϕ, e^U_ϕ) has a weaker effect on $E_\phi(U_i(w_\phi(e)))$. Therefore, decreasing e^P_δ becomes more costly. Hence, for the use of the second channel to be optimal, the pessimistic unproductive worker has to be sufficiently naïve.

This is summarised in Result 1.

Result 1 (Pooling of Pessimistic Workers)**.** *When beliefs and productivity are perfectly positively correlated, if a pessimistic unproductive worker is naïve enough, relative to the self-awareness of a optimistic productive worker, that is:*

$$\frac{\delta}{\phi} \geq \lambda, \tag{5}$$

then the employer uses an imaginary offer (w^P_δ, e^P_δ) for the unproductive type and exploitation is strong. If not, the employer uses no imaginary offer and exploitation is mild.

To fully understand Result 1, consider the following. Notice that the LHS of condition (5) corresponds to the naïveté of pessimistic unproductive workers over the self-awareness of optimistic productive workers. Hence, it can be interpreted as a measure of workers' relative naïveté in the population. The higher it is, the more the unproductive type believes himself to be productive and the more the productive worker believes himself to be unproductive. For the condition to hold, the probability a pessimistic unproductive worker assigns to himself of being productive has to be larger than the proportion of productive workers in the population. When condition (5) holds, $e_\delta^P = 1$ and $w_\delta^P = E_\delta(\theta)$ and exploitation is strong.

On the contrary, when condition (5) does not hold, either the optimistic productive worker is too self-aware or the pessimistic unproductive worker is not naïve enough. The latter's relative naïveté is lower than the proportion of productive workers in the population and it is, therefore, too costly to exploit him through the second channel. No imaginary offer is set in w_δ and pessimistic workers expect to be pooled in period 2.

Given the structure of the contracts described so far, the next result concerns workers' welfare.

Result 2 (Exploitation with Perfect Positive Correlation)**.** *When beliefs and productivity are perfectly positively correlated, productive workers enjoy a positive rent while unproductive ones are exploited. The rent enjoyed by the former is larger when exploitation of the latter is strong.*

The intuition behind Result 2 follows from the previous discussion. The employer wants to take advantage of the unproductive workers' naïveté. Since beliefs and productivity are perfectly positively correlated, however, she is forced to leave some positive surplus to the productive worker. The reason for this lies both in the difference in beliefs between belief types and in the way the two channels of exploitation affect the two contracts.

On the one hand, the first channel cannot be used to extract a surplus from the productive worker. The reason for this is that the latter expects to have a lower productivity than he actually has: he never accepts a contract that extracts a full surplus from a productive type because he would expect to get a strictly negative utility from doing so.

On the other hand, the role of the imaginary offer in the contract for the optimistic type is not to extract more surplus from him but rather to provide him with a form of 'insurance'. In other words, it represents a safe option for the optimistic productive worker to choose in the case that he turns out to be unproductive — an event that he deems possible with positive probability. Hence, the employer has no ability to exploit productive workers.

Finally, to understand the intuition behind optimistic workers' surplus, notice that an optimistic worker always assigns a larger probability to obtaining U_δ^P if he chooses $w_\delta(e)$ than the pessimistic worker. Hence, any change to $w_\delta(e)$ that increases U_δ^P keeping $E_\delta(U_j(w_\delta(e)))$ constant, increases $E_\phi(U_j(w_\delta(e)))$. For (IC_ϕ) to bind, therefore, the utility from the contract for the optimistic type has to increase when the imaginary offer is added to $w_\delta(e)$.

Notice that this is in accordance with the intuition behind (5). The greater is the general level of naïveté in the labour force population — i.e. (5) holds — the stronger is the exploitation of the unproductive worker.

4.2. **Perfect Negative Correlation.** In this section, I study the opposite case to that of section 4.1, namely when $p_P = 1$ and $p_U = 0$. This is the case in which all productive workers are pessimistic and all unproductive workers are optimistic.

To understand the framework I have in mind, consider the case of a population of newly graduated students looking for their first job. Grades and degrees can explain a lot about knowledge of topics and intelligence, but when it comes to innate ability, speed of adaptation, productivity and so on, there is nothing like true practice to give an indication of one's capabilities. Suppose that the students have a degree in financial economics and they are all looking for a job in the financial sector. They all apply for jobs according only to their expectations about their own productivity. Once a job is obtained, however, they learn their true productivity and choose the amount of effort to exert in the job accordingly. Some of the students have a passion for finance: they read the news, understand the mechanics and complexities of markets and would be perfect for a job in the financial sector (productive types). Others, instead, have chosen that specific course of study without having a deep interest in financial markets. Hence, they would be a less perfect match for a financial firm (unproductive types). In this section, I assume that understanding the complexities of the job and the mechanisms of financial markets without having a clear perception of one's own capability hurts self-confidence and creates a generally pessimistic feeling about one success in the market (pessimistic productive workers). A candidate who does not comprehend these complexities, instead, has a relatively arrogant attitude. He is convinced that the job will be easy (optimistic unproductive workers).

To derive the solution to this problem, it is easy to follow the same procedure as in section 4.1 but now, however, the period 2 binding constraints are $(IC_{P,\delta})$ and $(IC_{U,\phi})$. Hence, the problem becomes:

$$\max_{\{w_i^j\}_{i=\delta,\phi,j=P,L}} \lambda(y(e_\delta^P) - w_\delta^P) + (1-\lambda)(y(e_\phi^U) - w_\phi^U) \tag{6}$$

$$\text{s.t.} \quad \delta(w_\delta^P - \theta_P e_\delta^P) + (1-\delta)(w_\delta^U - \theta_U e_\delta^U) = 0 \qquad (IR_\delta)$$

$$\phi(w_\phi^P - \theta_P e_\phi^P) + (1-\phi)(w_\phi^U - \theta_U e_\phi^U) = \phi(w_\delta^P - \theta_P e_\delta^P) + (1-\phi)(w_\delta^U - \theta_U e_\delta^U) \qquad (IC_\phi)$$

$$w_\phi^P - w_\phi^U = \theta_U(e_\phi^P - e_\phi^U) \qquad (IC_{U,\phi})$$

$$w_\delta^P - w_\delta^U = \theta_P(e_\delta^P - e_\delta^U) \qquad (IC_{P,\delta})$$

$$w_\delta^P - w_\delta^U \leq \theta_U(e_\delta^P - e_\delta^U) \qquad (IC_{U,\delta})$$

$$w_\phi^P - w_\phi^U \geq \theta_P(e_\phi^P - e_\phi^U). \qquad (IC_{P,\phi})$$

Solving the binding constraints for $w_\delta^P, w_\delta^U, w_\phi^P, w_\phi^U$,

$$w_\delta^U = E_\delta(\theta)e_\delta^U, \tag{7}$$

$$w_\delta^P = E_\delta(\theta)e_\delta^U - \theta_P(e_\delta^U - e_\delta^P), \tag{8}$$

$$w_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^U + E_\phi(\theta)e_\phi^P + \theta_U(e_\phi^U - e_\phi^P), \tag{9}$$

$$w_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^U + E_\phi(\theta)e_\phi^P, \tag{10}$$

and substituting the relevant solutions in the maximisation, I obtain:

$$\begin{aligned} \max_{\{e_j^i\}_{j=\delta,\phi,i=P,L}} \quad & \lambda(y(e_\delta^P) - E_\delta(\theta)e_\delta^U + \theta_P(e_\delta^U - e_\delta^P))+ \\ & (1-\lambda)(y(e_\phi^U) - (E_\delta(\theta) - E_\phi(\theta))e_\delta^U - E_\phi(\theta)e_\phi^P - \theta_U(e_\phi^U - e_\phi^P)). \end{aligned} \tag{11}$$

From the maximisation problem, the chosen levels of effort e_δ^P and e_ϕ^U correspond to $y'(e_\delta^P) = \theta_P$ and $y'(e_\phi^U) = \theta_U$. As in section 4.1, all workers exert efficient levels of effort.

Proposition 1 (Full Efficiency)**.** *When beliefs and productivity are perfectly correlated, full efficiency is always achieved regardless of the direction of the correlation. That is, both productive and unproductive workers choose first-best levels of effort, regardless of their beliefs.*

If workers were naïve but the correlation between beliefs and productivity were not perfect, then even after updating her beliefs, the employer would not be able to precisely tell the productivity type of a worker. Hence, she would assign positive probability to all possible combinations of belief and productivity types. I derive the equilibrium for this case in section 5.

To derive the equilibrium values of imaginary offers in the case of a perfect negative correlation, notice from (11) that the effect of e_δ^U is always negative and the effect of e_ϕ^P is always positive. Hence, $e_\delta^U = 0$ while $e_\phi^P = 1$.[14]

To see why the effort level for the optimistic productive type has positive effects on profits, simply notice that in this framework optimistic workers are always unproductive. Hence, the second channel of exploitation is more powerful than ever

[14]This ensures that contracts are incentive compatibile.

and the employer uses an imaginary offer that grants the largest possible incentive compatible surplus to a hypothetical optimistic productive type. Also note that since the actual productive worker is always pessimistic he assigns to this offer a much smaller weight than the optimistic productive worker when evaluating $w_{\phi}(e)$.

As for the effort level of the pessimistic unproductive type, the intuition is unchanged from section 4.1.

Given all the above, I can derive the equivalent of Result 2 for the case of perfectly negatively correlated types.

Result 3 (Exploitation with Perfect Negative Correlation)**.** *When beliefs and productivity are perfectly negatively correlated, pessimistic productive workers obtain zero surplus while optimistic unproductive ones are exploited. Exploitation is always strong.*

Result 3 shows a peculiar feature of this case: pessimistic productive workers enjoy no positive rent. Their pessimistic naïveté is large enough to allow the employer to extract all their surplus, but not large enough for them to be exploited. This is because in period 1 pessimistic productive workers play the role of the *low type* — they obtain zero expected utility. Hence, the employer can extract a full surplus from them by offering a contract that ensures zero utility to both productivity types.

Optimistic unproductive types, on the other hand, can be screened away from $w_{\delta}(e)$ with the promise of a higher utility in the event of being productive and a lower one in the event of being unproductive, that is, by introducing an imaginary offer that grants positive utility to productive types and negative utility to unproductive types.

In other words, while in section 4.1 (IC_{ϕ}) acts as a 'proper constraint' on the level of exploitation of the unproductive type, here it acts as a means of exploitation. It is through (IC_{ϕ}) that the employer can separate the unproductive type and take advantage of his naïveté.

To conclude this section, I present a corollary to Results 2 and 3 that compares welfare findings.

Corollary 1 (Perfect Correlation Welfare)**.** *When beliefs and productivity are perfectly correlated, both productive and unproductive workers obtain lower utility when the correlation is negative.*

Corollary 1 is quite intuitive. When the correlation is negative, naïveté plays a much larger role and the productive worker loses all potential information rent. The employer uses the two channels of exploitation to their maximum effect.

To compare these findings with classical results, notice that if workers were not naïve but formed homogeneous unbiased expectations about their productivity (also known as the "selling of the firm" equilibrium, Laffont and Martimort, 2002; Harris and Raviv, 1979, where the worker becomes the residual claimant of the firm's profits), the employer would not be able to strongly exploit workers, but would still be able to achieve full efficiency. If instead workers were fully informed about their true productivity, the classical screening literature tells us that efficiency would only be achieved at the top and that productive workers would enjoy positive rents.

Hence, an agent's imperfect information about his productivity allows the employer to exploit agents. Naïveté on its own allows her to use the second channel of exploitation at the cost of full efficiency. A perfect correlation between beliefs and productivity enables her to achieve full efficiency and, if the correlation is negative, to extract all of the surplus from the productive types while still strongly exploiting unproductive workers.

5. Imperfectly Informative Beliefs

In this section, I study the general case where beliefs are imperfectly informative. More precisely, both productive and unproductive workers have a positive probability of being either optimistic or pessimistic, i.e. $p_P, p_U \notin \{0, 1\}$.[15]

As described in section 2, the basic intuition from the literature on contracting with naïve agents when there is no correlation between beliefs and productivity shows that the principal designs contracts that induce efficient effort in states — productivity types in my model — that agents deem less likely than the principal does (Eliaz and Spiegler, 2006, 2008). Exploitation also takes place in these states. Hence, in my model, efficiency would be at the top in the contract for the pessimistic worker and at the bottom in that for the optimistic worker.

After the introduction of type-dependent naïveté, however, this result carries over only partially and a new one holds for part of the parameter space. As I show in this section, the employer may find it optimal to induce a worker of belief i and productivity type j to exert the efficient level of effort, not because of a misalignment of beliefs between her and the worker but rather because of her posterior that a belief type i is indeed a productivity type j.

When beliefs are imperfectly informative, the employer has a prior that assigns positive probability to each possible combination of beliefs and productivity type.

[15]Notice that the case where the correlation between beliefs and productivity is perfect for only one productivity/belief pair is not analysed in the paper. Solutions for these cases, however, can be derived by combining the findings of this section with those of section 4. They present no further insights into the employer's optimal contracting behaviour.

Hence, she solves problem (2). Period 2 incentive compatibility this time, however, is not as straightforward as before.

First of all, which incentive compatibility constraint is binding depends on the posteriors of the employer, given that workers have self-selected in period 1.

Second, notice that, as in classical screening problems, if $(IC_{j,i})$ binds then the contract designed for an belief type i induces the efficient level of effort. Hence, deriving conditions for the IC constraints binding in period 2 also indicates whether efficiency for belief type i is at the top or at the bottom. I start with the contract designed for optimistic workers.

Result 4 (Efficiency for Optimistic Workers)**.** *If the employer has a strong updated belief that optimistic workers are productive, or the naïveté of unproductive optimistic workers is low enough, efficiency is at the top in the contract for optimistic workers. That is, if*

$$\Pr\{\theta = \theta_P | \mathcal{B} = \phi\} \tfrac{1-\phi}{\phi} \geq \Pr\{\theta = \theta_U | \mathcal{B} = \phi\}, \tag{12}$$

then $y'(e^P_\phi) = \theta_P$*.*

Result 4 (together with Result 5 below) represents the new trade-off that the principal faces on the use of the information granted by the type-dependancy of agents' naïveté.

The employer has two main objectives: to induce efficient levels of effort in workers — maximising the pie — and to extract as much surplus as she can — taking the pie away from the agents. The second is, of course, easier to achieve when facing agents with particularly wrong beliefs about themselves since, as shown in section 4, naïveté increases exploitation. When considering optimistic workers, these types are the unproductive ones and the ratio $\frac{\phi}{1-\phi}$ (which is increasing in ϕ) then becomes a measure of their naïveté. Its inverse measures the *self-awareness* of optimistic unproductive workers. This makes the intuition behind condition (12) easy to identify.

In order to maximise the pie, the principal would like to induce efficient effort in the type of worker she believes she is facing. Hence, she compares her posteriors about a worker being productive or unproductive given that he is optimistic. When the optimistic unproductive worker is naïve enough, however, inducing him to exert efficient effort becomes less costly. This creates a further incentive to set $y'(e^U_\phi) = \theta_U$. In order for efficiency at the top to be optimal in the contract for optimistic workers then, it is not enough for the monopolist to believe that an optimistic worker is most probably (in posterior terms) productive. It is also necessary for optimistic unproductive workers to be relatively self-aware (i.e. ϕ has to be low enough).

A graphical intuition for Result 4 is represented in Figure 1. In the figure constraints (IC_ϕ), $(IC_{U,\phi})$ and $(IC_{P,\phi})$ are plotted, together with isoprofits, in (w_ϕ^U, w_ϕ^P) space. Condition (12) holds in the graph on the right and fails in the graph on the left. Notice that profits increase towards the bottom left in each graph and that incentive compatible (both for period 1 and period 2) contracts lie in the area above (IC_ϕ) and between $(IC_{U,\phi})$ and $(IC_{P,\phi})$.

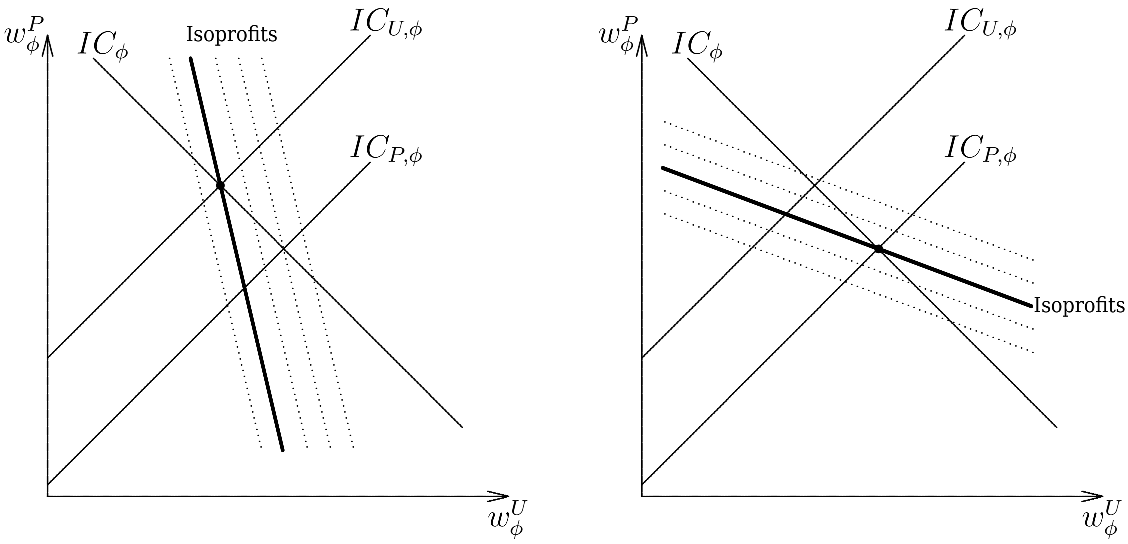

Figure 1. Optimistic Workers' Efficiency

In the figure (IC_ϕ), $(IC_{U,\phi})$, $(IC_{P,\phi})$ and isoprofits are plotted in (w_ϕ^U, w_ϕ^P) space. When condition (12) fails — left-hand graph — isoprofits are steeper than (IC_ϕ). When condition (12) holds — right-hand graph — they are flatter than (IC_ϕ). Profits of the employer increase towards the bottom left of the graphs.

In the graph on the left, the posterior of the employer on an optimistic worker being unproductive is strong. Hence, an increase in w_ϕ^U bites more on profits than the same increase in w_ϕ^P. Hence, the isoprofits are steeper than the (IC_ϕ) constraint and efficiency is at the bottom in the contract for optimistic workers. In the graph on the right, the opposite intuition applies.

A similar result is true for pessimistic workers:

Result 5 (Efficiency for Pessimistic Workers)**.** *If the employer has a strong updated belief that pessimistic workers are unproductive, or the naïveté of productive pessimistic workers is low enough, efficiency is at the bottom in the contract for pessimistic worker. That is, if*

$$\Pr\{\theta = \theta_U | \mathcal{B} = \delta\} \tfrac{\delta}{1-\delta} \geq \Pr\{\theta = \theta_P | \mathcal{B} = \delta\}, \tag{13}$$

then $y'(e_\delta^U) = \theta_U$*.*

Condition (13) is the mirror image of (12) for pessimistic workers. Notice that the naïveté of pessimistic productive workers is measured by $\frac{1-\delta}{\delta}$, which is decreasing in δ. Its inverse measures their self-awareness. The lower δ the larger the naïveté of productive pessimistic workers. Figure 2 below shows a similar graphical intuition to the one in Figure 1.

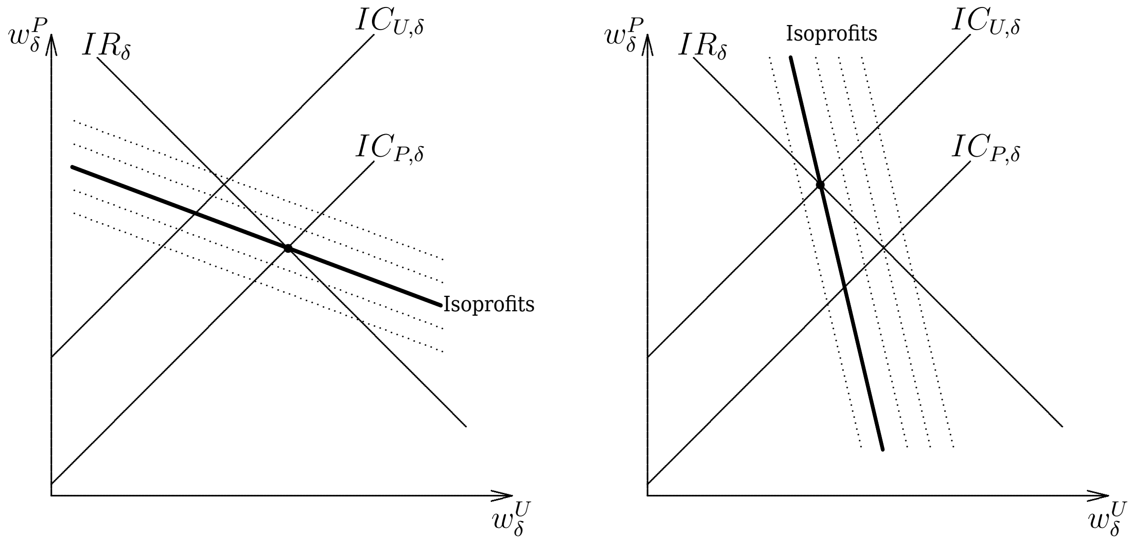

Figure 2. Pessimistic Workers' Efficiency
In the figure (IR_δ), $(IC_{U,\delta})$, $(IC_{P,\delta})$ and isoprofits are plotted in (w_δ^U, w_δ^P) space. When condition (13) fails — left-han graph — isoprofits are flatter than (IR_δ) and efficiency is at the top in $w_\delta(e)$. When condition (13) holds — right-hand graph — they are steeper than (IR_δ) and efficiency is at the bottom. The employer's profits increase towards the bottom left of the graphs.

To fully understand the importance of the correlation between beliefs and productivity, notice that, if they were perfectly independent conditions (12) and (13) would become $\frac{1-\lambda}{\lambda} \leq \frac{1-\phi}{\phi}$ and $\frac{\lambda}{1-\lambda} \leq \frac{\delta}{1-\delta}$ respectively. By assumption, these would never hold. Since the employer would gain no information from screening in period 1, she would focus on extracting surplus and inducing the most naïve workers in the population to exert efficient effort — following the intuition from Eliaz and Spiegler (2006, 2008).

Ultimately, conditions (12) and (13) define the equilibrium of the model. Together, they determine the optimal behaviour of the employer and identify which of the two competing effects (exploiting workers' naïveté vs. inducing efficiency in the most common productivity type) dominates. I present this result in Proposition 2 and then analyse it in the (p_U, p_P) space for given values of δ, λ and ϕ.

Proposition 2 (Imperfect Correlation Efficiency). *Given the correlation between beliefs and productivity, the principal uses the information provided by period 1 screening in order to:*

- *maximise efficiency — efficiency is at the top (bottom) in the contract for optimistic (pessimistic) workers — when both conditions (12) and (13) hold;*
- *better exploit the most naïve agents in the market — efficiency is at the bottom (top) in the contract for optimistic (pessimistic) workers — when both conditions fail.*

Proposition 2 describes the main trade-off faced by the employer and states the main contribution of the paper. It follows from combining Results 4 and 5. To understand the proposition, consider Figure 3.

First of all, notice that two further cases are implied by Proposition 2. When condition (12) holds and (13) fails, efficiency is at the top in both contracts, while in the opposite case, efficiency is at the bottom in both contracts. These are intermediate situations where the correlation between beliefs and productivity is such that the result of the trade-off is different for the two contracts.

The basic parameters of the model are $\delta, \lambda, \phi, p_P$ and p_U. The first three simply describe the relation between optimistic and pessimistic workers and the proportion of productive types in the population. The last two, instead, are the focus of the paper and determine the level of information about productivity obtained by knowing a worker's beliefs, i.e. the extent of naïveté type-dependence. In Figure 3, I assume $\delta = \frac{1}{3}, \lambda = \frac{1}{2}$ and $\phi = \frac{2}{3}$ and study the equilibrium of the model in (p_U, p_P) space.

The $45°$ line in the graph separates the area of positive correlation below the line, from that of negative correlation above the line. It also corresponds to a model without type-dependent naïveté.

In area A, optimal contracts feature efficiency at the top for the pessimistic worker and at the bottom for the optimistic one. Corollary 2 shows that this area always occupies the entire portion of the parameter space where beliefs and productivity are negatively correlated for every $\delta < \lambda < \phi$. This is perfectly in line with the findings in the literature, discussed in section 2, and agrees with the intuition that employers induce workers with wrongly biased beliefs to exert efficient effort, while distorting offers for more self-aware types. Notice that if the type dimensions were independent, p_P and p_U would be equal. Hence, the $45°$ line would be the parameter space and area A would characterise all the equilibria.

Corollary 2 (Efficiency with Negative Correlation). *When beliefs and productivity are negatively correlated, the contract for pessimistic workers features efficiency at the top while that for optimistic workers features efficiency at the bottom.*

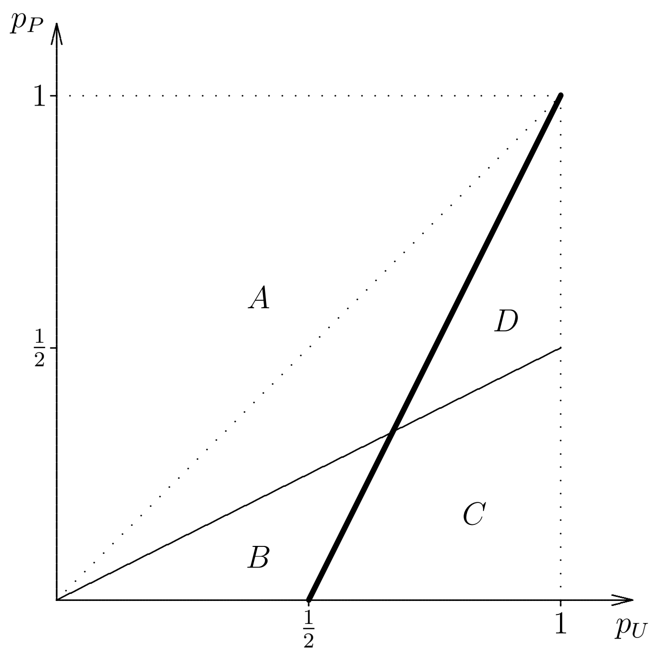

Figure 3. Efficiency in Optimal Contracting
In the figure, conditions (12) and (13) are plotted in (p_U, p_P) space. To the left of the bold line, condition (12) fails and the contract for optimistic workers features efficiency at the bottom. To the right of it the condition holds and the contract features efficiency at the top, instead. Above the thin line, condition (13) fails and the contract for pessimistic workers features efficiency at the top. Below it, the condition holds and the contract features efficiency at the bottom, instead.

This corollary is proven by studying conditions (12) and (13) when $p_P > p_U$. When beliefs and productivity are negatively correlated, the cost of extracting efficient levels of effort from optimistic unproductive workers and pessimistic productive workers is low. It is so low, in fact, that even when the chances of meeting such types are low (according to her updated beliefs) the employer still finds it optimal to extract a surplus from these workers nevertheless.[16]

The area below the $45°$ line is divided in four zones: a portion of area A; area B, where efficiency is at the bottom in both contracts; area C, where efficiency is at the top for optimistic and at the bottom for pessimistic workers; and area D, where efficiency is at the top for both belief types.

Consider a transition from area A to area B. The naïveté parameters remain unchanged from area A. What does change, however, is the *expected* naïveté of a pessimistic worker, since the probability of meeting a pessimistic worker who is

[16]Notice that a negative correlation between belief and probability happens in the area of the graph above the $45°$ line. Hence, however small, there is a strictly positive probability that these types exist.

productive is relatively low. When p_P is small, the number of pessimistic workers that turn out to be productive is low. Hence, after the screening of period 1, the employer updates her priors and induces pessimistic unproductive workers to exert efficient effort rather than the pessimistic productive ones. In other words, the chances that a pessimistic worker is productive are so small that the benefits of extracting efficient effort from such a type are negligible.

Contracts in area D follow from the exact opposite intuition. The employer's posterior on facing an unproductive worker given that the latter is optimistic are very small. Hence, she designs a contract with efficiency at the top for optimistic workers.

Finally, area C represents the case where workers have beliefs that are strongly positively correlated with their productivity — strong enough, that for the assumed levels of naïveté, the principal uses the information from period 1 screening to design more efficient contracts. She therefore disregards (to some extent) the exploitation opportunities created by naïveté. In this area, the optimal contracts resemble, in efficiency terms, those in section 4.1.

In Appendix A, I derive the optimal contracts for all the possible cases described. Given the solutions found, the next result studies the case of pessimistic types bunching, while the following corollary studies workers' welfare.

First of all, although in the case of imperfectly informative beliefs there are no imaginary offers, the offers set for workers whose period 2 IC are not binding play a similar role. If the employer wants to exploit the pessimistic unproductive type, for example, through offer (w_δ^P, e_δ^P), she is still capable of doing so. This time, however, offer (w_δ^P, e_δ^P) has a first-order direct effect on profits. This is because pessimistic productive workers *do* exist and the employer's priors on a worker being such a type are given by $p_P\lambda$. Hence, designing a contract that screens among differently productive pessimistic workers in period 2 may be suboptimal. Unlike section 4.1, this is *regardless* of the direction of the correlation between beliefs and productivities.

The higher the naïveté of pessimistic productive workers, relative to that of optimistic productive workers, the higher is the gain from using (w_δ^P, e_δ^P) for exploitation. When the proportion of optimistic workers is high, however, period 1 separation becomes too costly — in terms of the higher expected utility granted to optimistic workers.

Result 6 (Pooling of Pessimistic Workers)**.** *When beliefs and productivity are imperfectly correlated, if a pessimistic productive worker is naïve enough — relative to a optimistic unproductive one — or if the proportion of optimistic workers is* *small, that is*

$$\frac{\delta}{\phi} \geq \Pr\{\phi\}, \tag{14}$$

then the employer separates pessimistic workers on the basis of their productivity. Otherwise, they are bunched together.

Result 6 is reminiscent of Result 1. When the expected utility of $w_\delta(e)$ increases — as a consequence of a rise in the utility granted by (w_δ^P, e_δ^P) — in order to separate the optimistic workers from the pessimistic ones, the employer has to increase the expected utility coming from $w_\phi(e)$. This is regardless of the actual productivity of the optimistic workers. Furthermore, since the optimistic workers weight (w_δ^P, e_δ^P) more than the pessimistic ones, the increase in expected utility granted to optimistic workers can offset the profit gains from using (w_δ^P, e_δ^P) to exploit the pessimistic unproductive workers. This happens when condition (14) fails.

The next corollary shows that the qualitative results on workers' welfare are common to all four areas.

Corollary 3 (Imperfect Correlation Welfare)**.** *When beliefs are imperfectly informative about workers' productivity, unproductive workers are always exploited while productive workers enjoy a non-negative surplus.*

Corollary 3 follows from the proof of Lemma 1. Hence, the result is qualitatively unaffected by the direction and extent of the correlation between beliefs and productivity. Qualitative welfare results are unchanged regardless of the information gained from beliefs on productivity levels.

Since in period 1 the employer only has priors on agents' belief types, she cannot take advantage of the correlation between type dimensions. In period 2, however, she can update her beliefs and set offers that affect the extent of the exploitation of unproductive workers and the amount of surplus granted to productive ones.

6. Concluding Remarks

The purpose of this paper has been to study the optimal contracting behaviour of a principal who faces agents with type-dependent naÿvieté. In particular, I have focused on the role the information provided by the type dependency plays in the design of optimal contracts. There are two main implications of this new assumption.

First, when workers' beliefs are perfectly informative of their productivity, full efficiency is achieved. Unproductive workers are always exploited. Productive workers enjoy a positive surplus if their beliefs are positively correlated with their productivity. If the two are negatively correlated, instead, they enjoy no information rent.

The second main result highlights the trade-off faced by the principal in using the information gained by screening among different belief types in period 1, when the type dependency is not perfect. In this case, the principal designs contracts that extract efficient effort levels either from the most naïve workers in the population, maximising exploitation, or from the type of worker she deems she is most likely to face given her posteriors, maximising efficiency.

These findings connect the literature on sequential screening (Courty and Li, 2000, *inter alia*) with the literature on contracting with naïve agents (Eliaz and Spiegler, 2006, 2008) by assuming that agent naïveté depends on the agents' true nature (as in sequential screening problems).

Eliaz and Spiegler (2008) define a *speculative* contract as one that grants a worker with beliefs i expected utility of $\lambda(w_i^P - \theta_P e_i^P) + (1-\lambda)(w_i^U - \theta_U e_i^U) < 0$. In other words, a contract is speculative if it should not be signed by a worker with unbiased beliefs. In this paper, I have been able to prove that an optimistic (pessimistic) worker never (always) signs a speculative contract. This, however, does not save (condemn) him from (to) obtaining zero (negative) surplus.

This result follows from the assumption that one of the period 2 'states of the world', i.e. the two levels of productivity, dominates the other. In other words, for any level of effort e', the utility a productive worker obtains from $(w(e'), e')$ is always higher than the utility obtained by an unproductive worker. This assumption is not present in Eliaz and Spiegler (2006, 2008). The study of what would happen in a framework of type-dependent naïveté with unordered period 2 states is left for future work.

An extension of the present model with a continuum of belief types, and the assumption of heterogeneous distributions of beliefs among equally productive workers are work in progress. In addition, of interest for future research is a relaxation of the assumption on the perfect observability of e, introducing a moral hazard problem into the model.

References

Amador, M., I. Werning, and G. M. Angeletos (2006): "Commitment Vs. Flexibility," *Econometrica*, 2, 365–396.

Asheim, G. B. (2007): "Procrastination, Partial Naïvete, and Behavioral Welfare Analysis," Discussion paper, Memorandum, Department of Economics, University of Oslo.

Bagues, M., and M. J. Perez-Villadoniga (2012): "Do Recruiters Prefer Applicants with Similar Skills? Evidence from a Randomized Natural Experiment," *Journal of Economic Behavior & Organization*, 82, 12–20.

Banks, D. L., A. E. Lawson, and M. Logvin (2007): “Self-efficacy, Reasoning Ability and Achieving in College Biology,” *Journal of Research in Science Teaching*, 44(5), 706–724.

Banner, M., D. Dunning, J. Ehrlinger, J. Kerri, and J. Kruger (2008): “Why the Unskilled are Unaware: Further Exploration of (Absent) Self-insight among the Incompetent,” *Organizational Behavior and Human Decision Processes*, 105(1), 98–121.

Chi, M., T. Glaser, and R. Rees (1982): *Expertise in Problem Solving (in Advances in the Psychology of Human Intelligence)*. Erlbaum, Hillsdale, NJ.

Conlin, M., T. O’Donoghue, and T. J. Vogelsang (2007): “Projection Bias in Catalog Orders,” *The American Economic Review*, 97, 1217–49.

Courty, P., and H. Li (2000): “Sequential Screening,” *Review of Economic Studies*, 67, 697–717.

Deb, R., and M. Said (2015): “Dynamic Screening with Limited Commitment,” *Journal of Economic Theory*.

Dittrich, D. A., W. Güth, and B. Marciejovsky (2005): “Overconfidence in Investment Decisions: an Experimental Approach,” *The European Journal of Finance*, 11(6), 471–491.

Dunning, D., J. Ehrlinger, K. Johnson, and J. Kruger (2003): “Why People Fail to Recognize Their Own Incompetence,” *Current Directions in Psychological Science*, 12(3), 83–87.

Dunning, D., and J. Kruger (1999): “Unskilled and Unaware of it: How Difficulties in Recognising One’s Own Incompetence Lead to Inflated Self-assessments,” *Journal of Personality and Social Psychology*, 77, 1121–1134.

Eliaz, K., and R. Spiegler (2006): “Contracting with Diversely Naïve Agents,” *Review of Economic Studies*, 73, 689–714.

——— (2008): “Consumer Optimism and Price Discrimination,” *Theoretical Economics*, 3, 459–497.

Evans, R., and S. Reiche (2015): “Contract Design and Non-Cooperative Renegotiation,” *Journal of Economic Theory*, 157, 1159–1187.

Ferraro, P. J. (2010): “Know Thyself: Competence and Self-awareness,” *Atlantic Economic Journal*, 38, 183–196.

Gilpatric, M. S. (2008): “Present-bias Preferences, Self-awareness and Shirking,” *Journal of Economic Behaviour and Organisation*, 67, 735–754.

Grubb, M. D. (2015): “Selling to Overconfident Consumers,” *The American Economic Review*, 99, 1770–1807.

Harris, M., and A. Raviv (1979): “Optimal incentive Contracts with Imperfect Information,” *Journal of Economic Theory*, 20, 231–259.

Heidhues, P., and B. Köszegi (2010): "Exploiting Naïvete About Self-control in the Credit Market," *The American Economic Review*, pp. 2279–2303.

Kovác, E., and D. Krähmer (2013): "Optimal Sequential Delegation," *SFB/TR 15 Discussion Paper Series*, 427.

Laffont, J. J., and D. Martimort (2002): *The Theory of Incentives*. Princeton University Press, New Jersey, United States of America.

Lichtenstein, S., B. Fischoff, and L. D. Phillips (1982): *Calibration of Probabilities: The State of the Art to 1980*. Cambridge University Press, Cambridge, UK.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): "Projection Bias in Predicting Future Utility," *Quarterly Journal of Economics*, 118, 1209–48.

Moore, D. A., and P. J. Healy (2008): "The Trouble with Overconfidence," *Psychological Review*, 115(2).

O'Donoghue, T., and M. Rabin (2001): "Choice and Procrastination," *Quarterly Journal of Economics*, pp. 121–160.

Strotz, R. H. (1956): "Myopia and Inconsistency in Dynamic Utility Maximization," *The Review of Economic Studies*, 23(3), 165–180.

Svenson, O. (1981): "Are We Less Risky and More Skillful than Our Fellow Drivers?," *Acta Psachologica*, 47, 143–146.

Von Thadden, E.-L., and X. Zhao (2012): "Incentives for Unaware Agents," *Review of Economic Studies*, pp. 1–24.

Appendix A. Imperfectly Informative Beliefs — Optimal Contracts

In this appendix, I present the solutions of problem (2) for every value of p_P and p_U, that is, for all possible combinations generated by conditions (12) and (13). I also derive workers' utility in each equilibrium. Notice that the ranking and sign of workers' utility levels is proven in Lemma 1.

In what follows, I define U_i^j as the utility a worker of productivity type j and belief type i obtains at the end of the game: $U_i^j \equiv U_j(w_i^j, e_i^j) = w_i^j - \theta_j e_i^j$.

If conditions (12) and (13) fail together, the optimal contracts designed for area A are obtained by solving (2) with $(IC_{U,\phi})$ and $(IC_{P,\delta})$ binding. This results in:

$$w_\delta^U = E_\delta(\theta) e_\delta^U \tag{15}$$

$$w_\delta^P = E_\delta(\theta) e_\delta^U + \theta_P (e_\delta^P - e_\delta^U) \tag{16}$$

$$w_\phi^U = (E_\delta(\theta) - E_\phi(\theta)) e_\delta^U + E_\phi(\theta) e_\phi^P + \theta_U (e_\phi^U - e_\phi^P) \tag{17}$$

$$w_\phi^P = (E_\delta(\theta) - E_\phi(\theta)) e_\delta^U + E_\phi(\theta) e_\phi^P \tag{18}$$

and

$$e_\delta^U : y'(e) = \tfrac{E_\delta(\theta)-(1-E(p))E_\phi(\theta)-p_P\lambda\theta_P}{(1-\lambda)p_U} \tag{19}$$
$$e_\delta^P : y'(e) = \theta_P \tag{20}$$
$$e_\phi^U : y'(e) = \theta_U \tag{21}$$
$$e_\phi^P : y'(e) = \tfrac{(1-E(p))E_\phi(\theta)-(1-\lambda)(1-p_U)\theta_U}{(1-p_P)\lambda}. \tag{22}$$

This results in:

$$U_\delta^U = E_\delta(\theta)e_\delta^U - \theta_U e_\delta^U < 0 \tag{23}$$
$$U_\delta^P = E_\delta(\theta)e_\delta^U - \theta_P e_\delta^U > 0 \tag{24}$$
$$U_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^U - (\theta_U - E_\phi(\theta))e_\phi^P \leq 0 \tag{25}$$
$$U_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^U - (\theta_P - E_\phi(\theta))e_\phi^P > 0. \tag{26}$$

If condition (12) fails while (13) holds, the optimal contracts designed for area B are obtained by solving (2) with $(IC_{U,\phi})$ and $(IC_{U,\delta})$ binding. This results in:

$$w_\delta^U = E_\delta(\theta)e_\delta^P - \theta_U(e_\delta^P - e_\delta^U) \tag{27}$$
$$w_\delta^P = E_\delta(\theta)e_\delta^P \tag{28}$$
$$w_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P + E_\phi(\theta)e_\phi^P + \theta_U(e_\phi^U - e_\phi^P) \tag{29}$$
$$w_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P + E_\phi(\theta)e_\phi^P \tag{30}$$

and

$$e_\delta^U : y'(e) = \theta_U \tag{31}$$
$$e_\delta^P : y'(e) = \tfrac{E_\delta(\theta)-(1-E(p))E_\phi(\theta)-p_P(1-\lambda)\theta_U}{\lambda p_U} \tag{32}$$
$$e_\phi^U : y'(e) = \theta_U \tag{33}$$
$$e_\phi^P : y'(e) = \tfrac{(1-E(p))E_\phi(\theta)-(1-\lambda)(1-p_P)\theta_U}{(1-p_U)\lambda}. \tag{34}$$

This results in:

$$U_\delta^U = E_\delta(\theta)e_\delta^P - \theta_U e_\delta^P < 0 \tag{35}$$
$$U_\delta^P = E_\delta(\theta)e_\delta^P - \theta_P e_\delta^P > 0 \tag{36}$$
$$U_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P - (\theta_U - E_\phi(\theta))e_\phi^P \leq 0 \tag{37}$$
$$U_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P - (\theta_U - E_\phi(\theta))e_\phi^P > 0. \tag{38}$$

If conditions (12) and (13) holds together, the optimal contracts designed for area C are obtained by solving (2) with $(IC_{P,\phi})$ and $(IC_{U,\delta})$ binding. This results in:

$$w_\delta^U = E_\delta(\theta)e_\delta^P - \theta_U(e_\delta^P - e_\delta^U) \tag{39}$$
$$w_\delta^P = E_\delta(\theta)e_\delta^P \tag{40}$$
$$w_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P + E_\phi(\theta)e_\phi^U \tag{41}$$
$$w_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P + E_\phi(\theta)e_\phi^U + \theta_P(e_\phi^P - e_\phi^U). \tag{42}$$

and

$$e_\delta^U : y'(e) = \theta_U \tag{43}$$
$$e_\delta^P : y'(e) = \frac{E_\delta(\theta)-(1-E(p))E_\phi(\theta)-p_P(1-\lambda)\theta_U}{\lambda p_U} \tag{44}$$
$$e_\phi^U : y'(e) = \frac{(1-E(p))E_\phi(\theta)-\lambda(1-p_U)\theta_P}{(1-p_P)(1-\lambda)} \tag{45}$$
$$e_\phi^P : y'(e) = \theta_P. \tag{46}$$

This results in:

$$U_\delta^U = E_\delta(\theta)e_\delta^P - \theta_U e_\delta^P < 0 \tag{47}$$
$$U_\delta^P = E_\delta(\theta)e_\delta^P - \theta_P e_\delta^P > 0 \tag{48}$$
$$U_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P - (\theta_U - E_\phi(\theta))e_\phi^U \leq 0 \tag{49}$$
$$U_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P - (\theta_U - E_\phi(\theta))e_\phi^U > 0. \tag{50}$$

Finally, if condition (12) holds while (13) fails, the optimal contracts designed for area D are obtained by solving (2) with $(IC_{P,\phi})$ and $(IC_{P,\delta})$ binding. This results in:

$$w_\delta^U = E_\delta(\theta)e_\delta^U \tag{51}$$
$$w_\delta^P = E_\delta(\theta)e_\delta^U + \theta_P(e_\delta^P - e_\delta^U) \tag{52}$$
$$w_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P + E_\phi(\theta)e_\phi^U \tag{53}$$
$$w_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^U + E_\phi(\theta)e_\phi^U + \theta_P(e_\phi^P - e_\phi^U). \tag{54}$$

and

$$e_\delta^U : y'(e) = \frac{E_\delta(\theta)-(1-E(p))E_\phi(\theta)-p_U\lambda\theta_P}{(1-\lambda)p_P} \tag{55}$$
$$e_\delta^P : y'(e) = \theta_P \tag{56}$$
$$e_\phi^U : y'(e) = \frac{(1-E(p))E_\phi(\theta)-\lambda(1-p_U)\theta_P}{(1-\lambda)(1-p_P)} \tag{57}$$
$$e_\phi^P : y'(e) = \theta_P. \tag{58}$$

This results in:

$$U_\delta^U = E_\delta(\theta)e_\delta^U - \theta_U e_\delta^U < 0 \tag{59}$$

$$U_\delta^P = E_\delta(\theta)e_\delta^U - \theta_P e_\delta^U > 0 \tag{60}$$

$$U_\phi^U = (E_\delta(\theta) - E_\phi(\theta))e_\delta^P - (\theta_U - E_\phi(\theta))e_\phi^U \leq 0 \tag{61}$$

$$U_\phi^P = (E_\delta(\theta) - E_\phi(\theta))e_\delta^U - (\theta_P - E_\phi(\theta))e_\phi^U > 0. \tag{62}$$

Appendix B. Proof of Lemma 1

(1) First of all, $U_i^P \geq \max\{0, U_i^U\}$. To see this, notice that from $IC_{P,i}$: $U_i^P \geq w_i^U - \theta_P e_i^U \geq U_i^U$ for all e, with strict inequality for all $e > 0$. Moreover, suppose $U_i^U < 0$, $U_i^P > 0$ is implied by the (IR_J). Using the above, the (IR_δ), the (IC_ϕ) and the fact that $\phi > \delta$, I can write the following sequence of inequalities:

$$\phi U_\phi^P + (1-\phi)U_\phi^U \geq \phi U_\delta^P + (1-\phi)U_\delta^U \geq \delta U_\delta^P + (1-\delta)U_\delta^U \geq 0,$$

which proves that (IR_ϕ) holds.

Suppose now that (IR_δ) was not binding. Then, the principal can decrease all wages in the contract by $\epsilon > 0$ without affecting any of the other constraints but raising profits.

(2) Given the above, $U_\delta^P \geq 0 \geq U_\delta^U$. Rearrange the ICs in the following way:

$$\delta(U_\delta^P - U_\phi^P) + (1-\delta)(U_\delta^U - U_\phi^U) \geq 0 \tag{IC_δ}$$

$$\phi(U_\delta^P - U_\phi^P) + (1-\phi)(U_\delta^U - U_\phi^U) \leq 0. \tag{IC_ϕ}$$

This shows that the sign of the convex combination between $(U_\delta^P - U_\phi^P)$ and $(U_\delta^U - U_\phi^U)$ changes from non-negative to non-positive when the combination comes closer to $(U_\delta^P - U_\phi^P)$ instead of $(U_\delta^U - U_\phi^U)$. This implies that $(U_\delta^P - U_\phi^P) \leq 0$ and that $(U_\delta^U - U_\phi^U) \geq 0$, which implies $U_\phi^P \geq U_\delta^P \geq 0 \geq U_\delta^U \geq U_\phi^U$.

Suppose now (IC_ϕ) was not binding. Then the principal could decrease both e_ϕ^U and e_ϕ^P keeping period 2 incentive compatibility unchanged. In this way, profits would rise, (IC_δ) would be relaxed and (IR_ϕ) would still hold by the lemma above. To see that (IC_δ) is slack, rearrange the ICs in the following way:

$$\delta(U_\delta^P - U_\delta^U) + U_\delta^U \geq \delta(U_\phi^P - U_\phi^U) + U_\phi^U \tag{IC_δ}$$

$$\phi(U_\phi^P - U_\phi^U) + U_\phi^U = \phi(U_\delta^P - U_\delta^U) + U_\delta^U \tag{IC_ϕ}$$

From (IC_ϕ), $U_\phi^U = \phi(U_\delta^P - U_\delta^U) + U_\delta^U - \phi(U_\phi^P - U_\phi^U)$. Substitute it back into the (IC_δ) to get: $(U_\phi^P - U_\phi^U) \geq (U_\delta^P - U_\delta^U)$, which always holds given Lemma 2.

Appendix C. Proof of Result 1

Consider the principal objective function as in (4). Notice that the effect of e_δ^P is given by $\lambda E_\phi(\theta) + (1-\lambda)\theta_U - E_\delta(\theta)$, which is positive if and only if condition (5) holds. Hence, if that is the case, $(w_\delta^P, e_\delta^P) = (E_\delta(\theta), 1)$.

If, instead, (5), then the employer wants to set e_δ^P as low as possible. However, ex-post incentive compatibility implies that $e_\delta^P \geq e_\delta^U$. Hence, $(w_\delta^P, e_\delta^P) = (E_\delta(\theta) e_\delta^U, e_\delta^U)$ and the contract for δ induces (imaginary) pooling.

Appendix D. Proof of Result 2

To prove the statement, simply work out the wage levels and notice that: if (5) holds:

$$\begin{aligned} w_\delta^U - \theta_U e_\delta^U &= (E_\delta(\theta) - \theta_U) < 0 \\ w_\phi^P - \theta_P e_\phi^P &= (E_\delta(\theta) - E_\phi(\theta)) > 0. \end{aligned}$$

If it does not hold:

$$\begin{aligned} w_\delta^U - \theta_U e_\delta^U &= (E_\delta(\theta) - \theta_U) e_\delta^U \in [(E_\delta(\theta) - \theta_U), 0] \\ w_\phi^P - \theta_P e_\phi^P &= (E_\delta(\theta) - E_\phi(\theta)) e_\delta^U \in [0, (E_\delta(\theta) - E_\phi(\theta))]. \end{aligned}$$

Appendix E. Proof of Result 3

To prove the statement simply work out the wage levels and notice that:

$$\begin{aligned} w_\phi^U - \theta_U e_\phi^U &= (E_\phi(\theta) - \theta_U) < 0 \\ w_\delta^P - \theta_P e_\delta^P &= \theta_P e_\delta^P - \theta_P e_\delta^P = 0. \end{aligned}$$

Appendix F. Proof of Result 4

The employer wants to design incentive compatible contracts that maximise profits. From Lemma 1 I know that (IR_δ) and (IC_ϕ) have to bind in period 1. The first is irrelevant for the optimistic workers' contract.

I can represent incentive compatibility in a (w_ϕ^U, w_ϕ^P) space as in Figure 1 in the paper. Incentive compatible contracts lie above the (IC_ϕ) between $(IC_{U,\phi})$ and $(IC_{P,\phi})$. The expected utility increases towards the top right and profits towards the bottom left. Hence, an optimal contract always lies on the (IC_ϕ) binding line. In order to select the optimal contract, I study the slope of the isoprofits, in a (w_ϕ^U, w_ϕ^P) space, and compare it to that of the (IC_ϕ). The former is given by $-\frac{(1-\lambda)(1-p_U)}{\lambda(1-p_P)}$ while the latter is $-\frac{1-\phi}{\phi}$. Hence, isoprofits are flatter than the (IC_ϕ) if $(1-p_P) \geq (1-p_U)\frac{\phi}{1-\phi}\frac{1-\lambda}{\lambda}$ which can be rearranged to obtain (12). If the latter holds, the right-hand graph in Figure 1 shows that the optimal contract has $(IC_{P,\phi})$ binding and induces efficient effort in optimistic productive workers.

Appendix G. Proof of Result 5

The proof follows that for Result 4, but using the (IR_δ) instead of the (IC_ϕ) constraint. Notice that here I use a partial equilibrium argument. That is, I assume that, given the optimal contract for the pessimistic worker, the contract designed for the optimistic worker adjusts in equilibrium in order for the (IC_ϕ) to bind.

Appendix H. Proof of Corollary 2

To prove the corollary, simply notice that in area A both (12) and (13) must fail and that $p_P \geq p_U$. From the proofs of Result 4 and Result 5 the conditions are respectively equivalent to:

$$\frac{1-\phi}{\phi} \leq \frac{(1-\lambda)}{\lambda}\frac{(1-p_U)}{(1-p_P)} \qquad \text{and} \tag{63}$$

$$\frac{(1-\delta)\lambda}{(1-\lambda)\delta} \geq \frac{p_U}{p_P}. \tag{64}$$

Start by noticing that $\frac{1-\lambda}{\lambda} > \frac{1-\phi}{\phi}$. When $p_P \geq p_U$, $\frac{1-p_U}{1-p_P} \geq 1$. Hence, (63) always fails. For (64), notice that $\frac{(1-\delta)\lambda}{(1-\lambda)\delta} > 1$. Hence the condition always fails for $\frac{p_U}{p_P} \leq 1$. This proves that area A always takes up the entire space above the 45 degree line in Figure 3.

Appendix I. Proof of Result 6

Checking for $(e_i^P - e_i^U) > 0$, it is easy to see that this is true for all contracts and all types when the $(IC_{P,\delta})$ binds. As for the rest of the contracts, it is also easy to check that the offers for productive types are always incentive compatible when they induce separation. As for those for the pessimistic type:

$$e_\delta^P - e_\delta^U \geq 0 \text{ if and only if } \phi < \frac{\delta}{1-E(p)} \tag{65}$$

which generates (14).