The Cost of Incentives under Disagreement

Can an Employee be too Motivated?

This paper is currently begin revised

Eric Van den Steen*

March 22, 2007

Abstract

This paper identifies a new cost of pay-for-performance incentives when principal and agent may disagree on the optimal course of action. In particular, pay-for-performance gives the agent a reason to disobey the principal, and thus act against his principal's interests, when the two of them disagree. In other words, high-powered incentives may decrease the agent's 'zone of acceptance' (Simon 1947) when principal and agent disagree. As a consequence, disagreement forces a trade-off between motivation and authority.

This effect has a number of implications. First, and most importantly, agents who are subject to authority will have low-powered incentive pay. Second, intrinsically motivated agents with strong views will be more likely to disobey and thus, in equilibrium, less likely to be subject to authority and more likely to be independent entrepreneurs. A surprising result is that an increase in intrinsic motivation may actually decrease all players' expected utility. Finally, subjective performance pay will be optimal when (and only when) the principal tries to exert interpersonal authority, and not just second-best when the outcome is difficult to measure or contract. I also discuss some potential implications for the theory of the firm.

Through this analysis, the paper identifies an important difference between differing priors and private benefits (or private information): with differing priors, pay-for-performance may create agency problems rather than solving them.

JEL Codes: D81, J3, L22, M12, M52
Keywords: authority, interpersonal authority, differing priors, heterogeneous priors, agency theory, low-powered incentives, intrinsic motivation, subjective evaluation

1 Introduction

Motivation is supposed to be a good thing: motivated people work hard and exert themselves to make good decisions. This paper shows, however, that motivation also has its costs. In particular,

*MIT-Sloan School of Management (evds@mit.edu). This paper circulated earlier under the title 'Too Motivated?'. I thank Bob Gibbons for his extensive help and support, Bengt Holmstrom, John Roberts, Wouter Dessein, Paul Oyer, Roberto Rigobon, and Ravi Singh for the detailed feedback and conversations, and Joe Chen, Jerker Denrell, Mathias Dewatripont, Oliver Hart, Bentley MacLeod, Jean Tirole, and the seminar participants at the CRES Conference, Gerzensee, Harvard-MIT, INSEAD, Toulouse, the University of Chicago, ULB, and USC for the many useful suggestions.

when people openly disagree on the right course of action, intrinsic and extrinsic motivation may cause employees to disobey and, in doing so, may create, rather than solve, agency problems. Stated differently, the paper shows that disagreement induces a trade-off between motivation and interpersonal authority.

The idea is as follows. Consider a principal-agent setting in which principal and agent openly disagree on what course of action will lead to a success, i.e., they have differing priors. If the agent is paid a fixed wage and has no intrinsic motivation, then he is willing to follow the principal's orders with minimal inducement, even when he believes that these orders are the wrong thing to do. If, on the contrary, the agent has high intrinsic or extrinsic motivation to achieve a success, and if he disagrees with the principal's order, then he will be very tempted to disobey his principal. Such disobedience is costly to the principal since the agent acts against the principal's interest, which reduces the principal's expected payoff. Motivation thus creates an agency problem. Stated informally, paying employees a fixed wage makes them more willing to obey, since they don't bear the consequences. Low-powered incentives thus increase an employee's 'zone of indifference' (Barnard 1938) or 'zone of acceptance' (Simon 1947).

To study this mechanism, I consider a principal-agent model in which the agent, as part of his duties, has to decide whether to undertake a certain action. The principal and the agent openly disagree whether the action should be undertaken. While the principal can try to tell the agent what to do, the agent is a free person and can thus disobey the principal's orders. The principal, however, can invest in enforcement to make sure that such disobedience is costly to the agent. Note that disobedience itself is a costly agency problem from the principal's perspective: the principal believes that the agent takes the wrong action, which reduces her own expected payoff.

The paper then shows that the agent is more likely to disobey when he has higher-powered incentive pay or strong intrinsic motivation (modelled as a private benefit from success). This induces a trade-off between authority and pay-for-performance on two levels. First, when the principal invests in enforcement, high-powered incentive pay increases the likelihood that the agent disobeys, and thus weakens the principal's influence over the agent. This lower-level trade-off causes, on its turn, a higher-level trade-off. It leads, in particular, to the emergence of two equilibrium regimes: one in which the principal does not invest in enforcement, the agent has high-powered incentives, and the agent disregards what the principal wants him to do, and another in which the principal does invest in enforcement, the agent has low-powered incentives, and now the agent obeys the principal (most of the time). Using 'authority' in the sense of 'the power or right to give orders and enforce obedience' (Concise Oxford English Dictionary), I will then say that in the second equilibrium regime, the principal has (interpersonal) authority over the agent.[1]

The model then makes the following predictions.

1. People with high-powered incentives or with strong intrinsic motivation are less likely to obey orders.

2. Employees who are subject to authority will typically have low-powered incentives.

[1]Merriam-Webster Online defines authority as 'power to influence or command thought, opinion, or behavior', which is also consistent with its use in this paper. In particular, I will say that the principal has (interpersonal) authority when she can tell the agent what to do and the agent obeys with positive probability (but would have done something different if it weren't for the principal's order). While this is consistent with the use of 'authority' by, for example, Simon (1951), other parts of the economic literature, such as Aghion and Tirole (1997), have used authority more in the sense of a 'right to make an (impersonal) decision', such as setting a salary or allocating budget to a project. Such decisions, however, often need to get implemented by other people, in which case there is an implicit assumption that the principal also has interpersonal authority over these people. Note that the process of giving orders is eliminated from the main model, but is made explicit in appendix B.

3. People with high intrinsic motivation are less likely to be subject to authority

4. An increase in the agent's intrinsic motivation may lead to a Pareto inferior outcome.[2]

5. Subjective bonuses may be optimal, even when true performance is perfectly measurable and contractible. More importantly, subjective bonuses will go together with authority.

The model thus explains or predicts the following informal observations: employees typically have low-powered incentives, people who feel responsible for an outcome are difficult to manage or control, subjective bonuses are used much more within firms than between firms, and some firms avoid people with strong views. An interesting consequence is also that firms with intrinsically motivated people will need to rely on mechanisms other than authority to coordinate their employees, such as 'hiring for fit' or 'socialization'. I also discuss the paper's potential implications for the theory of the firm.

An important observation is that differing priors are a necessary ingredient to obtain the results in this analysis. Even stronger: in the model that I study, differing priors and private benefits lead to *opposite* conclusions. While giving an agent residual income reduces agency problems caused by private benefits, it exacerbates the agency problems caused by differing priors.

While the paper focuses on the disobedience mechanism described above, the model also gives rise to a different mechanism that generates some similar predictions (although, importantly, not the predictions on disobedience). In particular, Van den Steen (2006a) shows that differing priors make it socially efficient (by revealed preference) to allocate residual income to the person with most control and with the strongest convictions, and, in the other direction, to allocate control to the person with most at stake.[3] As I will show later, this effect is orthogonal to the 'disobedience effect' that is the focus of this paper, but cannot be eliminated from the model since it is intrinsic in the differing priors setup.

Some of the paper's implications, in particular the prevalence of low-powered incentive pay, have been derived in other contexts. The most important theory in this respect is that of multi-tasking. Holmstrom and Milgrom (1991), in their seminal contribution, argue that pay for performance on one activity can reduce effort on other activities that compete for attention, and as a consequence may bias effort towards more measurable activities. While, to my knowledge, the implications of multi-tasking for a principal's ability to exert authority have not been studied, it seems that the idea in this paper that incentives may lead to disobedience might be extended to a multi-tasking context. However, the fact that in Prendergast (2002) pay-for-performance becomes more attractive relative to monitoring when the complexity of a task increases, and some other considerations suggest that this may well depend on the particular structure of the setting. Moveover, the empirical predictions with regard to the role of disagreement and intrinsic motivation would be very different. Finally, the current theory predicts that these trade-offs will exist even when there are no problems with the relative measurability of different tasks.

There are also two arguments for low-powered incentives in firms that are based on traditional agency theory. First, if authority and pay-for-performance are substitute mechanisms to get the

[2]Efficiency comparisons in this paper are based on subjective expected utility of the players. It is this measure of utility that determines what contract the players negotiate, and thus what we will observe. However, it seems that many of the efficiency effects would continue to hold if we used reference beliefs to measure utility and considered a more extended model with coordination issues. For a discussion of different ways do measure utility in models with differing priors, see Van den Steen (2005a).

[3]Van den Steen (2006a) and the current paper were originally combined in one paper entitled 'Interpersonal Authority: A Differing Priors Perspective'.

agent to do the right thing, and each carries a fixed cost, then there will be a tendency to use only one of the two. While this idea can be found informally in the literature on franchising, such as Brickley and Dark (1987) or Martin (1988), Prendergast (2002) studies it formally as part of his analysis of the tenuous trade-off between risk and incentives. Second, Baker (1992) shows that incentives will be weaker when the objective measures of performance deteriorate. Neither of these theories makes predictions on disobedience (i.e., how incentives affect the effectiveness of authority), which is the focus of this paper, or implies that intrinsic motivation to perform well (objectively) can be bad. Note, for example, that in both papers the principal would never object to incentives that are paid for by some third party, since such incentives can only help him. In the current paper, on the contrary, the principal would sometimes be willing to pay to prevent such 'free' incentives, because incentives are not just ineffective but actively damaging. This difference illustrates the fundamental role that interpersonal authority plays in the current mechanism, an aspect that is absent from these earlier papers.

The paper is thus also related to the literature on authority as a way to control agency problems. The most important result is the idea of efficiency wages (Shapiro and Stiglitz 1984), which was studied in more detail in MacLeod and Malcomsom (1989) and MacLeod and Malcomsom (1998).[4] These papers relate to the model with exit in appendix C, although there are substantially differences both in focus and in results. In particular, one key outcome of appendix C is that the required efficiency wage increases as the agent has more high-powered incentives.

Within the behavioral finance literature, Barberis and Thaler (2003) note that 'since [overoptimistic managers] think that they are already doing the right thing, stock options or debt are unlikely to change their behavior.' Their argument is thus that equity-based pay loses its ability to solve agency problems. The current paper, on the contrary, implies that stock options *will* change an overoptimistic manager's behavior, but in the wrong direction. The manager may, for example, more forcefully resist limits imposed by the board. In other words, in this paper, equity-based pay does not simply lose its ability to solve agency problems, it creates new ones. There are also more distantly related contributions, such as Manove and Padilla (1999), who show that the signaling function of collateral breaks down when there may be overoptimistic entrepreneurs or managers.

The key contribution of this paper is to show that pay-for-performance and intrinsic motivation, which usually alleviate agency concerns, may instead create an agency conflict when people disagree on the optimal course of action. In particular, such outcome-based incentives may cause the agent to act against the principal's interests, by making the agent disobey the principal's orders. As a consequence, disagreement induces a trade-off between pay-for-performance and authority. The theory provides a novel explanation for the prevalence of low-powered incentives and predicts that such low-powered incentives and subjective performance pay will covary with the principal trying to exert interpersonal authority. It also makes new predictions regarding, for example, the effects of intrinsic motivation.

The next section explores the basic trade-off. It starts with a slightly simplified model to expose clearly the intuition and then completes the model to derive the full predictions. It also shows that intrinsic motivation may reduce Pareto-efficiency. Section 3 shows that subjective bonuses are optimal if (and only if) the principal tries to exert interpersonal authority. Section 4 shows that, in this model, differing priors give very different, even opposite, results from private benefits or private information. Section 5 considers the implications for governance and the theory of the firm, while section 6 concludes. The appendices contain some proofs and study useful extensions of, and

[4] Legros and Newman (2002) also show how the ability of players to jam the signals of their opponents to a judge in a legal dispute may lead to authority as the optimal solution.

variations on, this model.

2 The Basic Trade-offs

This section studies the mechanisms at the core of this paper: how disagreement on the optimal course of action may force a trade-off between interpersonal authority and outcome-based pay-for-performance incentives. To make the analysis maximally transparent, I will first analyze a subgame of a simplified model in subsections 2.1 and 2.2. Doing so makes one of the key forces in the paper very transparent. Then, in subsections 2.3 and 2.4, I complete the model and derive the overall predictions. Subsection 2.5, finally, shows that an increase in the agent's intrinsic motivation may lower all players' utilities.

The model that I will study tries to capture the very common situation in which a boss tells her employee what to do, but the employee may disobey the order. (Such disobedience can be very open, but can also take the form of feigned misunderstanding or forgetfulness.) While the subordinate can disobey for a variety of reasons, I'm interested here in the context where the employee disobeys because he cares about the outcome and disagrees with his boss about the right course of action. A rational employee then only obeys if he fears negative consequences from disobedience. Such negative consequences can consist, for example, of getting fired or being forced to correct or repeat the work. To keep the analysis as general and as simple as possible, I will take here the agent's cost of disobedience as exogenously given. To show that essentially the same result holds when these costs are more endogenous, appendix C considers the case where the principal can, at any point in time, discontinue the project and thus 'fire' the agent, while the working paper version of this paper (Van den Steen 2005b) considers the case where the principal can, at any point in time and with some exogenously given probability, force the agent to take a specific course of action by monitoring the agent closely.[5] The results are essentially the same as the ones derived here. Note also that these negative consequences (of disobedience for the agent) often impose a cost on the principal, such as finding a replacement employee or looking over the agent's shoulder when he repeats the work. I will therefore allow disobedience to impose a cost on both agent and principal.

2.1 The Simplified Model

Consider a setting in which a principal P hires an agent A for a project. As part of the project, A has to choose whether or not to undertake some specific action. In other words, A has to choose from the set $\{Y, N\}$ where Y denotes undertaking the action ('Yes') and N denotes not undertaking the action ('No'). The agent's decision can be either right or wrong, resulting in a project revenue of respectively 1 or 0. The decision is right if and only if it fits the state of the world, which is either y or n. That state of the world is unknown, however, and each player i has his or her own subjective belief μ_i that the state is y. The players have differing priors, i.e., μ_A and μ_P may differ even though no player has private information.[6] Let, finally, ν_i denote the strength of belief of

[5] The literal assumption in Van den Steen (2005b) is actually that the principal can change (i.e., force) the agent's decision at some cost, after the agent has chosen a course of action. It can be checked that that is indeed the equilibrium outcome of an extended model where the principal can force the agent's action (at a cost to both the principal and the agent) at any point in time. In particular, the principal will prefer to wait and see whether the agent takes the 'right' course of action without being forced to do so.

[6] Differing priors do not contradict the economic paradigm: while rational agents should use Bayes' rule to update their prior with new information, nothing is said about those priors themselves, which are primitives of the model. In

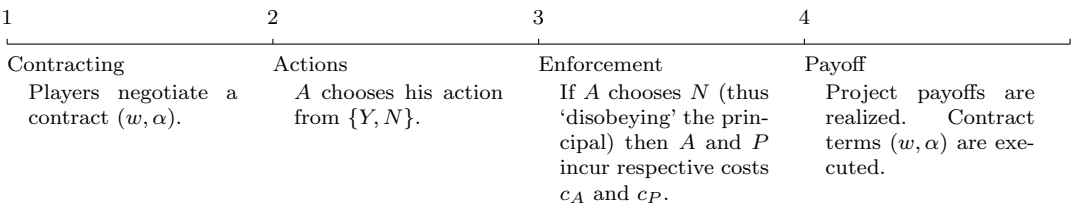

Figure 1: Time line of simplified model

player i, $\nu_i = \max(\mu_i, 1 - \mu_i)$, so that ν_i is also each player's belief in the state that he or she considers most likely.

The focus of the analysis will be on whether or when A will do what P wants him to do, i.e., whether or when A will choose whatever action P thinks is best, rather than what he himself thinks is best. I will interpret this as A 'obeying' P. For simplicity, however, the principal will not literally communicate an order to the agent. Instead, the players are simply assumed to know each others' beliefs, i.e., their beliefs are common knowledge. I will also assume, again for simplicity, that the players always disagree on the optimal course of action. In particular, let $1 > \mu_P > .5 > \mu_A > 0$ so that the principal believes that state y is most likely, while the agent believes that n is most likely. This assumption and interpretation do not affect the results. In particular, appendix B studies an extended model, in which the players sometimes agree and sometimes disagree on the optimal course of action, beliefs are private information, and P literally tells A what to do. It shows that the results are the same. Moreover, what I call 'obedience' here corresponds indeed with A literally obeying P's orders.

The timing of the game is indicated in figure 1. In period 1, the players negotiate a compensation contract for A that consists of a wage w and a share of the project revenue $\alpha \in [\,0, 1\,]$. (P's compensation is then $-w$ and the complementary share $(1-\alpha)$ of the project revenue.) Negotiation is according to axiomatic Nash bargaining with bargaining power λ and $1 - \lambda$ for P and A, and outside options of 0 for both. I motivate the $\alpha \in [\,0, 1\,]$ condition below.

In period 2, A publicly chooses his action from the set $\{Y, N\}$. This decision is non-contractible, and the ultimate control over the decision is always in the hands of the agent and cannot be contracted or otherwise moved around. It follows that the decision will always be taken by the agent, who will choose the action that is best from his perspective given his beliefs and the contract negotiated in period 1. If A chooses the action that P believes is wrong, then A and P incur in period 3 respective 'costs of disobedience' c_A and c_P. These costs are exogenously given. For A this cost represents, for example, the risk of getting fired or having to repeat the work. The corresponding costs for c_P would be the cost of finding a replacement for A or having to monitor A closely when he repeats the work.[7]

In period 4, the state gets realized, the principal receives the project's revenue, and she pays the agent according to the contract (w, α). To study the effects of the agent's intrinsic motivation, I

particular, absent any relevant information agents have no rational basis to agree on a prior. Harsanyi (1968) observed that 'by the very nature of subjective probabilities, even if two individuals have exactly the same information and are at exactly the same high level of intelligence, they may very well assign different subjective probabilities to the very same events'. For a more extensive discussion, see Morris (1995) or Van den Steen (2005a).

[7] As mentioned earlier, appendix C and Van den Steen (2005b) consider models where these costs are derived endogenously. In both cases, it is subgame-perfect for the principal to incur the cost c_P. For example, in appendix C the players agree on an efficiency wage that is sufficiently high to make it subgame perfect for the principal to fire a disobeying agent (whenever he can), even though doing so makes him lose the project.

will also allow that A gets a private benefit $\gamma_A \geq 0$ when the project is a success (and 0 otherwise). The idea here is that intrinsic motivation can be captured as a private benefit from success. For notational simplicity, I will denote player i's total benefit as α_i, so that $\alpha_P = 1-\alpha$ and $\alpha_A = \alpha+\gamma_A$.

Consider now the condition that $\alpha \in [\,0,1\,]$. Absent private benefits (i.e., $\gamma_A = 0$), this condition is in fact a no-wager condition: absent this condition, the players would bet on the state and, in doing so, generate infinite utility. This no-wager condition would follow endogenously if players had the ability to sabotage the project, i.e., if each player had the ability to make sure that the project fails. In that case, any contract with $\alpha \notin [\,0,1\,]$ would give one of the players a strict incentive to sabotage the project. Anticipating that, the other would never accept the 'bet'. To maintain generality and simplify the analysis, I simply impose the condition as an assumption.

The presence of private benefits γ_A makes this argument slightly more complex. In this case, the condition $\alpha \in [\,0,1\,]$ also excludes the possibility of eliminating the effect of γ_A by choosing a negative α. The earlier motivation, however, goes through nearly unchanged. In particular, the condition $\alpha \in [\,0,1\,]$ now follows endogenously from the ability to sabotage the project if A's private benefit is in fact a random variable that equals $\overline{\gamma}_A = \gamma_A/q$ with probability $q < \frac{1-\nu_P}{\nu_P}$ and 0 otherwise. In particular, in that case , $\alpha < 0$ will cause A to sabotage the project with probability $1-q$, which is never optimal. Alternatively, if there were a small fraction of potential employees with $\gamma_A = 0$, then these employees would be particularly attracted by a contract with negative α, and they would all try to sabotage the project. Adverse selection would then lead to disastrous results. Again, instead of including these elements explicitly in the model, I simply impose $\alpha \in [\,0,1\,]$ as an assumption so as to maintain maximum generality and simplicity. The condition is also a very natural one, in the sense that the two players simply split up the revenue from a success.

2.2 Subgame Analysis: The Effect of Motivation on Obedience

I consider now the subgame starting in period 2 to show the effect of motivation on obedience. In particular, I take the compensation contract (α, w) as exogenous and consider under what conditions A will do what P wants him to do. This builds intuition for the later results, but also delivers one of the main insights of the paper.

Proposition 1a *A will choose Y, and thus 'obey' P, if and only if $c_A \geq \alpha_A(2\nu_A - 1)$.*

Proof: A prefers to choose Y (rather than N) iff $\alpha_A(1-\nu_A) \geq \alpha_A\nu_A - c_A$. This implies the proposition. ■

This condition is central to the further analysis. It is the incentive compatibility constraint for the agent, i.e., the condition under which he is willing to 'obey' the principal. It says that obedience obtains only if the agent's penalty from disobedience is high enough.

The key insight here is that the minimal penalty c_A to keep the agent honest *increases* in α_A, the agent's benefit from a success.[8] The reason is that, as α_A increases, the agent cares more about making the right decision, thus increasing his temptation to disobey when he is asked to do

[8] In the model of appendix C, this result takes a slightly different form. There, the principal has to pay an efficiency wage to make the agent obey (Shapiro and Stiglitz 1984). The key result then is that the efficiency wage increases in α_A. As you pay the agent a higher share of the residual income, you also need to pay a higher efficiency wage (if you want to exert authority). In the equivalent model with private benefits instead of differing priors, the efficiency wage would *de*crease in α_A. MacLeod and Malcomsom (1989) and MacLeod and Malcomsom (1998) study efficiency wage models in much detail, including the issue of commitment by the principal.

something he disagrees with, and thus reducing his 'zone of indifference' (Barnard 1938) or 'zone of acceptance' (Simon 1947). As I will show later, this is the opposite result from the equivalent model with private benefits: with private benefits, the minimal penalty to keep the agent honest *decreases* in the agent's benefit from success, α_A.

This result has some important implications. First of all, all else constant, agents with high pay-for-performance are more likely to disobey their principal and just do what they themselves consider optimal. This can manifest itself as either visible disobedience or as restraint by the principal in giving orders (since she knows she will be disobeyed). Either way, it implies a loss of control for the principal, and thus creates an agency problem since the agent's action will be suboptimal from the principal's perspective. It will lead in section 2.4 to the result that agents who are subject to authority have low-powered incentives. This disobedience and loss of control is one of the distinguishing predictions of this paper.

While there are, to my knowledge, no data on disobedience that allow me to test this relationship, the result is at least consistent with the management literature on sales compensation, which cites 'loss of control' (of the manager over her salespeople) and disregard of authority among the most important negative effects of sales commissions. Oliver and Anderson (1994), for example, show that sales people who are evaluated on outcome, which includes pay-for-performance, are 'less accepting of authority/direction'. In this study, the use of outcome-based incentives was not significantly related to indicators for multi-tasking concerns, such as the ratio sell/non-sell time, the importance of planning, or the importance of call activity.

A second implication of proposition 1a is that intrinsic motivation is not always good. In particular, intrinsic motivation can be interpreted as the non-monetary benefit γ_A from achieving success. The result implies that people with higher intrinsic motivation will be more difficult to control in case of disagreement. This is consistent with personal observations that volunteer organizations tend to be very difficult to manage, with all participants going (very energetically) in all directions. This result will also lead to the result in section 2.4 that, in equilibrium, people with high intrinsic motivation will be less subject to authority.

A third implication of this result is that people are more likely to obey when they are not held responsible or accountable for the outcome, in the sense that they do not get the blame when things go wrong or do not get the praise when things go right.

Fourth, the condition implies that the agent is more likely to obey when ν_A is low. An agent with a low ν_A cares less about what course of action he follows and thus has less reason to disobey. This suggests that firms may sometimes prefer people with less experience since such people are easier to mold.

2.3 Complete Model

I now complete the model with three elements that will help the interpretation of both the model and the results. The timing of the complete game is indicated in figure 2.

The first addition is that disobedience is costly only if the principal invests at the start of the game in monitoring and other elements that make the agent bear some consequences from disobeying. I will refer to this as investing in 'enforcement'. Without such investment in enforcement, $c_A = c_P = 0$. If, on the contrary, P invests in enforcement, at cost $K \downarrow 0$ to the project, then c_A and c_P are as described below.[9]

[9]Taking the limit $K \downarrow 0$ simplifies the statement of results and the analysis, without really affecting the results.

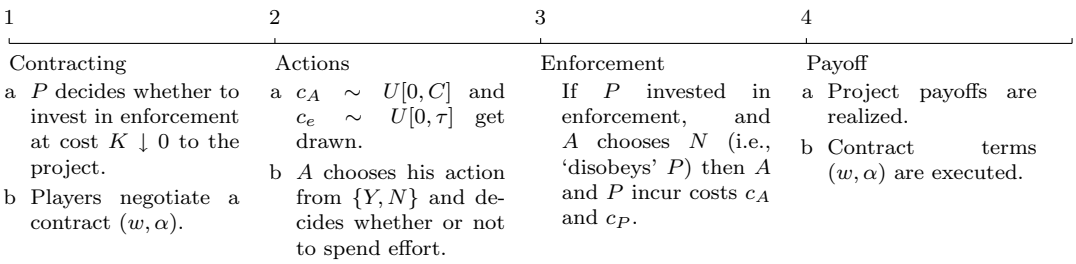

Figure 2: Time line of complete model

The second element is to introduce randomness in the agent's decision. In particular, instead of being a fixed parameter, A's cost of disobedience c_A will be a random variable with uniform distribution on $[0, C]$, with $C > 0$ if P invests in enforcement. The value of c_A will be publicly drawn at the start of period 2, i.e., after the contract negotiation but before the agent's action choice. This change to the model allows me to talk in a meaningful way about the likelihood that the agent obeys or disobeys, and how that affects equilibrium outcomes.

The third, and most important, addition to the model is to introduce an independent moral hazard component that creates a reason to give incentive pay. In particular, I will assume that project success depends not only on the agent's choice of action but also on whether or not the agent spends effort. Formally, assume that simultaneously with his choice of action (Y or N), the agent also decides whether or not to spend effort. The cost of effort to the agent is a random variable c_e with a uniform distribution on $[0, \tau]$, where $\tau \in (0, 1)$. The value of c_e will be publicly drawn, simultaneously with, but independently of, the value of c_A. With probability $(1 - \tau)$, the project is, as before, a success if and only if the agent's decision matches the state of the world. With the complementary probability τ, however, the project is a success if and only if the agent spent effort. The parameter τ thus captures the relative importance of effort versus decision making. The effect of this change is to make α (sometimes) take values other than the extremes (0 and 1), so that I can say meaningful things about how incentive pay affects behavior in equilibrium. I will assume that disobedience costs c_A and c_P are only incurred when the outcome is determined by decisions rather than by effort.[10]

2.4 Analysis

As mentioned before, the central result of this analysis is that disagreement causes a trade-off between authority and pay-for-performance incentives on two levels. First of all, the game has two types of equilibria: one type of equilibrium in which P has authority over A and A has low-powered incentives, and a second type of equilibrium in which A has high-powered incentives, but P has no authority over A. Since the equilibrium will be one or the other, this forces a high-level trade-off between authority and incentives. On a lower level, the trade-off also exists within the authority-type equilibrium, with the probability of disobedience increasing as the agent's incentives

[10]This corresponds to the assumption in the stories in appendix C and Van den Steen (2005b), that the principal observes whether the outcome will be dependent on decisions or effort at the start of period 3, i.e., just prior to deciding whether to fire the agent, respectively make him redo the work.

get stronger.[11] While these may seem, at least empirically, quite distinct results, I will argue below that the second effect is one of the causes of the first.

Before I get to the equilibrium outcome, let me first extend the result of proposition 1a to this context. In particular, the following proposition takes the subgame starting in period 2, and considers when A will do what P thinks is right, i.e., when A 'obeys' P.

Proposition 1b [Subgame Analysis] *If P invests in enforcement, then A chooses N (and disobeys) with probability* $\min\left(\frac{\alpha_A(2\nu_A-1)}{C},1\right)$. *The likelihood that A disobeys increases in α, γ_A, and ν_A.*

Proof : If P invested in enforcement, then A will choose Y iff $\alpha_A(1-\nu_A)+w \geq \alpha_A\nu_A + w - c_A$ or $c_A \geq \alpha_A(2\nu_A-1)$, which implies the proposition. ■

The following proposition then captures the dual trade-off between incentives and authority. It shows indeed that there are two types of equilibria. In the first type, P does not invest in enforcement and A always chooses N, i.e., A always disregards what the principal wants him to do and just follows his own beliefs. The principal thus has no interpersonal authority over the agent (in the sense discussed earlier). In this equilibrium, the agent will have very high-powered incentives. I will denote this type of equilibrium as *NAt*, which stands for 'No Authority'.

In the second type, P invests in enforcement and A chooses Y with strictly positive probability, i.e., he sometimes does as the principal wants him to do, going against his own beliefs. In this case, the principal thus has (some) interpersonal authority over the agent. The agent will typically have low-powered incentives. I will denote this type of equilibrium as *At*, which stands for 'Authority'. Moreover, within this type of equilibrium, stronger incentives cause more disobedience.

To state the result formally, let me define

$$\begin{aligned} f &= \tau - (1-\tau)\left[\nu_A+\nu_P-1\right] + (1-\tau)\frac{(2\nu_A-1)}{C}\left[\gamma_A(2\nu_A-1)+(\gamma_A-1)(2\nu_P-1)-c_P\right] \\ g &= \tau - (1-\tau)\frac{(2\nu_A-1)}{C}\left[(2\nu_A-1)+2(2\nu_P-1)\right] \end{aligned}$$

and let $\hat{\alpha}$ denote the equilibrium level of α.

Proposition 2 [Equilibrium] *There exists $\hat{\nu}_A$ such that the equilibrium is of type At if $\nu_A \leq \hat{\nu}_A$, and of type NAt otherwise.*

- *If the equilibrium is At and either $g \leq 0$ or $f \leq 0$ (which includes $\tau = 0$), then $\hat{\alpha} = 0$. Player A disobeys with constant probability $\frac{\gamma_A(2\nu_A-1)}{C}$.*
- *If the equilibrium is At and $f, g > 0$, then $\hat{\alpha} = f/g$. Moreover, $0 < \hat{\alpha} < 1$. Player A disobeys with probability $\frac{(\hat{\alpha}+\gamma_A)(2\nu_A-1)}{C}$, so that the level of disobedience will be high when $\hat{\alpha}$ is high.*
- *If the equilibrium is NAt, then $\hat{\alpha} = 1$. Player A always chooses N.*

The value of $\hat{\nu}_A$ decreases in γ_A and τ.

[11] Note that this is a statement about the probability of disobedience, not about the level of enforcement or monitoring. For simplicity, enforcement is a yes-no decision in this model. If, instead, P could choose the 'level' of enforcement and monitoring, then high pay-for-performance may elicit both strong monitoring/enforcement and more frequent disobedience. Strong enforcement may thus go together, in equilibrium, with weaker effective authority, in the sense of more disobedience.

Proof : The proof is in appendix A. ■

To explore the intuition behind the proposition, let me focus first on the case where P invested in enforcement. As mentioned earlier, there is a local trade-off between authority and incentives: low levels of α go together with strong authority (lower disobedience) while high levels of α go together with weaker authority (more disobedience). There are two mechanisms that cause this trade-off. The most direct mechanism is the result from proposition 1b that more incentives will cause A to disobey more. As discussed in the introduction, however, there is also a second mechanism that is explored in depth in Van den Steen (2006a). In particular, as A disobeys more, A also values the residual income more (by revealed preference) while P values the residual income less. This will, on its turn, favor a higher α. The two effects reinforce each other: as α increases, A disobeys more, thus increasing his valuation of the residual income, which favors increasing α even more. This virtuous/vicious circle makes that α and authority will tend toward extremes. But the trade-off would exist even without this second effect.

This local trade-off between authority and incentives is, on its turn, at the basis of the second, higher-level, trade-off between authority and incentives that gives rise to the two equilibrium regimes. In particular, as a higher α increases the level of disobedience, two things happen. First of all, P's benefits from investing in enforcement are lower, since A often disobeys anyways. Second, since c_A and c_P are incurred more often, the cost associated with enforcement also increases. Both these effects, which are directly caused by the increase in disobedience, make investing in enforcement less attractive. Note that there is also a third, independent, effect: as α increases, P's share from the payoffs decrease, which also reduces P's incentives to invest in enforcement. This third effect would disappear if the cost of enforcement were proportional to α_P. In that case, the dual equilibrium regime would be caused exclusively by the disobedience effect at the core of this paper.

The proposition thus predicts that, in the presence of disagreement, people who are subject to authority will usually have low-powered incentive pay, or often even fixed salaries. Since nearly all employees are subject to authority, it thus provides a new explanation for the (informally observed) lack of high-powered incentives in firms. In the other direction, the model also predicts that agents with high-powered incentives should be less subject to authority. A nice illustration of this phenomenon can be found in the HBS case on Lincoln Electric (Berg and Fast 1983), a firm famous for its high-powered incentive systems. At one point in the case, two employees are interviewed regarding their opinion about the company. Both employees start out by saying how much they like being their 'own boss' or their 'own man'.

A different take on the dual regime result is that in the first regime, the principal decides on the course of action and bears the risk of her decisions, while in the second regime the agent decides on the course of action and bears the risk of his decisions. I will come back to this interpretation and its potential relevance for the theory of the firm in section 5.

Consider next the comparative statics on the prevalence of authority. In particular, *At* becomes more prevalent (in the sense of obtaining in a strictly larger subset of the appropriate section of the $(\nu_A, \nu_P, \gamma_A, c_P, \tau)$ parameter space) as γ_A, ν_A, and τ are smaller.[12]

[12]For both γ_A and ν_A, there is again a second mechanism, beyond the one focused on in this paper, that works in the same direction. In particular, when γ_A and ν_A are higher, then A's utility from following his own views increases. This makes it more attractive to move control to A, and thus to implement the *NAt* equilibrium. In the other direction, the *NAt* equilibrium favors, by revealed preference, shifting residual income to A. As mentioned earlier, these effects are studied formally in Van den Steen (2006a). They are orthogonal to the 'obedience effect' that is the focus of this paper. This can be seen from writing the derivative of the joint utility as $\frac{\partial U}{\partial z_\alpha}\frac{dz_\alpha}{dx} + \frac{\partial U}{\partial x}$, where

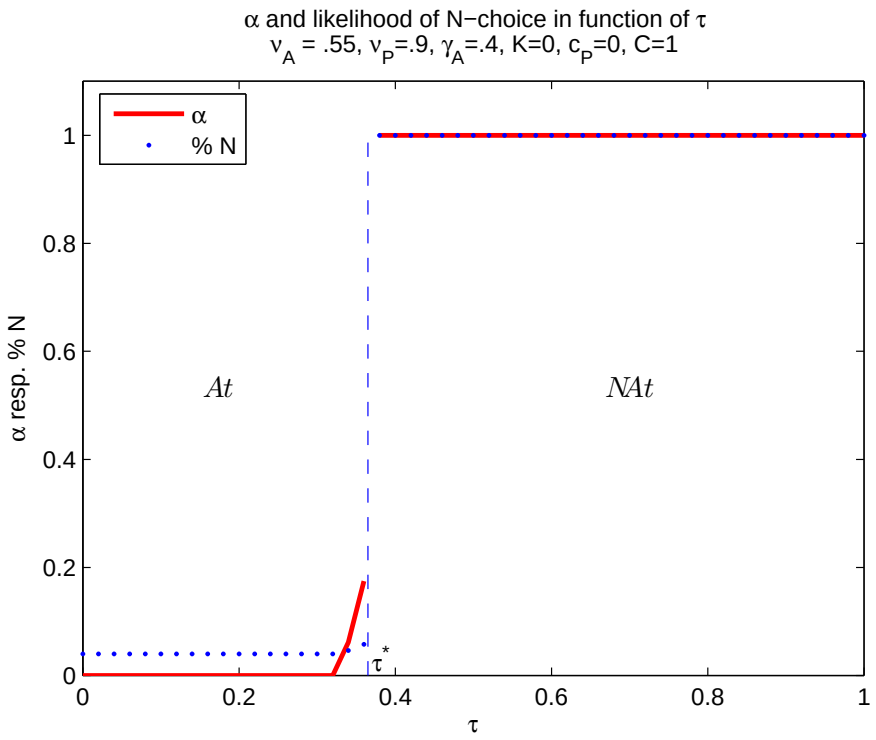

Figure 3: Evolution of $\hat{\alpha}$ and the likelihood that A chooses N ('disobeys'), as a function of τ.

Consider first the comparative static on τ, which captures the importance of personal effort by the agent. Figure 3 shows, for one specific case, the evolution of $\hat{\alpha}$ (and of disobedience) as a function of τ. At low levels of τ, the effort that would be induced by outcome-based incentives is not worth the costs from disobedience and from shifting the residual income. At some point, however, effort becomes sufficiently important for the firm to give explicit outcome-based incentives. But since the costs of providing such incentives under At are high, the equilibrium quickly shifts to the NAt equilibrium, which leads to a sudden jump in $\hat{\alpha}$. This result thus predicts that we are more likely to observe authority when the importance of making the right decisions is high relative to the importance of effort. This is discussed further in section 5.

Consider next the comparative static on ν_A, the agent's belief strength. If ν_A is higher, then A cares more about which action is taken and is thus less willing to obey P, making authority more difficult to implement. In other words, people with strong beliefs about the right course of action have difficulties following the ideas and 'orders' of others and will, in equilibrium, be less subject to authority. This predicts that people with strong beliefs should be more likely to become entrepreneurs. This is consistent with the evidence that entrepreneurs have relatively strong beliefs about being right (Cooper, Dunkelberg, and Woo 1988, Landier and Thesmar 2007). It also predicts that people with strong beliefs are more likely to be in leadership or management positions, while people with weak beliefs are more likely to be in the position of follower or subordinate.

Consider next the comparative static on γ_A. The private benefit from success, γ_A, gives A a

the first term is the disobedience effect, while the second term captures the other effect. The intuition in this section regarding disobedience is based on the fact that $\frac{\partial U}{\partial z_\alpha} < 0$.

reason to decide according to his own insights, and thus to disobey P. This makes authority both less effective (and thus less beneficial) and more expensive to implement. If we interpret γ_A as intrinsic motivation then the theory predicts that people with high intrinsic motivation are less likely to be subject to authority. Note that this result is *not* caused by the fact that motivated people need less supervision, but by the fact that it is costly to make them obey. This prediction is consistent with McClelland's (1964) theory that entrepreneurs will be people with a high need for achievement.

Another implication (of the role of γ_A) is that a firm with highly motivated employees will need to rely on other methods than authority to achieve coordination. One important alternative is to hire people with similar beliefs (Van den Steen 2006b). The prediction would thus be that firms that put a lot of weight on intrinsic motivation when hiring new employees will also hire more on 'fit' than other firms and will invest more in socialization and training.

2.5 Counter-Productive Intrinsic motivation

Propositions 1b and 2 suggest that intrinsic motivation may actually be dysfunctional in a world with differing priors, since it makes interpersonal authority less effective. I will show, in particular, that an increase in the intrinsic motivation of A may decrease both players' expected utility, despite the fact that the direct effect of an increase in intrinsic motivation is to increase A's utility. That is the content of the following proposition.

Proposition 3 *There exists $(\nu_A, \nu_P, \gamma_A, \tau)$ such that each player's utility strictly decreases in γ_A.*

Proof : Let $\tau = \gamma_A = 0$ and $\nu_P > \nu_A$, so that $\hat{\alpha} = 0$ and $U_{At} > U_{NAt}$. So we are left to show that there exist $\nu_P > \nu_A$ such that U_{At} decreases in γ_A.
For what follows remember from the proof of proposition 2 that $z_\alpha = \alpha_A(2\nu_A - 1)$, and that A disobeys (when P invested in monitoring) if and only if $c_A \leq z_\alpha$. Consider then the derivative of U_{At}

$$\begin{aligned}
\frac{dU_{At}(\alpha)}{d\gamma_A} &= \frac{\partial U_{At}(\alpha)}{\partial z_\alpha}\frac{dz_\alpha}{d\gamma_A} + \frac{\partial U_{At}(\alpha)}{\partial \gamma_A} \\
&= \left[(1-\tau)\frac{1}{C}\left(\alpha_A\nu_A + \alpha_P(1-\nu_P) - c_P - \alpha_A(2\nu_A - 1)\right) - (1-\tau)\frac{1}{C}\left(\alpha_A(1-\nu_A) + \alpha_P\nu_P\right)\right](2\nu_A - 1) \\
&\quad + \left[(1-\tau)\int_0^{z_\alpha} \nu_A \frac{1}{C}\,du + (1-\tau)\left(1 - \frac{z_\alpha}{C}\right)(1-\nu_A) + \frac{2\alpha_A + 2\alpha_P}{2}\tau\right] \\
&= (1-\tau)\frac{1}{C}\left[\alpha_A(2\nu_A - 1) - \alpha_P(2\nu_P - 1) - c_P - \alpha_A(2\nu_A - 1)\right](2\nu_A - 1) \\
&\quad + \left[(1-\tau)z_\alpha(2\nu_A - 1)\frac{1}{C} + (1-\tau)(1-\nu_A) + (1+\gamma_A)\tau\right] \qquad (1)
\end{aligned}$$

At $\tau = \gamma_A = 0$ and $\nu_P > \nu_A$ (so that $\alpha_A = 0$), this becomes

$$\frac{dU_{At}(\alpha)}{d\gamma_A} = (1-\nu_A) - \frac{2\nu_A - 1}{C}\left[(2\nu_P - 1) + c_P\right]$$

which is strictly negative for sufficiently large ν_A. This proves the proposition. ■

The intuition is as follows. Increasing γ_A has two effects. On the one hand, its direct effect is to increase the utility of A and therefore the total surplus that can be shared. This is the second term in equation 1 in the proof above. On the other hand, increasing γ_A leads to more disobedience, which is costly for the principal since she believes that the agent takes the wrong decision. This

is the first term in equation 1. The gain will be arbitrarily small when A always obeys P (which is the case when α and γ_A equal zero) and A has strong beliefs (so that he is nearly sure that the project will fail). This arbitrarily small gain is then outweighed by the cost of A disobeying more. This intuition also suggests that this utility-lowering effect of intrinsic motivation is most likely when both A and P have strong beliefs, and A's intrinsic motivation is small to begin with.

The key implication of this section on counter-productive intrinsic motivation, then, is that it is not always optimal for a firm to try to hire employees with high intrinsic motivation.

3 Subjective Performance Pay

Up to this point, I have considered pay-for-performance based only on objective measures of true performance. There are, however, some elements in the analysis that suggest that subjective performance measures (where the principal pays the agent according to how well she thinks the agent performed) may be more effective in this context. In particular, the agency problem originates here in the fact that the principal and the agent have different perspectives on the same issue. The problem might be solved if the agent were to see the problem 'through the eyes of the principal', which is exactly what subjective performance pay may accomplish. In what follows, I will show indeed that subjective performance pay dominates objective performance pay when (and only when) P tries to exert authority over A. The insight of this section is *not* so much to show that subjective performance pay may be optimal, but to show how and why it co-varies with authority.[13] In combination with the arguments of section 5, this argument provides a potential rationale why subjective performance pay is so often an explicit part of employment contracts (as opposed to business-to-business contracts, which nearly never specify completely discretionary payments).

To study this issue formally, I will assume that the contract between the principal and the agent may, apart from w and α, also contain a provision that the agent gets a bonus at the end of period 3 that equals β times the principal's expected value (at that time), $\beta E_P[R]$, with $\beta \in [0, (1-\alpha)]$ and R the revenue from the project.[14] I abstract here from the (very important) questions how the principal could commit to such a payment and whether other subjective compensation schemes may be even better. My only purpose is to show that, if the parties can find a way to implement this scheme, such subjective measures of performance completely dominate the objective ones in (and only in) the authority equilibrium. In other words, I want to show that subjective bonuses are not just a second-best solution when objective bonuses are unfeasible, but may actually be better even when objective payments are available, and that this will precisely be the case when the principal wants to exert authority.

The following proposition says indeed that, when subjective pay for performance is available, objective pay for performance will never be used in an *At* equilibrium, and *vice versa* for *NAt*. I will discuss the intuition in more detail after the proposition and its proof. Let $\hat{\alpha}$ and $\hat{\beta}$ denote the equilibrium values for α and β.

Proposition 4 *In any equilibrium where P invests in enforcement, $\hat{\alpha} = 0$. In any equilibrium where P does not invest in enforcement, $\hat{\beta} = 0$.*

[13] There is a considerable literature on subjective bonuses, such as Baker, Gibbons, and Murphy (1994) or MacLeod and Malcomsom (1998), but these papers typically assume that the objective outcome is non-contractible, as opposed to here.

[14] The restriction that $\beta \leq (1-\alpha)$ is again the no-wager condition, but now from the subjective perspective of the principal: the principal cannot promise more than what he expects to get from the project.

Proof : Consider first an equilibrium where P invests in enforcement and $(\alpha, \beta) = (a, b)$ with $a > 0$. In this case, A obeys if $c_A \geq (a + \gamma_A)(2\nu_A - 1) - b(2\nu_P - 1) = z$. This gives a total payoff as a function of a (using $\alpha = a + b$) of

$$\begin{aligned} U_{At}(a) &= (1-\tau)\int_0^z [(a+\gamma_A)\nu_A + b(1-\nu_P) + (1-a-b)(1-\nu_P) - c_P - u]\,\frac{1}{C}\,du \\ &\quad +(1-\tau)\left(1-\frac{z}{C}\right)[(a+\gamma_A)(1-\nu_A) + b\nu_P + (1-a-b)\nu_P] + \frac{(\alpha+\gamma_A)^2 + 2(\alpha+\gamma_A)(1-\alpha)}{2}\tau \end{aligned}$$

with derivative (assuming $0 < a < \alpha$ and keeping α constant)

$$\frac{dU_{At}(\alpha)}{da} = \frac{(1-\tau)}{C}\left(2(a+\gamma_A)(2\nu_A - 1) - 2(1-a)(2\nu_P - 1) - 2c_P - C\right)[\nu_A + \nu_P - 1]$$

Since the second derivative is positive, we either have $a = 0$ or $a = \alpha$. So now I want to argue that the latter can never be true in an *At* equilibrium. I will do so by showing that when $\frac{dU_{At}}{da} \geq 0$ at $a = \alpha$, then $\frac{dU_{At}}{d\alpha} > 0$ so that this can only hold at $\alpha = 1$, at which point *At* is dominated by *NAt*. To see this note that $\frac{dU_{At}}{da} \geq 0$ at $a = \alpha$ implies that $2\alpha_A(2\nu_A - 1) - 2\alpha_P(2\nu_P - 1) - 2c_P - C \geq 0$, but

$$\begin{aligned} \frac{dU_{At}(\alpha)}{d\alpha} &= \frac{(1-\tau)}{C}\left(2\alpha_A(2\nu_A - 1) - 2\alpha_P(2\nu_P - 1) - 2c_P - C\right)[\nu_A + \nu_P - 1] + \alpha_P\tau \\ &\quad +\frac{(1-\tau)}{C}[+2\alpha_P(2\nu_P - 1) + 2c_P]\left(\nu_P - \frac{1}{2}\right) \end{aligned}$$

which is then strictly positive. This proves the first part of the proposition.
For the second part, consider an equilibrium where P does not invest in enforcement. In that case, A will always choose N. The total utility from a contract with $(\alpha, \beta) = (a, b)$ with $b > 0$ is then

$$U_{NAt}(\alpha) = (1-\tau)((a+\gamma_A)\nu_A + b(1-\nu_P) + (1-a-b)(1-\nu_P)) + \frac{(\alpha+\gamma_A)^2 + 2\alpha_P\alpha_A}{2}\tau$$

which clearly increases in a (keeping α constant). This proves the second part of the proposition. ■

To see the intuition behind the result, note first that objective pay-for-performance makes disagreement costly in an *At* equilibrium: not only is one of the players always disappointed with the decision, and thus expects a low payoff, but disagreement also causes A to disobey and thus the costs c_A and c_P to be incurred. Consider now what happens if, instead, A gets paid a subjective bonus. Since he will evaluate actions and outcomes as if he had the principal's beliefs, he will do as the principal wants him to do, eliminating costly disobedience. Moreover, A will also evaluate payoffs using P's beliefs so that he will expect a high payoff from such decisions.

In a *NAt* equilibrium, on the contrary, where A always chooses N, subjective bonuses can never be optimal. On the one hand, a subjective bonus will be smaller (in the eyes of the agent) than the equivalent objective one, since the principal believes that the project is likely to fail and thus has a low expected value. On the other hand, making the agent think more like the principal also doesn't help as long as the equilibrium stays *NAt*.[15]

[15] At first sight, it may seem that there is an alternative intuition to this result. In particular, it may seem that subjective bonuses implicitly allow 'contracting on actions'. The extended model in appendix B makes clear that this is not the right intuition. While the subjective bonus result holds in that case, it is not known in advance what action the principal will believe is right, so contracting on actions could never replicate the subjective bonus.

This result thus establishes two things. First, subjective pay-for-performance may be optimal here, rather than a second-best solution when outcomes are difficult to measure. Second, and most important, authority (for the principal) will go together with subjective pay-for-performance (for the agent). This latter prediction is consistent with the observation that subjective bonuses are often used within firms, where managers exert authority over employees, but that such discretionary payments are only rarely used between firms, where authority is often much weaker or non-existent.

4 Differing Priors versus Private Benefits or Private Information

When working with differing priors, it is useful to consider whether the results are unique to differing priors or could also be obtained in a model where agency problems are modelled as private benefits or private information. There are at least two reasons for this. First, although differing priors may sometimes be a more intuitive or more relevant way to model agency problems, the argument for modelling with differing priors is obviously stronger if the results are different from what would be obtained with common priors and private benefits. Second, answering this question deepens our understanding of the underlying mechanism, and the role that differing priors play in it.

In this case, the result is striking: for the model in this paper, differing priors and private benefits give essentially *opposite* results, while private information gives no result like the one derived in this paper. This is an important outcome from a methodological point of view, since it implies that results obtained with private benefit models do not necessarily extend to the case of differing priors.

4.1 Private benefits

To compare differing priors with private benefits, I will analyze the simple model of section 2.1, but now with private benefits. The resulting model is in fact very much in the style of Prendergast (2002). As in section 2.1, assume that the agent has to choose a course of action from $\{Y, N\}$, and that the action is a success, giving payoff 1 instead of 0, if and only if it fits the state of the world, which is either y or n. However, to make this a common-prior model, now assume that the probability that the state is y is commonly known to be $\rho > .5$. The players thus agree that action Y is more likely to succeed than action N. The agent, however, has a commonly known private benefit b from undertaking N. The timing of the game remains that of figure 1, including player i incurring cost c_i when A undertakes N, i.e., when A disobeys P. Note that in this case, as in section 2.1, c_A is given.

I focus again on the contract terms under which the agent will do as his principal prefers. I discuss this condition and its implications after the proof.

Proposition 5 *There exists an equilibrium for the subgame starting in period 2 in which A does what P wants him to do if and only if $c_A \geq b - \alpha(2\rho - 1)$.*

Proof : Given that he incurs a cost c_A when he chooses N, A will choose Y if and only if $\alpha\rho + w \geq \alpha(1-\rho) + b + w - c_A$ or $c_A \geq b - \alpha(2\rho - 1)$. ■

The condition identifies the subgame equilibrium that is the analogue of the subgame equilibrium identified in section 2.2 in which the agent obeys. In particular, in both cases A chooses the action that P considers best, in order to avoid the disobedience cost c_A. The condition that $c_A \geq b - \alpha(2\rho - 1)$ is the IC constraint that makes A 'obey': it specifies the minimal penalty that

induces obedience. In section 2.2, this minimal penalty increased in α, so that pay for performance hindered authority. Here, on the contrary, the minimal penalty *decreases* in α, so that pay for performance now facilitates obedience. That is the key result of this subsection: differing priors and private benefits have *opposite* effects in this model.

Note that when A's stake α is high enough, in particular when $\alpha \geq \frac{b}{2\rho-1}$, then A obeys even without any disobedience costs. In other words, pay-for-performance aligns the objectives of principal and agent when the agency issue is private benefits, while it further misaligns their objectives when the agency issue is differing priors.

4.2 Private Information

At first sight, differing priors may also seem similar to private information that cannot be communicated: in both cases, the principal and agent have different beliefs about the right course of action. It may seem, in particular, that an agent with private information may also care more about control when he has a stake in the outcome. While it is true that incentive pay will make the agent care more about the decision and about the allocation of control, private information in itself causes no agency conflict: there is no conflict in objectives between principal and agent and there is never open disagreement (between the principal and the agent) on the optimal decision or on the optimal allocation of control. For example, the principal and the agent always agree that whoever has the best information should make the decision. It follows that private information acts indeed very different in this context than differing priors.[16]

5 Potential Implications for Governance and for a Theory of the Firm

Grocers versus Employees Alchian and Demsetz (1972) famously argued that a manager exerts no more authority over his employee than a customer exerts over his grocer. In particular, you can 'fire' your grocer, just as you can fire an employee. Some people have suggested even worse: that contractors are *more* likely to obey than employees, since firing an employee can be difficult, while you can easily walk away from your grocer.

The analysis in this paper suggests that this argument overlooks an important aspect: the fact that a typical employee has a very limited financial stake in the outcome of the project, relative to a contractor. As a consequence, employees and contractors will obey under different circumstances.

If there is disagreement on which course of action is most likely to succeed, and the principal orders a specific course of action, then a contractor will be more reluctant to obey than an employee, since the contractor bears a large share of the consequences while the employee doesn't. This is often reflected in employees' complaints in the following style: 'This is a stupid decision. But you know what? I don't care. It's his project and if he wants it this way, we'll do it that way.' If, on the other hand, the 'order' by the principal requires private effort from the agent or causes a private cost for the agent, then a contractor may well be more responsive than an employee.

In line with this argument, section 2.4 concluded that authority will be more prevalent when the importance of making the right decisions is high relative to the importance of effort. This suggests that in equilibrium we will observe employees when such obedience is important.

[16] The working paper, Van den Steen (2005b), shows this formally in a variation on this model that combines private benefits and private information.

Two Distinct Regimes The paper has a second implication that suggests that this mechanism may play a role in a theory of the firm. In particular, the analysis shows that the trade-off between authority and incentives induced by disagreement gives rise to two qualitatively different regimes:

1. A gets nearly all the residual income and P does not interfere with A's decisions.

2. P gets nearly all the residual income and tells A what to do.

Like Holmstrom and Milgrom (1994), then, this paper suggests that there are two different systems that resemble to some degree the 'firms versus markets' distinction. Moreover, in this paper the key elements of these systems are interpersonal authority and residual income, which fits well with the intuitive view that many people have of the distinction between firms and markets.

While these observations clearly do not add up to a theory of the firm (lacking even a definition), they do suggest that the combination of differing priors and authority may play an important role in such theory.

6 Conclusion

Disagreement induces a trade-off between incentives and authority since pay-for-performance will tempt the agent to disobey orders he disagrees with. As a consequence, employees subject to authority will typically have low-powered incentives, while firms with intrinsically motivated employees may have to rely on alternative means of coordination, such as hiring people with similar beliefs or preferences. Moreover, subjective bonuses will be optimal (and not just a second-best solution when outcomes are difficult to measure) exactly when the principal tries to exert interpersonal authority over the agent.

The analysis also has an important methodological implication: differing priors can have very different, even opposite, results from private benefits. We can therefore not assume that differing priors are just a special case of private benefits, and that results derived under private benefits extend to a context with differing priors. This opens up important new areas of research, since open disagreement is an important aspect of life.

A Proofs

Proof of Proposition 2: Note that in the Nash bargaining in period 1, the players will agree on the value of α that gives the highest total continuation utility, and bargain over w to allocate that total utility between them. It thus suffices to determine which α gives the highest total utility.

For backwards induction, consider first the choice of effort (which is independent of any decision on the course of action). The agent will spend effort if $\tau\alpha_A \geq c_e$, so the agent's utility from the effort portion is

$$\int_0^{\tau\alpha_A} (\tau\alpha_A - u)\,\frac{1}{\tau}\,du = \tau\alpha_A^2 - \frac{\tau\alpha_A^2}{2} = \frac{\alpha_A^2}{2}\tau$$

while the principal's utility from the effort portion is

$$\int_0^{\tau\alpha_A} (\tau\alpha_P)\,\frac{1}{\tau}\,du = \tau\alpha_A\alpha_P$$

so that the total utility from effort equals

$$\frac{\alpha_A^2 + 2\alpha_A\alpha_P}{2}\tau$$

Consider now the choice of a course of action. If P did not invest in enforcement, then the unique equilibrium is for A to always choose N (and thus disobey). This is the *NAt* equilibrium. The total utility equals

$$U_{NAt}(\alpha) = (1-\tau)(\alpha_A\nu_A + \alpha_P(1-\nu_P)) + \frac{\alpha_A^2 + 2\alpha_P\alpha_A}{2}\tau$$

with derivative

$$\begin{aligned}\frac{dU_{NAt}(\alpha)}{d\alpha} &= (1-\tau)(\nu_A + \nu_P - 1) + \frac{2\alpha_A + 2\alpha_P - 2\alpha_A}{2}\tau \\ &= (1-\tau)(\nu_A + \nu_P - 1) + (1-\alpha)\tau > 0\end{aligned}$$

so that U_{NAt} is maximized at $\alpha = 1$ with joint utility

$$U_{NAt} = (1-\tau)(1+\gamma_A)\nu_A + \frac{(1+\gamma_A)^2}{2}\tau$$

Consider next the case that P did invest in enforcement (and remember that we look at the limit $K \downarrow 0$). Proposition 1b implies that A will choose Y (and thus obey) with probability $\max\left(1 - \frac{\alpha_A(2\nu_A-1)}{C}, 0\right)$. This is the *At* equilibrium. Let now $z_\alpha = \alpha_A(2\nu_A - 1)$. If $z_\alpha \geq C$, then A will always choose N (and thus disobey) and the total utility equals

$$U_{At}(\alpha) = (1-\tau)\left[\alpha_A\nu_A + \alpha_P(1-\nu_P) - \frac{C}{2} - c_P\right] + \frac{\alpha_A^2 + 2\alpha_P\alpha_A}{2}\tau$$

The derivative for α is, in this case,

$$\begin{aligned}\frac{dU_{At}(\alpha)}{d\alpha} &= (1-\tau)\left[\nu_A - (1-\nu_P)\right] + \frac{2\alpha_A + 2\alpha_P - 2\alpha_A}{2}\tau \\ &= (1-\tau)\left(\nu_A + \nu_P - 1\right) + (1-\alpha)\tau > 0\end{aligned}$$

so that U_{At} is maximized at $\alpha = 1$ with utilities

$$U_{At} = (1-\tau)\left((1+\gamma_A)\nu_A - \frac{C}{2} - c_P\right) + \frac{(1+\gamma_A)^2}{2}\tau$$

and is thus always dominated by the case where P does not invest in enforcement. So in all equilibria where P invests in enforcement, $z_\alpha < C$.

If $z_\alpha < C$, then A will choose N (and thus disobey) with probability z_α/C. The total utility now equals,

$$\begin{aligned}
U_{At}(\alpha) &= (1-\tau)\int_0^{z_\alpha} [\alpha_A\nu_A + \alpha_P(1-\nu_P) - c_P - u]\,\frac{1}{C}\,du \\
&\quad +(1-\tau)\left(1-\frac{z_\alpha}{C}\right)[\alpha_A(1-\nu_A)+\alpha_P\nu_P] + \frac{\alpha_A^2+2\alpha_A\alpha_P}{2}\tau \\
&= (1-\tau)\frac{\alpha_A(2\nu_A-1)}{C}\left[\frac{\alpha_A(2\nu_A-1)}{2} - \alpha_P(2\nu_P-1) - c_P\right] + (1-\tau)[\alpha_A(1-\nu_A)+\alpha_P\nu_P] \\
&\quad +\frac{\alpha_A^2+2\alpha_A\alpha_P}{2}\tau
\end{aligned}$$

Note that

$$\begin{aligned}
U_{At}(\alpha=1) &= (1-\tau)\frac{(1+\gamma_A)(2\nu_A-1)}{C}\left[\frac{(1+\gamma_A)(2\nu_A-1)}{2} - c_P\right] + (1-\tau)(1+\gamma_A)(1-\nu_A) + \frac{(1+\gamma_A)^2}{2}\tau \\
&< (1-\tau)(1+\gamma_A)\frac{1}{2} + \frac{(1+\gamma_A)^2}{2}\tau < U_{NAt}
\end{aligned}$$

The derivative is

$$\begin{aligned}
\frac{dU_{At}(\alpha)}{d\alpha} &= (1-\tau)\frac{(2\nu_A-1)}{C}\left[\frac{\alpha_A(2\nu_A-1)}{2} - \alpha_P(2\nu_P-1) - c_P\right] \\
&\quad (1-\tau)\frac{(2\nu_A-1)}{C}\left[\frac{\alpha_A(2\nu_A-1)}{2} + \alpha_A(2\nu_P-1)\right] \\
&\quad -(1-\tau)[\nu_A+\nu_P-1] + \alpha_P\tau \\
&= \tau - (1-\tau)[\nu_A+\nu_P-1] + (1-\tau)\frac{(2\nu_A-1)}{C}[\gamma_A(2\nu_A-1) + (\gamma_A-1)(2\nu_P-1) - c_P] \\
&\quad -\alpha\left(\tau - (1-\tau)\frac{(2\nu_A-1)}{C}[(2\nu_A-1)+2(2\nu_P-1)]\right) \\
&= f - g\alpha
\end{aligned}$$

where

$$\begin{aligned}
f &= \tau - (1-\tau)[\nu_A+\nu_P-1] + (1-\tau)\frac{(2\nu_A-1)}{C}[\gamma_A(2\nu_A-1) + (\gamma_A-1)(2\nu_P-1) - c_P] \\
g &= \tau - (1-\tau)\frac{(2\nu_A-1)}{C}[(2\nu_A-1)+2(2\nu_P-1)]
\end{aligned}$$

The second derivative is

$$\frac{d^2U_{At}(\alpha)}{d\alpha^2} = -g = (1-\tau)\frac{(2\nu_A-1)}{C}[(2\nu_A-1)+2(2\nu_P-1)] - \tau.$$

If $g \le 0$, then the solution is either $\alpha = 0$ or $\alpha = 1$.[17] Since $U_{At}(\alpha=1) < U_{NAt}$, it follows that, using $\hat{\alpha}$ to denote the optimal α, $\hat{\alpha} = 0$ whenever the equilibrium is At.

Consider next $g > 0$, so that the unrestricted optimal α is unique and determined by the FOC, and would thus equal f/g. If $f \le 0$, then $\hat{\alpha} = 0$. When $f > 0$, then $\hat{\alpha} = \min(f/g, 1)$. But from before we know that $U_{At}(\alpha=1) < U_{NAt}$, so in equilibrium, the constraint will never bind and we have $\hat{\alpha} = f/g$.

[17] Actually, if $g = f = 0$, then all $\alpha \in [0,1]$ are solutions, but given indifference, it is sufficient to look at the extremes, and since $U_{At}(\alpha=1) < U_{NAt}$, At will always be dominated in this case.

I will now show that $U_{At}(\hat{\alpha}) - U_{NAt}$ strictly decreases in ν_A and γ_A. Moreover, at the point where $U_{At}(\hat{\alpha}) = U_{NAt}$, $U_{At}(\hat{\alpha}) - U_{NAt}$ strictly increases in ν_P and strictly decreases in τ. These four results imply the rest of the proposition. Note that the envelope theorem applies, so that it is not necessary to consider how the optimal α changes with any of these parameters.

Consider first ν_A. Note that

$$\frac{dU_{NAt}(\alpha)}{d\nu_A} = (1-\tau)(1+\gamma_A)$$

while, using $z_\alpha < C$ or $\alpha_A(2\nu_A - 1) < C$,

$$\begin{aligned}
\frac{\partial U_{At}(\alpha)}{\partial \nu_A} &= (1-\tau)(-\alpha_A) + (1-\tau)\frac{\alpha_A 2}{C}\left(\frac{1}{2}\alpha_A(2\nu_A-1) - \alpha_P(2\nu_P-1) - c_P\right) \\
&\quad + (1-\tau)\frac{\alpha_A(2\nu_A-1)}{C}\alpha_A \\
&= -(1-\tau)\alpha_A + (1-\tau)\frac{\alpha_A}{C}\left(2\alpha_A(2\nu_A-1) - 2\alpha_P(2\nu_P-1) - 2c_P\right) \\
&< -(1-\tau)\alpha_A + (1-\tau)\frac{\alpha_A}{C}2\alpha_A(2\nu_A-1) \\
&< -(1-\tau)\alpha_A + (1-\tau)2\alpha_A < (1-\tau)(1+\gamma_A)
\end{aligned}$$

so that $\frac{dU_{At}(\hat{\alpha}) - U_{NAt}}{d\nu_A} < 0$.

Consider next γ_A. In this case,

$$\frac{dU_{NAt}(\alpha)}{d\gamma_A} = (1-\tau)\nu_A + (1+\gamma_A)\tau$$

while

$$\begin{aligned}
\frac{\partial U_{At}(\alpha)}{\partial \gamma_A} &= (1-\tau)(1-\nu_A) + (1-\tau)\frac{1}{C}\left(\alpha_A(2\nu_A-1)^2 - \alpha_P(2\nu_A-1)(2\nu_P-1) - (2\nu_A-1)c_P\right) + \frac{2\alpha_A + 2\alpha_P}{2}\tau \\
&= (1-\tau)(1-\nu_A) + (1-\tau)\frac{1}{C}(2\nu_A-1)\left(\alpha_A(2\nu_A-1) - \alpha_P(2\nu_P-1) - c_P\right) + (1+\gamma_A)\tau \\
&< (1-\tau)(1-\nu_A) + (1-\tau)\frac{1}{C}(2\nu_A-1)\alpha_A(2\nu_A-1) + (1+\gamma_A)\tau \\
&< (1-\tau)(1-\nu_A) + (1-\tau)(2\nu_A-1) + (1+\gamma_A)\tau \\
&= (1-\tau)\nu_A + (1+\gamma_A)\tau
\end{aligned}$$

which implies again $\frac{dU_{At}(\hat{\alpha}) - U_{NAt}}{d\gamma_A} < 0$.

Consider next ν_P. Note that U_{NAt} is independent of ν_P. So it suffices to show that $\frac{\partial U_{At}(\hat{\alpha})}{\partial \nu_P} > 0$ whenever $U_{At}(\hat{\alpha}) = U_{NAt}$. To this end, note

$$\begin{aligned}
\frac{\partial U_{At}(\alpha)}{\partial \nu_P} &= (1-\tau)\alpha_P + (1-\tau)\frac{\alpha_A(2\nu_A-1)}{C}(-2\alpha_P) \\
&= (1-\tau)\alpha_P\left(1 - 2\frac{1}{C}\alpha_A(2\nu_A-1)\right) \\
&= (1-\tau)\alpha_P\left(1 - 2\frac{z_\alpha}{C}\right)
\end{aligned}$$

When (by contraction) this is (weakly) negative, then $2\frac{z_\alpha}{C} \geq 1$ so that (using also $z_\alpha \leq C$)

$$\begin{aligned} U_{At}(\alpha) &= (1-\tau)\left(\alpha_A(1-\nu_A)+\alpha_P\nu_P\right)+(1-\tau)\frac{z_\alpha}{C}\left(\frac{1}{2}\alpha_A(2\nu_A-1)-\alpha_P(2\nu_P-1)-c_P\right)+\frac{\alpha_A^2+2\alpha_A\alpha_P}{2}\tau \\ &< (1-\tau)\left(\alpha_A(1-\nu_A)+\alpha_P\nu_P\right)+(1-\tau)\alpha_A(\nu_A-\frac{1}{2})-(1-\tau)\frac{1}{2}\alpha_P(2\nu_P-1)+\frac{(1+\gamma_A)^2}{2}\tau \\ &= (1-\tau)\frac{1+\gamma_A}{2}+\frac{(1+\gamma_A)^2}{2}\tau \\ &< (1-\tau)(1+\gamma_A)\nu_A+\frac{(1+\gamma_A)^2}{2}\tau = U_{NAt} \end{aligned}$$

It follows that, whenever $U_{At}(\hat{\alpha}) \geq U_{NAt}$, $\frac{dU_{At}(\hat{\alpha})-U_{NAt}}{d\nu_P} > 0$.
Consider finally τ. Note that we can write $U_{At} = (1-\tau)A+\tau B$ and $U_{NAt} = (1-\tau)E+\tau F$, where $F > B$. It follows that when $U_{At} = U_{NAt}$, $E < A$. The derivative can now be written

$$\frac{\partial U_{At}(\hat{\alpha})-U_{NAt}}{\partial\tau} = -A+B+E-F < 0$$

The second part of the proposition then follows. ■

B A Model with Explicit Orders

The models in section 2 trade off realism for transparency and simplicity. One of the key concessions in this respect is the assumption that the players' beliefs are common knowledge and that they are known to disagree. In reality, people cannot read each others' minds, and beliefs are thus private information. Moreover, people sometimes agree and sometimes disagree, and many of the contentious issues arise during the execution of the project, long after the contract has been signed. The purpose of this appendix is to present an extension of the model in this sense and show that the results still hold.

Consider therefore the model of section 2.3 with the following modifications. The most important modification is that beliefs are not common knowledge, but are randomly drawn and private information to the players. The players can communicate about these beliefs through cheap talk. In particular, the beliefs get drawn at the start of period 2 from a random distribution, with μ_i denoting i's belief that the state is y. The idea of having the beliefs being drawn after the contract negotiation, is that the contentious issues arise only after the project has been started, so that it is only at that time that it becomes clear which issues and thus which beliefs are relevant. To keep the analysis transparent and tractable, I assume a very simple degenerate distribution for these prior beliefs. In particular, for some given parameters $\nu_A, \nu_P \in (.5, 1)$, μ_i will equal either ν_i or $1-\nu_i$, with equal probability. In other words, μ_i is drawn from a 2-point distribution with half its weight on ν_i and half on $1-\nu_i$. It follows that the player always has the same strength of belief, ν_i, in the state that he considers most likely. Moreover, each player will believe half the time that y is the most likely state, and half the time that n is the most likely state. The prior beliefs will be independent draws, so that the players will disagree half the time. The expected project revenue according to i is ν_i when i believes that the decision is right, and $(1-\nu_i)$ otherwise.

The beliefs are originally private information. In the second step of period 2, however, P has a chance to tell A what to do. In particular, P chooses whether and, if so, what message to send from the set $\{Y, N\}$. These messages can best be interpreted as respectively 'you should do Y' and 'you

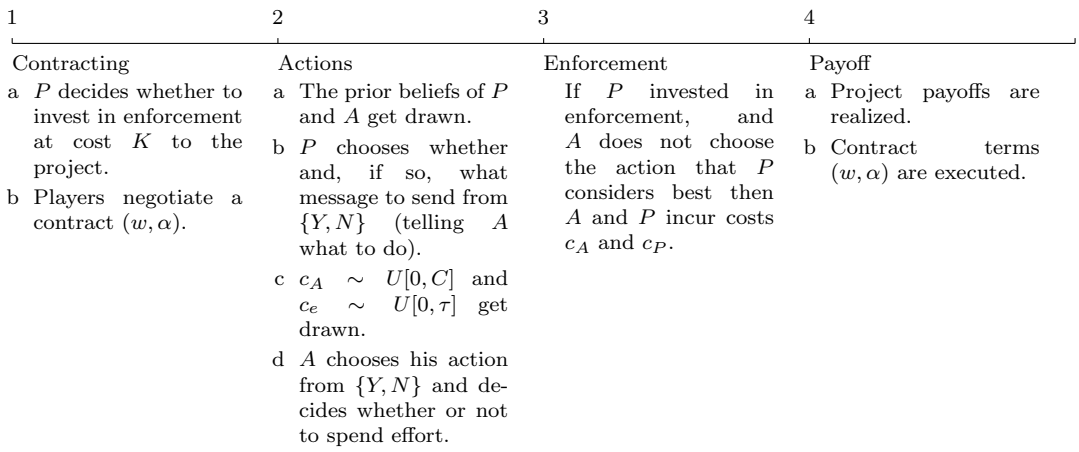

Figure 4: Time line of basic model with explicit orders

should do N'. While sending these messages is costless, I will assume that P has a lexicographic preference for being obeyed. In particular, I will assume that P, when otherwise indifferent, prefers 'giving an order and being obeyed (most of the time)' over 'not giving an order', and prefers 'not giving an order' over 'giving an order and being disobeyed (most of the time)'. For simplicity, I will also limit attention to pure-strategy equilibria that are not Pareto-dominated (in any subgame), which will imply that, in equilibrium, the communications will reveal the players' beliefs truthfully, if at all. The full timing of the game is shown in figure 4.

I have to be careful here when I use the term 'obey'. In particular, I will reserve the term 'obey' for the case in which A does what P (literally) tells him to do. Note that this is not necessarily the same as what P wants him to do (since the 'order' is cheap talk and may thus differ from P's true preferences). Whenever confusion is possible and A does what P *wants* him to do, I will simply say so.

As in section 2.3, I will assume that c_P and C are exogenously given. I will now define the *At* and *NAt* equilibria as follows. In the *NAt* equilibrium, P does not invest in enforcement, P never tells A what to do, and A always chooses the action that he considers most likely to succeed. In the *At* equilibrium, P invests in enforcement, P always tells A what to do (and the order is always what P really wants A to do), and A sometimes obeys P's order against his own beliefs. The following proposition shows that the result is identical to that in section 2.3. Let now

$$\begin{aligned} f &= \tau - \frac{(1-\tau)}{2}\left[2\nu_P - 1\right] + \frac{(1-\tau)}{2}\frac{(2\nu_A - 1)}{C}\left[\gamma_A(2\nu_A - 1) + (\gamma_A - 1)(2\nu_P - 1) - c_P\right] \\ g &= \tau - \frac{(1-\tau)}{2}\frac{(2\nu_A - 1)}{C}\left[(2\nu_A - 1) + 2(2\nu_P - 1)\right] \end{aligned}$$

Proposition 6 *There exists $\hat{\nu}_A$ such that the equilibrium is of type At if $\nu_A \leq \hat{\nu}_A$, and of type NAt otherwise.*

- *If the equilibrium is At and either $g \leq 0$ or $f \leq 0$ (which includes $\tau = 0$), then $\hat{\alpha} = 0$. Player A disobeys with constant probability $\frac{\gamma_A(2\nu_A - 1)}{2C}$.*
- *If the equilibrium is At and $f, g > 0$, then $\hat{\alpha} = f/g$. Moreover, $0 < \hat{\alpha} < 1$. Player A disobeys with probability $\frac{(\hat{\alpha}+\gamma_A)(2\nu_A-1)}{2C}$, so that the level of disobedience increases in $\hat{\alpha}$.*
- *If the equilibrium is NAt, then $\hat{\alpha} = 1$. Player A always chooses the action that he considers best.*

The value of $\hat{\nu}_A$ decreases in γ_A and τ.

Proof : The proof is completely analogous to the one of proposition 2 and is available from the author. ■

C A Model with Exit

A very natural setting is one in which the principal and the agent can exit the project at any point in time. In this appendix, I show that such model has very similar results to the model studied in section 2. A key benefit of this model is that it easily relates to well-known ideas about authority, in particular to the ideas on efficiency wages (Shapiro and Stiglitz 1984). I will show, for example, that the efficiency wage will increase in the agent's share of the project: as you pay the agent more residual income, you also need to pay him a higher fixed wage. Moreover, the outcome of this model is about the best that can be achieved if actions are non-contractible and the agent has limited liability.

The formal setting is essentially identical to that of section 2.1, except for the following changes:

1. Players do not incur exogenous disobedience costs: $c_A = c_P = 0$.
2. Prior to stage 4, either player can, at any point in time, try to exit the project. Such attempt does not always succeed.[18] Whether exit is possible does not get revealed until one player really tries to exit. If such attempt (by either player) is successful, then the game is over, the contract gets cancelled, and both players get 0. The ex-ante probability that exit is possible is exogenously given and equal to p.
3. In the first stage, I will use the extension of Nash bargaining by Zhou (1997) (in which the bargaining solution is always the point that maximizes the Nash product) when the feasible set is non-convex.

While formally each player can try to exit at any point in time before period 4, it is straightforward (given that the outside options in the bargaining are the same as the payoff upon exit) that this can only be (strictly) optimal in period 3 (after A has made his decision). I will therefore simplify the game here and only consider that option. Figure 5 then depicts the timing of this game. For simplicity, I restrict the analysis to pure-strategy, Pareto-optimal equilibria.

As mentioned earlier, the negotiated wage will now also play the role of efficiency wage: the principal pays the agent more than the market wage, in order to have something to punish the

[18] One could imagine, for example, that it requires that you get the contract declared void. Whether this will happen depends on specific wording in the contract etc.

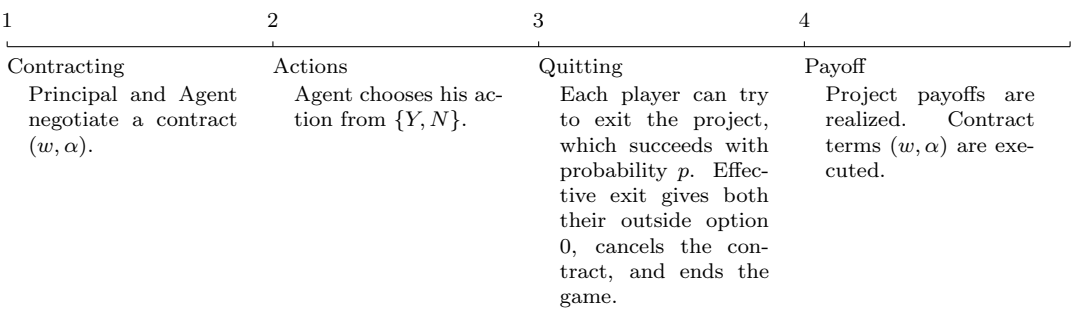

Figure 5: Time line of the model with exit

agent. In particular, the following proposition identifies the contract terms that cause the agent to obey the principal. The proof is available from the author.

Proposition 7 *There exists a (pure-strategy, Pareto-optimal) equilibrium (for the subgame that starts in period 2) in which A does what P wants him to do, and neither player tries to exit (in equilibrium) if and only if the following conditions are satisfied:*

$$w \geq \alpha_A(\theta(2\nu_A - 1) - (1 - \nu_A)) \tag{2}$$

$$w \geq \alpha_P(1 - \nu_P) \tag{3}$$

$$w \leq \alpha_P \nu_P \tag{4}$$

where $\theta = \frac{(1-p)}{p}$.

The first condition is the efficiency wage condition that makes it incentive compatible for the agent to 'obey' the principal. The second condition commits the principal to firing an agent who disobeys: it guarantees that the wage is so high that the principal only wants to continue if the agent did obey. The third condition is a simple individual rationality condition. Note that the efficiency wage in equation (2) indeed increases in α_A (for sufficiently low values of p), so that higher pay-for-performance makes it more difficult to generate authority and obedience.

There are again two equilibria. In analogy to before, I will use *At* ('Authority') to describe the following equilibrium (in this game *without moral hazard component*):

- Principal and agent agree on a contract in which $\alpha = 0$.
- The agent does what the principal thinks is best.
- The principal tries to exit if (and only if) she observes that the agent did not act as the principal wanted him to do.

Note that disobedience is possible with $\alpha = 0$ since $\gamma_A \geq 0$.

I will use *NAt* ('No Authority') to describe the following equilibrium:

- Principal and agent agree on a contract in which $\alpha = 1$.
- The agent chooses the action that he believes has the highest probability of success.
- Neither player tries to exit in equilibrium or as a response to a deviation by the other.

The following proposition then says that these two equilibria are the only possible (pure-strategy, Pareto-dominant) equilibria, and that interpersonal authority will be *more* likely when the agent has weaker beliefs and less private benefits at stake. The proof is again available from the author.

Proposition 8 *For any set of parameters, the (pure-strategy, Pareto-optimal) equilibrium exists and is either At or NAt. There exists $\hat{\nu}_P$ such that the (only) equilibrium is At when $\nu_P > \hat{\nu}_P$ and the (only) equilibrium is NAt when $\nu_P < \hat{\nu}_P$. The value of $\hat{\nu}_P$ increases as γ_A or ν_A increase.*

References

Aghion, P., AND J. Tirole (1997): "Formal and real authority in organizations," *Journal of Political Economy*, 105(1), 1– 29.

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

Baker, G., R. Gibbons, AND K. J. Murphy (1994): "Subjective Performance Measures in Optimal Incentive Contracts," *Quarterly Journal of Economics*, 109(4), 1125– 1156.

Baker, G. P. (1992): "Incentive contracts and performance measurement," *Journal of Political Economy*, 100(3), 598– 614.

Barberis, N. C., AND R. H. Thaler (2003): "A Survey of Behavioral Finance," in *Handbook of the Economics of Finance 1B*, ed. by G. M. Constantinides, M. Harris, and R. M. Stulz. Elsevier North Holland.

Barnard, C. (1938): *The Functions of the Executive*. Harvard University Press, Cambridge MA.

Berg, N. A., AND N. D. Fast (1983): "Lincoln Electric Co," HBS Case 9-376-028.

Brickley, J. A., AND F. H. Dark (1987): "The Choice of Organizational Form: The Case of Franchising," *Journal of Financial Economics*, 18, 401– 420.

Cooper, A. C., W. C. Dunkelberg, AND C. Y. Woo (1988): "Entrepreneurs' Perceived Chances for Success," *Journal of Business Venturing*, 3(3), 97– 108.

Harsanyi, J. C. (1968): "Games with Incomplete Information Played by 'Bayesian' Players, I-III, Part III. The Basic Probability Distribution of the Game," *Management Science*, 14(7), 486– 502.

Holmstrom, B., AND P. Milgrom (1991): "Multi-Task Principal Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7, 24– 52.

——— (1994): "The firm as an incentive system," *American Economic Review*, 84(4), 972– 991.

Landier, A., AND D. Thesmar (2007): "Financial Contracting with Optimistic Entrepreneurs," *Review of Financial Studies*, Working Paper NYU Stern - ENSAEForthcoming.

Legros, P., AND A. Newman (2002): "Courts, Contracts, and Interference," *European Economic Review*, 46(4), 734– 744.

MacLeod, W. B., AND J. M. Malcomsom (1989): "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57(2), 447– 480.

——— (1998): "Motivation and Markets," *American Economic Review*, 88, 388– 411.

Manove, M., AND A. J. Padilla (1999): "Banking (Conservatively) with Optimists," *Rand Journal of Economics*, 30(2), 324– 350.

Martin, R. (1988): "Franchising and Risk Management," *American Economic Review*, 78(5), 954– 968.

McClelland, D. C. (1964): *Power: The Inner Experience*. Irvington Publishers, New York.

Morris, S. (1995): "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11, 227– 253.

Oliver, R. L., AND E. Anderson (1994): "An Empirical Test of the Consequences of Behavior- and Outcome-Based Sales Control Systems," *Journal of Marketing*, 58(4), 53– 67.

Prendergast, C. (2002): "The Tenuous Trade-off between Risk and Incentives," *Journal of Political Economy*, 110(5), 1071– 1102.

Shapiro, C., AND J. E. Stiglitz (1984): "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74(3), 433– 444.

Simon, H. (1947): *Administrative Behavior*. Free Press, New York.

——— (1951): "A Formal Theory of the Employment Relationship," *Econometrica*, 19, 293– 305.

Van den Steen, E. J. (2005a): "Notes on Modelling with Differing or Heterogeneous Priors," Working Paper, MIT-Sloan.

——— (2005b): "Too Motivated?," MIT Sloan Working Paper No. 4547-05.

——— (2006a): "Disagreement and the Allocation of Control," Working Paper MIT Sloan.

——— (2006b): "On the Origin of Shared Beliefs (and Corporate Culture)," Working Paper MIT-Sloan.

Zhou, L. (1997): "The Nash Bargaining Theory with Non-Convex Problems," *Econometrica*, 65(3), 681– 685.