Implicit Incentives and Delegation in Teams

Harvey Upton*

27th August 2021

Abstract

This paper studies an infinitely-repeated game of team production, in which agents are required to supply costly effort in the presence of moral hazard. The principal also has the option to delegate an additional production-relevant decision to a member of the team. We provide conditions under which delegation changes the scope of peer sanction and thus influences the implicit incentives which are generated by the agents' repeated interaction. Delegation can then be strictly preferred by the principal to centralisation, despite misalignment of preferences and the absence of asymmetric information with respect to the efficient decision. We provide a comparative static analysis and use our results to discuss various aspects of organisational design, including the implications of our findings for optimal team composition, choice of leader and transparency within the firm.

1 Introduction

The last few decades have seen a consistent shift in how firms choose to organise the production process, away from traditional individual working practices and towards the widespread use of teams.[1] The advantages of teamwork are numerous. By combining the specialised knowledge and skills of many workers, teams are able to create complementarities in the production process which enhance performance (Lazear & Shaw 2007). A closely integrated production process also improves workers' abilities to help one another, share production-relevant information and provide colleagues with timely feedback. Importantly, team members often interact repeatedly over time and are typically responsible for monitoring and motivating one another, which can help mitigate the well-known freeriding problems which are often associated with team production (Holmström 1982). Workers in teams understand that their behaviour today can affect the actions of others tomorrow, and take this into account when choosing their productive contributions. Che & Yoo (2001) show that this so-called 'mutual monitoring', when paired with an appropriate team-level reward scheme, can create strong implicit incentives which reduce firms' costs of motivating employees.

At the same time, firms are also becoming flatter (Rajan & Wulf 2006) and, in response to rapid changes in information technology, more decentralised, so that autonomy over various aspects of the production process is increasingly granted to lower-level employees (Bresnahan et al. 2002). At the intersection of these two trends are self-organised teams (SOTs). SOTs typically bring together individuals with diverse skill and

*Humboldt-Universität zu Berlin. Contact: harvey.upton@hu-berlin.de
This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 403041268 – TRR 266 Accounting for Transparency. I would like to thank Anja Schöttner, Amir Habibi and audiences at the TRR 266 Annual Conference 2020 and the HU Berlin Microeconomic Theory Colloquium for valuable comments.

[1]For instance, Lazear & Shaw (2007) report that extensive adoption of teams in large firms grew from 37% to 61% between 1987 and 1999.

knowledge sets, who are given collective autonomy as a unit and are responsible for planning, managing and executing tasks interdependently in order to attain their goals (Magpili & Pazos 2018). Douglas & Gardner (2004) report that SOTs are utilised by almost three quarters of the top 1000 US firms, while recent evidence suggests that their use has become even more frequent over time.[2] Their discretion over the production process is wide-ranging and can extend from project management and formulation of strategy, to the development of skills and even the self-evaluation of team performance.

The economic approach to the study of organisations postulates that decisions regarding the organisation of the production process should be made holistically, so that a firm's choice of whether to allocate authority to lower-level workers should depend on whether they operate as part of a team. Yet the majority of the economic literature which studies the allocation of decision rights within a firm limits attention to the interaction between a principal and a single agent. This differs from an interaction between a principal and a team of agents in several important aspects, not least because workers in teams often make decisions which can impact their teammates in a myriad of different ways. For instance, in the context of product design, decisions over one aspect of the product (e.g. functionality) necessarily shape other aspects of the design (e.g. aesthetics) and thus constrain the choices of other employees. The fact that workers make decisions which directly affect their colleagues has important implications for the informal relationships between team members, and, by extension, for the implicit incentives which are generated by their repeated interaction. The goal of this paper is to study the allocation of decision rights to a team of workers, the ensuing impact of their increased authority on these implicit incentives, and the resulting implications for the firm's wage structure and choice of organisational design.

For that purpose, we study a dynamic model of team production in which workers must be motivated to provide effort in the presence of moral hazard. The firm can also choose whether or not to delegate an additional production-relevant decision to a member of the team. We show that delegation can be strictly profitable for the firm — even when parties have symmetric information regarding the efficient decision and when preferences over the decision between the firm and the team are misaligned — due to its positive effect on the implicit incentives which are generated by the workers' interaction. By highlighting this potential benefit of delegating decision rights, our framework is also able to generate a number of novel insights and empirical predictions regarding various aspects of organisational design, including the optimal composition of teams, the role of leadership and the implications of delegation for transparency within firms. More generally, the paper contributes to the literature which explores the combined used of innovative management practices within organisations, by outlining complementarities between the delegation of decision rights, the adoption of teamwork and the encouragement of mutual monitoring between workers (see also, for instance, the discussions in Chalioti 2016 and Ishihara 2017).

We formalise team production as an infinitely-repeated game à la Che & Yoo (2001), featuring a principal and two agents. In each period, the agents are required to provide costly effort in order to increase the probability of a binary-outcome project being successful. Since the agents' effort provision cannot be observed by the principal, a moral hazard problem arises, so that the agents must be motivated to work hard by payments which condition on project success.[3]

While many studies have considered similar environments, our framework differs in two important ways. First, there are two possible 'production methods' which can be utilised by the team in order to achieve

[2]See for instance Deloitte's 2017 report on Global Human Capital Trends.

[3]The lack of individual measures of performance implies that the agents are incentivised using a group-reward scheme. In practice, workers are often compensated according to a team-level performance measure; see Hamilton et al. (2003) and Boning et al. (2007) for examples. Nonetheless, we discuss the case of individual performance measurement in the conclusion to the paper.

their goal. Second, in each period one of two non-verifiable states of the world is realised, corresponding to variations in production-relevant information over time. Both of these factors influence the probability of project success. Specifically, in a particular state of the world, only one of the two production methods is 'correct', with implementation of the 'incorrect' production method reducing the probability that the project is successful. In addition, each agent incurs a cost associated with a particular production method being adopted, with the difference in these (heterogeneous) costs representing the agents' preferences over the production process; the principal, on the other hand, cares only about the final outcome of the project.

These assumptions aim to capture the idea that, in many work environments, there are several ways in which a particular goal can be achieved. Examples abound. Software engineers can develop a program using many different programming languages; successful advertising campaigns take various forms and can reach consumers through a wide range of media channels; scientists can approach a particular research question using a variety of different methods; businesses can attract new clients by emphasising the quality of their services, or by attempting to cultivate a strong relationship using interpersonal skills. More generally, in situations where a series of tasks need to be completed in order to achieve a particular objective, there are often many possible allocations of different tasks to different workers. Production can also vary in the extent to which workers help one another, share information and provide valuable feedback to others.

In all of the foregoing examples, the 'best' approach to the production process will typically be sensitive to the particular circumstances at hand, so that the most appropriate production method varies over time. Moreover, workers will often have preferences over which approach is adopted. These preferences may represent their fields of expertise and thus how comfortable they are with a particular production method or may reflect their subjective feelings over different approaches, such as a desire to avoid being allocated an especially unpleasant task. Preferences could even incorporate other elements, such as private benefits, or valuable training or experience, which they might expect to gain from the adoption of a particular working method.

In this framework, the core of our analysis is concerned with the following question: should authority over the production method to be implemented in each period lie with the principal (*centralisation*) or with one of the agents (*delegation*)? One possible rationale for delegating the decision would be if the agents had superior information over the state of the world in each period. However, we do not make this assumption. We instead assume that the state of the world is common knowledge, so that all parties — including the principal — can correctly identify the appropriate production method for the current period. Nonetheless, we show that delegation can still become optimal as an organisational design, due to the implications of decision-making authority for the implicit incentives generated by the repeated interaction between the agents.

Intuitively, as touched upon in the foregoing discussion, the fact that agents interact repeatedly over a prolonged time period has important implications for team production under moral hazard. Che & Yoo (2001) analyse an environment in which agents can observe one another's effort decisions and condition their future behaviour on current-period actions. In their framework, agents who renege on an implicit agreement to undertake high effort are subjected to an extreme form of 'peer sanction', in which all other agents repeatedly shirk for the remainder of the game. This creates implicit incentives, since agents understand that their effort decisions will affect the future behaviour of others and take this into account when deciding whether to work or shirk.[4]

The allocation of an additional production-relevant decision to one of the agents affects the situation in two key ways. First, the agent with decision making authority (the 'team leader') must be motivated to work

[4]Hamilton et al. (2003) provide empirical evidence for increased cooperation within teams, with increased monitoring ability offering a possible explanation. See also the discussion and literature cited in Lazear & Oyer (2012).

hard *and* to choose the correct production method in each period. The misalignment of preferences between the principal and team leader — due to the fact that the latter's adoption costs vary depending on which production method is selected — then imply that additional constraints must be satisfied if the wage scheme is to remain incentive compatible. This is the drawback of delegation.

Second, the team leader's increased authority expands the potential scope of peer sanction. Following a deviation from the agents' implicit agreement, workers may not only shirk, but now also have the possibility to vary which production method is implemented in a particular period. In other words, delegation implies an increased set of potential 'punishment paths' which can be followed after agents deviate from their agreed-upon actions. Since these punishment paths determine the implicit incentives which are generated by the agents' repeated interaction, adopting an organisational structure of delegation can thus relax the incentive compatibility constraints, allowing for lower explicit incentives in the form of wage payments. However, this is only the case if the agents are actually willing to follow a particular punishment path after a deviation; put differently, the additional sanctions offered by delegation can only benefit the principal if they represent feasible play for the agents.[5]

We provide conditions under which delegation allows the agents to utilise alternative punishment paths, which are not available in a centralised organisation. Moreover, we show that there exist cases in which the increased implicit incentives associated with these alternative sanctions are sufficiently large such that delegation becomes strictly optimal for the principal. This highlights a novel benefit of delegating decision rights to teams. Conferring increased authority over the production process has the ability to generate implicit incentives, via the informal relationships between team members, motivating workers and reducing the necessary explicit incentives in the form of monetary compensation.

By providing a comparative static analysis of the principal's wage costs under each organisational structure, we also use our results to make various predictions regarding the optimal allocation of decision rights within organisations. In particular, our findings suggest that the implicit incentives created by delegation are greatest when the agents have conflicting preferences over the choice of production methods. This follows because, in such cases, the team leader's self-interested behaviour in choosing the production method following a breakdown in team cohesion yields particularly unfavourable outcomes for the subordinate agent and thus represents a harsh sanction. In contrast, when the two agents have similar preferences, the team leader's choice of production method will typically benefit both parties. Our findings therefore provide an additional explanation for the increased real-world adoption of work groups comprised of employees from diverse technical backgrounds (Van Knippenberg & Mell 2016). We also argue that the creation of implicit incentives is complimented most by delegation for larger teams. In addition, our results yield insights regarding various aspects of organisational design, including the role of leadership and the characteristics of leaders, the types of decisions which are most likely to be allocated to empowered teams and the implications of our findings for transparency in organisations and the boundary of the firm.

The analysis proceeds as follows. After the model setup is presented in Section 2, Section 3 studies the benchmark case of a one-shot game under each organisational structure. In the absence of a repeated interaction, the principal cannot rely on implicit incentives to help motivate the agents to undertake effort; accordingly, there is no rationale for allocating decision rights to a member of the team and delegation can never be strictly optimal. In Section 4, we next study the infinitely-repeated game and solve for the principal's choice of payment schedule under each organisational form. We also provide a comparative static

[5] Technically, the expanded action set under delegation allows for a wider range of possible Nash Equilibria in the underlying stage games. This is important, since it is these equilibria which determine the feasible play in the punishment phase, which occurs after either agent has deviated from their implicit agreements when the parties utilise grim-trigger strategies.

analysis of the wage schemes, illustrating how the agents' compensation varies in the parameters of the model. Section 5 compares the principal's costs associated with each of the organisational structures, showing that there exist cases in which delegation can lead to strictly lower total wage costs. We also identify changes in the parameters of the model which make delegation more likely to be optimal for the principal, as well as studying how these changes affect the (asymmetric) wages of the two agents under a delegated organisational structure, which yields insights into which agent should be selected for the team leader position. In Section 6, the foregoing results are used to discuss numerous aspects of organisational design.

In Section 7, we study the possibility of collusion between the agents under a regime of delegation. Specifically, we investigate whether there exists an alternative outcome (which can be sustained as an equilibrium) that offers a Pareto improvement to the workers, given the principal's choice of wage scheme. We show that in many cases implementing the principal's desired outcome in each period also maximises the agents' surplus from the relationship and use this result to provide sufficient conditions for the associated equilibrium to be collusion-proof. However, we also show that there are cases in which the agents' wages become sufficiently low under delegation that they could benefit from coordinating on an alternative equilibrium in which they shirk at least some of the time. Appendix II provides a numerical example of such a situation. In such cases, the principal's design of incentives is scuppered by the possibility of collusion, so that she must either adjust the wage scheme associated with delegation, or revert to a centralised organisational design. Section 8 concludes the paper with a brief discussion of the model's assumptions. All proofs are relegated to Appendix I.

Related Literature. This paper is closely related to two distinct strands of literature. First, a number of papers study the implicit incentives which are created by long-term interactions between workers in teams and investigate the implications for various aspects of organisational design. Che & Yoo (2001) study the interaction between these implicit incentives and the design of the principal's optimal compensation scheme. They show that the agents' abilities to monitor one another provides a rationale for the use of joint performance evaluation, whereby the workers are rewarded as a team. They also discuss the implications of the agents' mutual monitoring for job design, arguing that firms may find it optimal to assign a set of tasks to a team rather than an individual worker, especially if there is a large degree of interdependency in the production process.

Kvaløy & Olsen (2006) build on Che & Yoo's (2001) model by assuming that performance measures are non-verifiable, so that the contract between the principal and agents must be self-enforcing. They show that since the design of the incentive scheme affects the principal's commitment power, the optimal contractual form varies with the abilities of the agents. For low productivity workers, the principal is more likely to utilise relative performance evaluation, improving her commitment ability; however, for high productivity workers she is less tempted to renege on the contract and is thus more likely to use a team-reward scheme.[6]

Ishihara & Muramoto (2021) show that providing agents with rents can influence the strength of implicit incentives if parties can terminate the relationship in response to a deviation from the agreed upon actions. They use this result to study the optimal form of incentive scheme (RPE vs. JPE) when performance measures are unverifiable. Similarly, Ishihara (2017) studies the optimal incentive scheme and degree of cooperation between agents in a multitasking model with helping when contracts are relational.

Two recent papers by Glover & Kim (2020, 2021) investigate how a principal's choice of team composition influences the implicit incentives between a team of agents and use their results to discuss features of organ-

[6]See also Baldenius et al. (2016), who study the case of a verifiable team measure and non-verifiable individual measures of performance.

isational design, while Au (2020) explores the role of severance compensation for team members. Villas-Boas (2020) studies an environment in which team productivity is decreasing over time, showing that the optimal lifespan of a team is chosen to balance productive efficiency with the need to create incentives through a long-term interaction. In contrast to our framework, none of the foregoing studies explore the impact of the delegation of decision making authority to workers within a team.

Second, a small number of recent papers also study the implications of delegating a production-relevant decision to a team of workers. Adrian & Möller (2020) study a team production environment in which a project of uncertain quality must first be selected before agents undertake effort. They show that the principal may prefer to delegate the selection decision if the agents have better information, but also that delegation can prohibit the efficient sharing of information. Intuitively, an agent may choose to conceal 'bad news' regarding a project's quality in order to maintain the effort incentives of his team mates, even though this can lead to a suboptimal choice of project adoption. They show that this breakdown in information sharing within the team is exacerbated by pay disparity, providing a rationale for wage equality in self-organised teams.

Rohlfing-Bastian & Schöttner (2020) analyse the delegation of job design in a multitasking environment with incongruent performance measurement. While the majority of their analysis studies a setting with a single agent, they also extend their analysis to a team setting, showing that this can increase the principal's propensity to delegate the decision. Finally, Kräkel (2017) studies an environment in which a group of workers can be delegated the decision of how to organise themselves into teams. He shows that workers may choose to abuse their authority and purposefully match inefficiently, in order to manipulate the firm's choice of incentive pay and thereby increase their rents. All of these papers consider static environments, so that implicit incentives arising from repeated interaction play no role.

2 Model Setup

We consider a dynamic environment in which a principal employs two agents (denoted by A and B) to undertake team production, which we formalise as an infinitely repeated game. All parties are assumed to be risk-neutral and discount the future according to the common factor $\delta \in (0,1)$. One can think of δ as capturing parties' time preferences, or, alternatively, as a measure of the expected lifetime of the team.

In each period, the agents work on a binary-outcome project which yields verifiable output $R > 0$ to the principal in the case of success, and nothing in the case of failure. The probability of the project being successful is determined by three factors.

- First, each agent makes a choice over their effort provision. Specifically, an agent can choose to either 'work' and incur effort costs of $e > 0$, or 'shirk' at a cost of zero. It is assumed that effort inputs of the two agents are perfect substitutes, such that only the aggregate amount of effort influences the success of the project.

- Second, there exist two possible methods of production, which we denote by γ_1 and γ_2. We assume that agent $i \in \{A, B\}$ incurs a cost $c_i^j \in \mathbb{R}$ when production method γ_j, $j \in \{1,2\}$, is adopted; these 'adoption costs' along with the aforementioned effort costs enter additively into an agent's utility function. No such costs are incurred by the principal. Throughout the paper, we find it convenient to write $\Delta_i = c_i^2 - c_i^1$ as the difference in costs between the two production methods for Agent i. Without loss of generality, we assume that $c_A^2 \geq c_A^1$ so that $\Delta_A \geq 0$; we make no such restriction on the sign of Δ_B. The choice of production method in a particular period cannot be verified to third parties.

– Third, we assume that there are two possible states of the world, ω_1 and ω_2. In each period, state ω_1 is realised with probability r and ω_2 with probability $1-r$. It is assumed that the state of the world is independent between periods and is common knowledge to all parties, but non-verifiable.

As discussed in the introduction, the key idea here is that in many environments, the most appropriate production method can vary between projects or over time, as the information available to the team changes. In our model, these differences in production relevant information are captured by the state of the world in a particular period, which influences the effectiveness of the adopted production method. The notion that the agents may have diverging preferences over alternative production methods is captured by our assumption of adoption costs.[7] Moreover, since the principal has no such direct preference over which method of production is adopted, the existence of these costs creates an additional source of tension between the principal and the agents.[8]

In order to capture the foregoing ideas, we say that γ_j is the 'correct' production method in state ω_k if and only if $j = k$, so that γ_1 is correct in ω_1 and γ_2 is correct in ω_2. Letting $l = \{0, 1, 2\}$ denote aggregate effort provision (i.e. the number of agents who exerted high effort), the probability of the project being successful when the correct production method is adopted is given by p_l. In contrast, we say that a production method γ_j is 'incorrect' in state ω_k when $j \neq k$; the probability of project success in this case, conditional on aggregate effort provision l, is then given by q_l. The following assumption imposes restrictions on these probabilities.[9]

Assumption 1. *The probabilities of project success satisfy the following conditions:*

(i) $p_2 > p_1 > p_0$ *and* $q_2 > q_1 > q_0$

(ii) $p_l > q_l$ *for all* $l \in \{0, 1, 2\}$

(iii) $p_2 - p_1 > p_1 - p_0$ *and* $q_2 - q_1 > q_1 - q_0$

(iv) $p_2 - p_1 > q_2 - q_1$ *and* $p_1 - p_0 > q_1 - q_0$

Part *(i)* of Assumption 1 simply states that effort is always productive, regardless of which production method is adopted in a particular state, while part *(ii)* formalises the notion that adopting the correct production method in each state increases the probability of project success. Part *(iii)* imposes supermodularity upon both the correct and incorrect production technologies and implies that the agents' marginal products are increasing in the effort of teammates. This is a common assumption in the literature, and provides a possible rationale for team production. Similarly, part *(iv)* implies that the agents' marginal products are higher when the correct production technology is adopted. This assumption seems appropriate for many of the situations in which we are interested, such as the assignment of tasks among team members.[10]

The payoff of each agent in each period consists of their income, net of any effort and adoption costs which they incur. The principal's payoff is simply output, net of wage payments made to the agents. All parties are assumed to have an outside option of zero. Throughout the paper, we assume that R is sufficiently large

[7] While it is natural to think of these costs as being positive, we also allow for the possibility that they are negative, so that an agent receives positive utility from adopting a particular production method.

[8] Note that the principal has indirect preferences over which production method is chosen, via its implications for the probability of project success.

[9] A special case in which the probability structure satisfies Assumption 1 is when $q_l = \kappa p_l$, with $\kappa \in (0, 1)$, $p_2 > p_1 > p_0$ and $p_2 - p_1 > p_1 - p_0$. At various points throughout the paper, we impose this particular restriction in order to investigate the implications of changes in the parameter κ, which can be thought of as measuring the relative effectiveness of the incorrect production method, or the importance of choosing the correct production method.

[10] It seems natural that efficient task assignment typically involves workers performing the tasks in which they are most productive.

such that the principal always wishes to achieve the maximum probability of project success. Accordingly, we study the cost minimisation problem for the principal of implementing the correct production technology and high effort from both workers in each period. In the ensuing, we refer to this as being the *principal's desired outcome*.

We assume that the effort supply of both agents is unobservable to the principal and any potential third party, resulting in a moral hazard problem. The principal's choice of wage scheme must then provide appropriate incentives to the agents, and must also satisfy the following conditions. First, due to limited liability of the agents, all wages must be non-negative so that transfers from the agents to the principal are forbidden. Second, our assumptions state that the agents' effort provision, the adopted production method and the state of the world are all non-verifiable, and thus uncontractable. Accordingly, wage payments can be conditional on only the outcome of the principal's projects, success or failure. Third, we assume that the wage scheme must be stationary, so that payments can only depend on the outcome of the current-period project. This also requires that the wage scheme, once chosen at the beginning of the relationship, applies to all subsequent periods.

Due to their close interaction, it is assumed that the agents are perfectly able to observe each other's effort provision in each period. While this introduces the possibility that agents could be required to report their observations to the principal, we assume that such an arrangement is infeasible due to prohibitive costs. We also rule out the possibility of explicit side-contracting between the agents in the form of monetary transfers à la Itoh (1992, 1993). Instead, we focus on the implicit incentives which stem from the informal relationship between the agents in the dynamic game.[11]

In this framework, we are interested in comparing the principal's costs associated with two organisational structures. Under centralisation, the principal observes the state of the world and implements the correct production method in each period; in this case, the wage structure must be designed in such a way that both agents are incentivised to repeatedly provide high effort. Under an organisational design of delegation, decision rights over the choice of production method are assigned to Agent A, who we refer to as the 'team leader'. The wage scheme must then provide incentives not only for the repeated provision of high effort, but also for Agent A to adopt the correct production method in each period.

The timing of the game is as follows. At time $t = 0$, the principal chooses an organisational structure and offers the agents a wage scheme, which is either accepted or rejected. In case of rejection by either party, the game ends. Otherwise, the game continues. At the beginning of each following period, the state of the world is realised and observed by all parties; agents then simultaneously undertake all actions (effort inputs and, potentially, choice of production method), after which the project is either a success or a failure. The principal pays the agents the appropriate wages as specified by the contract and a new period begins.[12]

3 The Static Benchmark

In this section, we analyse the single-period game played under each organisational form, beginning with centralisation. From a technical perspective, we wish to find the wage scheme which implements the principal's desired outcome as a Nash Equilibrium in both states of the world, while minimising total expected wage costs. Since each agent's wage can condition only on the outcome of the principal's project, a wage scheme

[11] Many of these assumptions (e.g. stationary wage contracts, no communication between the principal and agents, the exclusion of formal side-contracting) follow the existing literature and have proved useful for studying various organisational issues, such as job design (Che & Yoo 2001), relational contracts (Kvaløy & Olsen 2006) and team composition (Glover & Kim 2020, 2021). Further justification and discussion of these restrictions can be found in these papers.

[12] Additional discussion of the model's assumptions is provided in the conclusion to the paper.

specifies two payments for each agent, for the cases of success and failure. It is straightforward to verify that the principal cannot benefit from positive transfers following a failed project, since this simultaneously increases wage costs and undermines incentives. We therefore set payments following project failure equal to zero and concentrate on wage schemes of the form $\{w_A, w_B\}$, which specify bonuses for each agent to be paid following a successful project.[13]

3.1 Centralisation

In a centralised organisational structure, the principal observes the state of the world and always chooses to implement the correct method of production. Accordingly, the agents' only decisions concern whether to work or shirk. Given a wage scheme $\{w_A, w_B\}$, Figure 1 illustrates the game played by the agents in state ω_k of the world, which we denote by Γ_k^{cent}.

	work	*shirk*
work	$p_2 w_A - e - c_A^k,\ p_2 w_B - e - c_B^k$	$p_1 w_A - e - c_A^k,\ p_1 w_B - c_B^k$
shirk	$p_1 w_A - c_A^k,\ p_1 w_B - e - c_B^k$	$p_0 w_A - c_A^k,\ p_0 w_B - c_B^k$

Figure 1: The game under centralisation in state ω_k, Γ_k^{cent}.

Since the agents do not decide on the production method, the principal's goal is to design a wage scheme which implements $(work; work)$ as a Nash Equilibrium in both states of the world. This implies that, in state ω_k, Agent i's wage must satisfy the constraint:

$$p_2 w_i - e - c_i^k \geq p_1 w_i - c_i^k \iff w_i \geq \frac{e}{p_2 - p_1} \tag{1}$$

That is, the wage w_i must be sufficiently large to induce high effort, conditional on the other agent working. From (1), it is clear that the incentive compatibility constraints are identical in both states of the world and for both agents. Accordingly, to minimise costs the principal sets $w_i = \frac{e}{p_2 - p_1}$ for $i = A, B$.

3.2 Delegation

When the organisational structure takes the form of delegation, the production method is chosen not by the principal, but solely by Agent A. This implies that Agent A has a total of four possible strategies, consisting of combinations of his binary choice of production method and binary effort decision. Agent B's strategy set, however, is the same as in the case of centralisation. Accordingly, the principal must design the wage scheme $\{w_A, w_B\}$ such that, in both states of the world, *(i)* Agent A chooses to implement the correct method of production and *(ii)* both agents choose to exert high effort. Technically, the requirement is that $(\gamma_1\ and\ work; work)$ and $(\gamma_2\ and\ work; work)$ form Nash Equilibria in states ω_1 and ω_2, respectively. These games are respectively denoted by Γ_1^{del} and Γ_2^{del} and are illustrated by Figures 2 and 3.

[13] Note that whether agents prefer to participate will depend, in particular, on the adoption costs c_i^1 and c_i^2 for $i \in \{A, B\}$. However, since we allow for negative costs, it follows that for any pair (Δ_A, Δ_B), there exists adoption costs which are consistent with both agents' preferring to participate in the production process. Since it is the values Δ_A and Δ_B which turn out to be important for our analysis of incentive provision, rather than the adoption costs themselves, we therefore take the agents' participation as given for the remainder of the paper. In all of the numerical examples considered, it is verified that the agents do indeed wish to participate.

	work	*shirk*
γ_1 *and work*	$p_2 w_A - e - c_A^1,\ p_2 w_B - e - c_B^1$	$p_1 w_A - e - c_A^1,\ p_1 w_B - c_B^1$
γ_1 *and shirk*	$p_1 w_A - c_A^1,\ p_1 w_B - e - c_B^1$	$p_0 w_A - c_A^1,\ p_0 w_B - c_B^1$
γ_2 *and work*	$q_2 w_A - e - c_A^2,\ q_2 w_B - e - c_B^2$	$q_1 w_A - e - c_A^2,\ q_1 w_B - c_B^2$
γ_2 *and shirk*	$q_1 w_A - c_A^2,\ q_1 w_B - e - c_B^2$	$q_0 w_A - c_A^2,\ q_0 w_B - c_B^2$

Figure 2: The game under delegation in state ω_1, Γ_1^{del}.

We begin by deriving the resulting constraints on Agent A's wage. First, let us consider the game played in state ω_1. Note that any strategy which implements production method γ_2 can never be a best response for Agent A in this state, since, for a given level of effort provision, implementing γ_1 leads to both a higher probability of project success and lower adoption costs (as $\Delta_A \geq 0$, and therefore $c_A^2 \geq c_A^1$, by assumption). Accordingly, in state ω_1, w_A needs to satisfy only (1) in order to implement γ_1 *and work* as a best response to high effort from Agent B.

	work	*shirk*
γ_1 *and work*	$q_2 w_A - e - c_A^1,\ q_2 w_B - e - c_B^1$	$q_1 w_A - e - c_A^1,\ q_1 w_B - c_B^1$
γ_1 *and shirk*	$q_1 w_A - c_A^1,\ q_1 w_B - e - c_B^1$	$q_0 w_A - c_A^1,\ q_0 w_B - c_B^1$
γ_2 *and work*	$p_2 w_A - e - c_A^2,\ p_2 w_B - e - c_B^2$	$p_1 w_A - e - c_A^2,\ p_1 w_B - c_B^2$
γ_2 *and shirk*	$p_1 w_A - c_A^2,\ p_1 w_B - e - c_B^2$	$p_0 w_A - c_A^2,\ p_0 w_B - c_B^2$

Figure 3: The game under delegation in state ω_2, Γ_2^{del}.

Next, we consider the game played in state ω_2. In this case, in order to implement γ_2 *and work* as a best response to high effort from Agent B, w_A must satisfy (1), along with the following additional constraints:

$$p_2 w_A - e - c_A^2 \geq q_2 w_A - e - c_A^1 \iff w_A \geq \frac{c_A^2 - c_A^1}{p_2 - q_2} = \frac{\Delta_A}{p_2 - q_2} \tag{2}$$

$$p_2 w_A - e - c_A^2 \geq q_1 w_A - c_A^1 \iff w_A \geq \frac{e + c_A^2 - c_A^1}{p_2 - q_1} = \frac{e + \Delta_A}{p_2 - q_1} \tag{3}$$

which guarantee that the correct production method, γ_2, will be implemented. (2) and (3) require that γ_2 *and work* yields a higher expected utility to Agent A when Agent B works hard than γ_1 *and work* and γ_1 *and shirk*, respectively. Altogether, in order to minimise costs while satisfying (1)-(3), the principal sets w_A equal to:

$$\max\left\{\frac{e}{p_2 - p_1}, \frac{\Delta_A}{p_2 - q_2}, \frac{e + \Delta_A}{p_2 - q_1}\right\}$$

under an organisational structure of delegation.

We next consider the requirements on w_B. Since Agent B has no control over the chosen production method, it is sufficient that his wage is large enough to induce high effort in either state of the world, given that Agent A implements the correct method of production and also undertakes high effort. Accordingly, as before, w_B needs only to satisfy (1), and is therefore identical to the wage paid under centralisation in the static benchmark.

3.3 Discussion

Under either organisational structure, the wage payment for each agent must be sufficiently large to induce high effort, given that the correct production method has been adopted and that the other agent is similarly working hard. Accordingly, the wage payments (to both agents) must satisfy (1) under both centralisation and delegation. Since Agent B does not select the production method under either organisational structure, for him (1) is the sole requirement and hence his wages are identical in both cases. In contrast, delegation imposes additional constraints on the wage paid to Agent A, since he must be induced to both work hard and to select the correct method of production; this implies that (2) and (3) must also hold.

Intuitively, when selecting the production method, Agent A takes two separate considerations into account. First, by choosing the correct method of production, he can increase the probability of project success and therefore his chances of receiving the wage payment. Second, due to the difference in adoption costs $\Delta_A \geq 0$, he always has an inclination to adopt production method γ_1, regardless of whether it maximises the probability of project success. In state ω_1, γ_1 is the correct production method and hence there is no tension between the two objectives. However, in state ω_2, implementing the correct method of production requires adopting γ_2 and incurring higher costs. It follows that in state ω_2, Agent A will only implement the correct production method γ_2 if the wage payment is sufficiently large to compensate for this increase in adoption costs, i.e., if (2) and (3) are satisfied.

Whether either of these additional constraints bind will depend in particular on the size of Δ_A. For $\Delta_A = 0$, Agent A faces the same adoption costs for each production method and hence always prefers to implement the correct one. Similarly, for low values of Δ_A, the wage payment required to induce high effort is also sufficiently large to motivate the agent to always select the correct production method. In these cases, there is no drawback to a regime of delegation: the interests of the principal and Agent A are sufficiently aligned such that wages are the same under both organisational structures. However, as Δ_A grows large, eventually one of the constraints (2) or (3) will bind; in these cases, Agent A's preference for γ_1 is sufficiently strong that further incentives are required to ensure that the correct production method is always chosen. Delegation then becomes strictly worse than centralisation, due to the higher wage payment which is required for Agent A.

In summary, delegation can never be strictly optimal in the static environment. At best, it can replicate the wage structure under centralisation — but only when Δ_A is low. For higher values of Δ_A, delegation is strictly inferior as an organisational design, due to the necessary increase in Agent A's wage.

4 The Dynamic Game

In this section, we study the full dynamic game under each organisational form in turn. At the start of the dynamic game, the principal fixes the organisational structure. Thereafter, the state of the world is realised at the start of each period and the agents play the stage game which is associated with this combination of state and organisational form. A strategy for each agent is a function which maps from any possible history of past play (including the realisation of states) and the current state into a probability distribution over actions.

As is well-known, infinitely repeated games typically admit a vast set of subgame-perfect equilibria. We shall follow the literature and restrict attention to equilibria in which agents condition on only the state and actions undertaken in the previous period and utilise grim-trigger strategies. In such equilibria, any deviation from an agreed-upon action profile is punished by repeated play of stage game equilibria for the remainder of the game, maximising the implicit incentives generated by the agents' interaction. Technically, our goal is then to derive the incentive scheme which minimises the principal's expected wage costs, while implementing a subgame-perfect equilibrium in which both agents work hard and the correct production method is selected in each period. As outlined in Section 2, the model's assumptions imply that the principal selects a stationary wage scheme, which conditions payments to each agent only on current-period project success and, once set at the beginning of the relationship, applies to all subsequent periods. As before, we set payments following a failed project equal to zero and concentrate on the wages $\{w_A, w_B\}$ which are paid following project success.

As shown by Che & Yoo (2001), an infinitely-repeated relationship between agents in a team-production context introduces substantial differences compared to the one-shot game. This stems from the fact that agents are able to monitor one another's effort decisions in each period and adjust their future play accordingly. Intuitively, agents understand that shirking will be observed by other agents and can lead to a breakdown in team cohesion and peer sanction, whereby all other agents shirk in future periods. This so-called *mutual monitoring* between agents creates implicit incentives, allowing for a reduction in the explicit incentives provided by the principal.

In our framework, it is always optimal for the principal to design the wage scheme such that both agents repeatedly shirk following a deviation from the agreed-upon action profile, as in Che & Yoo (2001).[14] However, implementing an organisational structure of delegation changes the scope of peer sanction, since Agent A can now control the production methods which are adopted during the punishment phase of the game. This modifies the implicit incentives provided by the agents' mutual monitoring, leading to differences in the required wage payments under the two organisational designs.

4.1 Centralisation

Similar to the one-shot game, under an organisational structure of centralisation, in each period the principal observes the state of the world and implements the correct production method. Accordingly, the agents' only decisions are whether to *work* or *shirk*, so that the stage game played in state ω_k is Γ_k^{cent}, illustrated by Figure 1. From the foregoing discussion, we wish to derive the wage scheme which implements high effort from both agents in all periods as a subgame-perfect equilibrium. Technically, this requires that $(work; work)$ is played in each period, regardless of the state of the world. As we restrict attention to grim-trigger strategies, we proceed in two steps. In the first step, we derive the incentive compatibility constraints which must be satisfied in order to induce high effort from both workers, taking as given the play on the 'punishment path' following a deviation from the principal's desired outcome (i.e. shirking by either agent). In the second step, we identify the circumstances under which repeated shirking by both agents can be feasibly sustained during this punishment phase of the game and use this analysis to solve for the principal's optimal wage scheme.

Step One. Let P_i denote the expected per-period utility of Agent i during the punishment phase of the

[14] Intuitively, this follows from the fact that each agent's effort has a positive effect on their teammate's expected payoff, so that designing a wage scheme that can sustain a stage-game Nash Equilibrium in which one of the agents works with a positive probability must be associated with a weakly higher expected payoff for both parties. Clearly, this cannot provide a harsher punishment than repeated shirking and thus cannot be optimal.

game. For a particular P_i, in order to induce high effort the wage payment w_i must satisfy the following incentive compatibility constraints:

$$(1-\delta)(p_2 w_i - e - c_i^1) + \delta(p_2 w_i - e - r c_i^1 - (1-r)\, c_i^2) \geq (1-\delta)(p_1 w_i - c_i^1) + \delta P_i \tag{4}$$

$$(1-\delta)(p_2 w_i - e - c_i^2) + \delta(p_2 w_i - e - r c_i^1 - (1-r)\, c_i^2) \geq (1-\delta)\left(p_1 w_i - c_i^2\right) + \delta P_i \tag{5}$$

for states ω_1 and ω_2 respectively. The first of the two terms on the LHS of (4) represents the current-period utility from choosing high effort, while the second describes the discounted sum of expected future utility from repeatedly playing the principal's desired strategy in all subsequent periods, where both of these terms have been multiplied by $1-\delta$. Similarly, the first term on the RHS of the inequality shows the current-period utility from shirking, while the second represents the discounted sum of expected future utility from the punishment outcome being played repeatedly, P_i; again, both terms have been multiplied by $1-\delta$. (5) can be explained similarly, for the second state of the world. Incentive compatibility therefore requires that each agent's discounted sum of expected utility is higher when sticking to the agreed-upon action profile, than it is when the agent shirks in one period with the punishment equilibrium being played thereafter. Cancelling terms, it is easily verified that the two constraints (4) and (5) are equivalent.

Step Two. Next, we consider the play on the punishment path after shirking by either agent. As discussed in the introduction to this section, the harshest possible punishment that workers can impose on one another under centralisation is repeated shirking for the remainder of the game. Under what circumstances is this play feasible? To investigate this issue, first note that any wage payment w_i for which:

$$p_2 w_i - e < p_1 w_i \iff w_i < \frac{e}{p_2 - p_1} \tag{6}$$

also satisfies:

$$p_1 w_i - e < p_0 w_i \iff w_i < \frac{e}{p_1 - p_0} \tag{7}$$

by our assumption of supermodularity (Assumption 1, *iii*). From observation of Figure 1, it follows that when w_i satisfies (6), shirking is a strictly dominant strategy for Agent i in the stage game in both states of the world (i.e. in both Γ_1^{cent} and Γ_2^{cent}). Intuitively, in the one-shot game, if agents do not find it beneficial to exert high effort when their teammate works hard, then by complementarity of efforts they will also shun high effort when their teammate is shirking. Accordingly, when both w_A and w_B satisfy (6), the outcome $(shirk; shirk)$ is the unique Nash Equilibrium of both stage games and thus repeated shirking on the punishment path is feasible.

This play yields an expected per-period utility to Agent i during the punishment phase of:

$$P_i^{cent} := p_0 w_i - r c_i^1 - (1-r)\, c_i^2 \tag{8}$$

Setting $P_i = P_i^{cent}$ in (4) and rearranging then yields the constraint:

$$w_i \geq \frac{e}{p_2 - (1-\delta)p_1 - \delta p_0} := w^{cent} \tag{9}$$

It is straightforward to verify that w^{cent} satisfies (6), so that when $w_A = w_B = w^{cent}$, $(shirk, shirk)$ is the unique Nash Equilibrium of the stage-game in both states; accordingly this wage structure implements repeated play of $(work; work)$ in both states as a subgame-perfect equilibrium under centralisation at the lowest possible costs to the principal.

As shown by Che & Yoo (2001), under a centralised organisational structure the wage paid to both agents in the repeated game, w^{cent}, is strictly lower than that paid in the one-shot game, $\frac{e}{p_2 - p_1}$. This reflects the discussion in the introduction to this section, namely that the implicit incentives provided by the agents' mutual monitoring allows for lower explicit incentives in the form of wage payments, reducing the principal's costs. To conclude the subsection, Proposition 1 outlines the comparative statics associated with the principal's wage costs under centralisation.

Proposition 1. *The principal's optimal wage payment to the agents under centralisation, w^{cent}, is:*

(i) Increasing in e.

(ii) Decreasing in δ.

(iii) Invariant to r, c_i^1 and c_i^2 for $i \in \{A, B\}$, and changes in q_l, $l \in \{0, 1, 2\}$.

The intuition behind part *(i)* of the proposition is straightforward, since increased effort costs necessitate higher payments in order for the wage scheme to remain incentive compatible. For part *(ii)*, note that as emphasised in the foregoing discussion, in the dynamic game the threat of a breakdown in team cohesion (and the ensuing reduction in future-period expected wages) provides implicit incentives to workers. When δ is large, the agents place a higher value on wages received in future periods. This increases these implicit incentives, allowing for a reduction in payments. Finally, under an organisational structure of centralisation, the correct production method is implemented in each period by the principal, so that the wage scheme needs only to provide sufficient incentives for the agents to provide high effort. Accordingly, the agents choices cannot influence which production method is adopted, either in current or future periods, so that the variables outlined in part *(iii)* of the proposition play no role.

4.2 Delegation

Under an organisational structure of delegation, it is the duty of Agent A to choose which production method is adopted, after observing the state of the world. The underlying stage games in states ω_1 and ω_2 are again respectively given by Γ_1^{del} and Γ_2^{del}, displayed in Figures 2 and 3. Our goal is to derive an incentive scheme which implements high effort from both agents in all periods, as well as inducing the correct choice of production method from Agent A, as a subgame-perfect equilibrium. Technically, this requires that $(\gamma_1 \textit{ and } work; work)$ is always played in state ω_1 and $(\gamma_2 \textit{ and } work; work)$ is always played in state ω_2.

As in the previous subsection, we proceed in two steps. We first study the incentive compatibility constraints which the wages paid to each agent must satisfy, taking as given the play on the punishment path. We then consider the outcomes which can be credibly sustained during the punishment phase of the game, in particular showing that delegation allows for alternative punishment equilibria to those studied in the case of centralisation. We then use these observations to solve for the principal's optimal wage scheme.

Step One. Let us again denote P_i as the expected per-period utility of Agent i during the punishment phase of the game. Since under delegation the agents play an asymmetric role in the production process, we consider the incentive compatibility constraints for w_A and w_B in turn, taking P_i as given. In state ω_1, the incentive compatibility constraints for Agent A are (4), along with:

$$(1-\delta)(p_2 w_A - e - c_A^1) + \delta(p_2 w_A - e - r c_A^1 - (1-r)\, c_A^2) \geq (1-\delta)(q_2 w_A - e - c_A^2) + \delta P_A \tag{10}$$

$$(1-\delta)(p_2 w_A - e - c_A^1) + \delta(p_2 w_A - e - r c_A^1 - (1-r)\, c_A^2) \geq (1-\delta)(q_1 w_A - c_A^2) + \delta P_A \tag{11}$$

The LHS of all three constraints is identical, and consists of the current-period utility from choosing high effort and implementing γ_1, as well as the discounted sum of expected utility from repeatedly playing the principal's desired strategy in all future periods. The RHS of the three constraints comprises the respective immediate benefits from the three possible deviations from this strategy, along with the discounted sum of future expected utility from the punishment outcome being played repeatedly. Similarly, in state ω_2 Agent A's incentive compatibility constraints are (5), along with:

$$(1-\delta)(p_2 w_A - e - c_A^2) + \delta(p_2 w_A - e - r c_A^1 - (1-r)\, c_A^2) \geq (1-\delta)\left(q_2 w_A - e - c_A^1\right) + \delta P_A \tag{12}$$

$$(1-\delta)(p_2 w_A - e - c_A^2) + \delta(p_2 w_A - e - r c_A^1 - (1-r)\, c_A^2) \geq (1-\delta)\left(q_1 w_A - c_A^1\right) + \delta P_A \tag{13}$$

which can be explained analogously to those above.

Altogether, the wage w_A must satisfy six constraints. (4) and (5) guarantee that Agent A prefers to work hard rather than shirking, under implementation of the correct production method, and are thus identical to those under centralisation. The constraints (10)-(13) are the additional restrictions associated with delegation, which stem from Agent A's increased discretion over the production process. Since Agent B only faces one decision in each state — whether to *work* or *shirk* — his role in the production process is unchanged from the case of centralisation; accordingly, the necessary incentive compatibility constraints are simply (4) and (5), as before.

Step Two. Next, we move onto the question of which outcomes can be feasibly sustained in the punishment phase of the game. As discussed in the introduction to this section, the principal will always find it optimal to design the wage scheme such that both agents shirk repeatedly on the punishment path. However, a delegated organisational structure allows for alternative punishment equilibria compared to the case of centralisation, due to Agent A's increased decision making authority.

To begin the analysis, first note that any punishment equilibrium which involves Agent A implementing production method γ_2 in state ω_1 is infeasible. To see this, observe that in the stage game Γ_1^{del} illustrated by Figure 2, the strategies γ_2 *and work* and γ_2 *and shirk* are strictly dominated by γ_1 *and work* and γ_1 *and shirk*, respectively. Intuitively, implementing γ_2 in state ω_1 always results in a strictly worse outcome for Agent A, since not only is the probability of project success reduced due to the incorrect production method being chosen, but he also incurs higher adoption costs as $\Delta_A \geq 0$. It follows that any threat to implement γ_2 in state ω_1 during the punishment phase of the game is incredible.

We therefore only need to consider the production method which is implemented in state ω_2 during the punishment phase of the game. This is determined by Agent A's best response to Agent B's shirking in this state, which itself depends on two factors: the size of w_A and the size of Δ_A. To see this, note that in state

ω_2, Agent A's best response will depend on the direction of the following inequality:

$$p_0 w_A - c_A^2 \lesseqqgtr q_0 w_A - c_A^1 \tag{14}$$

The left- and right-hand sides of (14) are the expected utilities from implementing the correct (γ_2) and incorrect (γ_1) production methods, respectively. Rearranging yields:

$$w_A \left(p_0 - q_0\right) \lesseqqgtr \Delta_A \tag{15}$$

Intuitively, by choosing the correct production method in state ω_2, Agent A can maximise the probability of project success, but must also incur higher adoption costs since $\Delta_A \geq 0$. His best response therefore depends on the relative magnitude of these two effects. Since w_A must be set by the principal to be sufficiently large to induce high effort, for low values of Δ_A it must be the case that the LHS of (15) is greater than the RHS. Accordingly, when Δ_A is small, Agent A always prefers to select γ_2 in state ω_2 during the punishment phase of the game.

In contrast, when Δ_A becomes sufficiently large, it may be possible that the principal can choose a w_A which is incentive compatible, but still sufficiently small such that the RHS of (15) is greater. However, in this case the principal also has the option of choosing a higher w_A, so that the LHS of is greater. In other words, for Δ_A sufficiently large, the principal can choose which punishment outcome is implemented by Agent A, by varying the size of the wage payment w_A. In the following, we consider these two cases — low and high values of Δ_A — in turn.

Low Δ_A. From the foregoing discussion, we know that when Δ_A is small, production method γ_1 will be adopted in state ω_1 during the punishment phase of the game, and production method γ_2 in state ω_2. Since the correct production method is implemented in both states, this mirrors the punishment equilibrium under centralisation, so that the expected per-period utility to Agent i is once again given by P_i^{cent}. Moreover, the proof of the following proposition shows that the additional constraints on w_A associated with delegation, (10)-(13), never bind in this case, so that the wage scheme takes an identical form to that under a centralised organisational structure.

Proposition 2. *Suppose we have:*

$$\widehat{\Delta}_A := \frac{e\left[p_0 - q_0\right]}{p_2 - (1-\delta)p_1 - \delta p_0} > \Delta_A \tag{16}$$

Then under an organisational structure of delegation, the principal minimises the costs of implementing her desired outcome as a subgame-perfect equilibrium by setting:

$$w_A = w_B = w^{cent} = \frac{e}{p_2 - (1-\delta)p_1 - \delta p_0} \tag{17}$$

The intuition behind this result is straightforward. As in the case of centralisation, the principal sets the wage sufficiently high to induce *work* from both agents in each period, conditional on the correct production method always being implemented. Moreover, since the difference in costs between the two production methods for Agent A is relatively small, this wage is also sufficient to always induce the correct choice of production method — both in the initial phase of the game, and in the punishment phase after a deviation by either party. In other words, when Δ_A is low, the interests of Agent A and the principal regarding the choice of production method are sufficiently aligned such that delegation essentially replicates centralisation.

Moreover, as outlined in the foregoing discussion, it is not possible for the principal to design the wage scheme in such a way that an alternative punishment equilibrium is utilised by the agents when Δ_A is small, since offering a lower wage is inconsistent with inducing high effort from Agent A.[15] This implies that the foregoing wage scheme minimises the principal's costs.

High Δ_A. From the foregoing discussion, for higher values of Δ_A the principal can design the wage scheme to induce either of the two punishment equilibria under consideration. However, in order to implement the same punishment equilibrium as under centralisation (i.e. the correct production methods in both state) it must be the case that the LHS of (15) is greater than the RHS; for $\Delta_A > \widehat{\Delta}_A$, it is easily verified that this requires $w_A > w^{cent}$, so that Agent A's wage is strictly higher than under centralisation. Moreover, since play on the punishment path is identical to centralisation, Agent B's wage must continue to satisfy $w_B \geq w^{cent}$. It then follows that the principal's wage expenditure would be strictly higher under delegation than under centralisation; accordingly, designing a wage scheme which implements the same punishment strategy as centralisation when we have $\Delta_A > \widehat{\Delta}_A$ is strictly suboptimal for the principal relative to centralisation.

We therefore focus on the alternative punishment equilibrium, in which Agent A chooses to implement γ_1 in both states of the world along the punishment path. In this case, the expected per-period utility for Agent i during the punishment phase of the game is:

$$P_i^{del} := r\left(p_0 w_i - c_i^1\right) + (1-r)\left(q_0 w_i - c_i^1\right) = rp_0 w_i + (1-r)\, q_0 w_i - c_i^1 \tag{18}$$

This differs from the expected per-period utility under the centralisation punishment in two respects. First, since the incorrect production method is implemented whenever state ω_2 is realised, the agents face a strictly lower probability of receiving their bonuses. Second, since γ_1 is now implemented in every period during the punishment phase, the expected adoption costs of the agents changes following a deviation by either agent. While the first of these effects is negative for both agents, the second can be positive or negative, depending on their preferences over the different production methods.

In order to analyse the implications of this change in expected utility during the punishment phase for the incentive compatibility constraints, let us first define:

$$\Delta_A^{max} := \frac{e\left[p_1 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\, q_0\right]\right]}{(1-\delta r)\,(p_2 - p_1)} \tag{19}$$

$$\Delta_B^{max} := \frac{e\left[p_1 - rp_0 - (1-r)\, q_0\right]}{(1-r)\,(p_2 - p_1)} \tag{20}$$

both of which are strictly positive. We then have the following result.

Proposition 3. *Suppose we have $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$. Then under an organisational structure of delegation, the principal can implement her desired outcome as a subgame perfect equilibrium by*

[15] This argument is formalised in the proof of Proposition 2.

setting:

$$w_A^{del} = \max\left\{\frac{e+\delta(1-r)\Delta_A}{p_2-(1-\delta)p_1-\delta[rp_0+(1-r)q_0]}, \frac{e+(1-\delta r)\Delta_A}{p_2-(1-\delta)q_1-\delta[rp_0+(1-r)q_0]}\right\} \tag{21}$$

$$w_B^{del} = \frac{e+\delta(1-r)\Delta_B}{p_2-(1-\delta)p_1-\delta[rp_0+(1-r)q_0]} \tag{22}$$

It is often more convenient for us to rewrite (21) as:

$$w_A^{del} = \begin{cases} \frac{e+\delta(1-r)\Delta_A}{p_2-(1-\delta)p_1-\delta[rp_0+(1-r)q_0]} & \Delta_A \leq \Delta_A^{crit} \\ \\ \frac{e+(1-\delta r)\Delta_A}{p_2-(1-\delta)q_1-\delta[rp_0+(1-r)q_0]} & \Delta_A > \Delta_A^{crit} \end{cases} \tag{23}$$

where:

$$\Delta_A^{crit} = \frac{(p_1-q_1)\,e}{p_2-p_1(1-\delta r)+q_1\delta(1-r)-\delta[rp_0+(1-r)q_0]} \tag{24}$$

is strictly positive and satisfies $\Delta_A^{crit} \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$.

The proof of Proposition 3 shows that w_A^{del} and w_B^{del} are the lowest possible wages which satisfy the required incentive compatibility constraints, conditional on the punishment equilibrium in which (γ_1 *and shirk*; *shirk*) is repeatedly played in both states of the world. The proof also shows that when $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$, ($\gamma_1$ *and shirk*; *shirk*) is the unique stage game equilibrium of the games Γ_1^{del} and Γ_2^{del} when $w_A = w_A^{del}$ and $w_B = w_B^{del}$, so that this punishment path is indeed feasible for the agents. As emphasised in the foregoing discussions, we require Δ_A to be sufficiently large ($\Delta_A > \widehat{\Delta}_A$) such that Agent A prefers to implement γ_1 in state ω_2. However, we also require that Δ_A and Δ_B are not too large ($\Delta_A < \Delta_A^{max}$, $\Delta_B < \Delta_B^{max}$). To see why, note that w_i^{del} is increasing in Δ_i for both agents, so that large values of Δ_i are associated with high wages. For $\Delta_A > \Delta_A^{max}$ or $\Delta_B > \Delta_B^{max}$, it is possible that wages become so large that agents prefer to work even when their coworker is shirking, in which case (γ_1 *and shirk*; *shirk*) fails to be a Nash Equilibrium of the underlying stage games.[16]

The proof also shows that since $0 < \widehat{\Delta}_A < \Delta_A^{max}$, the interval $\left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ is always non-empty. Hence, for any set of remaining model parameters, there will always exist values of Δ_A and Δ_B such that the foregoing wage scheme can implement the principal's desired outcome as a subgame-perfect equilibrium under delegation.

There are two reasons why these payments to the agents differ to w^{cent}, the wage offered under centralisation. First, as discussed in the foregoing analysis, the fact that the outcome (γ_1 *and shirk*; *shirk*) is implemented on the punishment path whenever state ω_2 is realised, rather than (γ_2 *and shirk*; *shirk*) in the case of centralisation, leads to differences in expected utilities between the two organisational structures during the punishment phase of the game. This affects the strength of the implicit incentives, provided by the agents' mutual monitoring, and therefore changes the wages which are necessary to ensure incentive compatibility.

The second reason for the difference in wages associated with the two organisational forms is related to

[16]Moreover, as these cases are associated with high wages for at least one agent, delegation is unlikely to be the optimal organisational design for the principal. Since our goal is not to provide an exhaustive analysis of the principal's optimal choice of organisational structure for all possible cases, but rather to analyse the trade-offs inherent in the choice, a formal analysis of cases where $\Delta_A > \Delta_A^{max}$ or $\Delta_B > \Delta_B^{max}$ is thus omitted.

Agent A's increased responsibility under delegation. In this case, his wage must be designed to induce both high effort *and* the correct choice of production method in each period, as we saw in the static benchmark. When his preferences over the production method are relatively weak ($\Delta_A \leq \Delta_A^{crit}$), this additional concern does not play a role. Intuitively, if there is a small difference between the adoption costs associated with each production method, then the wage necessary to induce high effort is already sufficient to also induce the correct production method choice; i.e. (4) binds. In other words, his best deviation from the principal's desired strategy does not involve selecting the incorrect production method. However, for $\Delta_A > \Delta_A^{crit}$, Agent A's preferences over the production methods are relatively strong; in this case, his optimal deviation *does* involve choosing the incorrect production method, so that (13) now binds and the wage must be adjusted further. Proposition 4 provides a comparative static analysis of the agents' wages $\{w_A^{del}, w_B^{del}\}$, followed by a discussion of the associated intuitions.

Proposition 4. *Let $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$. The wages w_A^{del} and w_B^{del} have the following properties:*

(i) w_A^{del} and w_B^{del} are both strictly increasing in e.

(ii) w_A^{del} and w_B^{del} are strictly increasing in Δ_A and Δ_B, respectively.

(iii) w_A^{del} and w_B^{del} are both strictly decreasing in δ.

(iv) w_A^{del} is strictly decreasing in r for all Δ_A. w_B^{del} is strictly decreasing in r iff. Δ_B is sufficiently large; in particular, for all $\Delta_B \leq 0$, w_B^{del} is strictly increasing in r.

(v) In the special case where $q_l = \kappa p_l$ for $l \in \{0, 1, 2\}$, w_A^{del} and w_B^{del} are both strictly increasing in κ.

Part *(i)* of Proposition 4 states that, analogous to the payment under centralisation, higher wages are required to maintain incentive compatibility as effort costs increase. Similarly, part *(iii)* implies that wages can be lowered as δ increases, since the agents put a higher weight on future outcomes, increasing the strength of the implicit incentives provided by the informal relationship between the agents.

The intuition behind part *(ii)* of the proposition follows directly from the foregoing discussion, since the punishment equilibrium prescribes implementing production method γ_1 in both states of the world. The stronger the agents' preferences for this production method are, the less harsh the punishment and hence the higher wages need to be to induce effort. Moreover, when $\Delta_A > \Delta_A^{crit}$, there is an additional effect, since a higher w_A^{del} is required in order to prevent Agent A deviating from the principal's desired strategy by adopting the incorrect production method.

Part *(iv)* of the proposition is a little more complicated. Observing the constraints (4) and (13) along with (18), one can see that an increase in r has two distinct effects. The first of these is to increase the expected utility associated with the punishment phase, (18), which has a positive effect on wages since implicit incentives are reduced. The second effect is to change the agents' expected utilities associated with sticking to the principal's desired strategy: as r increases, state ω_1 occurs more often and hence γ_1 is implemented at a higher rate. For $\Delta_i > 0$, this effect increases expected utility, allowing for a reduction in wages, while the reverse is true for $\Delta_i < 0$. For Agent A, one can show that the restriction $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ implies that Δ_A is always sufficiently large such that the second effect dominates and thus wages become lower. For Agent B, however, the direction of the total effect is unclear and depends on both the sign and magnitude of Δ_B; in particular, when we have $\Delta_B < 0$, both of the aforementioned effects are positive and hence w_B^{del} must be increased.

Finally, for part *(v)*, the effect of an increase in κ is to increase the agents' expected utilities in the punishment phase, since the incorrect production technology becomes less inferior and thus the probability of project success increases. Moreover, for $\Delta_A > \Delta_A^{crit}$, the increase in κ also makes deviating to the incorrect production method more attractive for Agent A. Both of these effects work to increase the wages necessary to maintain incentive compatibility.

5 Centralisation vs. Delegation

The previous section outlined the key trade-off between centralisation and delegation in the current framework. While delegation imposes additional constraints upon the principal's choice of wage scheme, due to Agent A's increased control over the production process, it can also allow for alternative punishment equilibria for the same reason. By influencing the implicit incentives associated with the agents' mutual monitoring, this alternative play during the punishment phase of the game implies that the wages under the two organisational forms can become disparate. In this section, we are primarily interested in the following question: can this change in the implicit incentives provided by the agents' long-term interaction ever lead to delegation becoming optimal as an organisational structure?

For low values of $\Delta_A < \widehat{\Delta}_A$, we showed that delegation essentially replicates the case of centralisation, so that the principal's wage costs are identical under the two organisational designs. Accordingly, we focus on the case of $\Delta_A > \widehat{\Delta}_A$, where the wage scheme induces an alternative punishment equilibrium to that under centralisation. In this case, the key question is whether this alternative punishment outcome is able to provide a harsher sanction to the agents.

Recall that under delegation with $\Delta_A > \widehat{\Delta}_A$, during the punishment phase of the game the outcome (γ_1 *and shirk*; *shirk*) is implemented in state ω_2, compared to (γ_2 *and work*; *work*) in the case of centralisation. Subtracting (8) from (18) yields the ensuing difference in expected punishment utility for Agent i across the two organisational structures:

$$P_i^{del} - P_i^{cent} = (1-r)\left[-\left(p_0 - q_0\right) w_i + \Delta_i\right] \tag{25}$$

This difference is made up of two distinct effects. First, in state ω_2 the probability of project success falls from p_0 to q_0 due to the incorrect production method being implemented. This has a negative impact on the expected punishment utilities of both workers, since the probability with which they receive their wages is reduced. Second, in ω_2 the production method γ_1 is implemented rather than γ_2, leading to a change in adoption costs. For Agent A this effect is positive, since $\Delta_A \geq 0$, but it can be either positive or negative for Agent B depending on the sign of Δ_B.[17]

Whether delegation is able to provide a more or less harsh punishment to the agents depends on the aggregation of these effects. For Agent A, we know that expected utility in the punishment phase must be higher under delegation than under centralisation, since $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ guarantees that (γ_1 *and shirk*; *shirk*) is a Nash Equilibrium of the stage game Γ_2^{del}, by Proposition 3. Intuitively, since in the punishment phase Agent B repeatedly shirks, Agent A's actions must form a best response to this action. For (γ_1 *and shirk*; *shirk*) to be a Nash Equilibrium in state ω_2, it must then be the case that Agent A's expected payoff is higher from playing γ_1 *and shirk* rather than γ_2 *and shirk*. Since this latter action is associated with the punishment equilibrium under centralisation, it follows that Agent A must be strictly better off in the game's punishment

[17] To be precise, under delegation the expected probability of project success during the punishment phase is reduced from p_0 to $rp_0 + (1-r)\, q_0$, while the difference in expected adoption costs for Agent i is $(1-r)\left(c_i^2 - c_i^1\right) = (1-r)\Delta_i$.

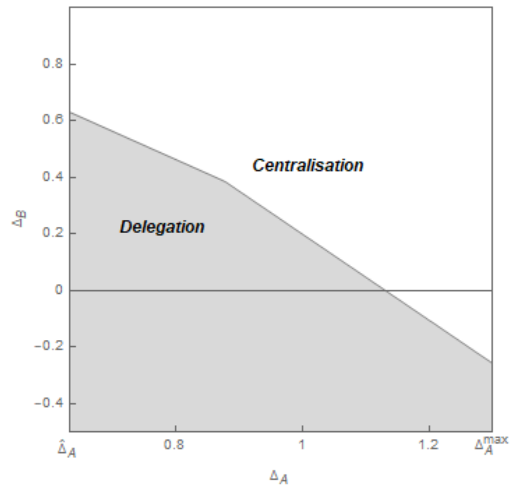

Figure 4: The optimal organisational structure as Δ_A and Δ_B vary.

phase under delegation. In other words, since Agent A controls the choice of production method and thus the equilibrium played in the punishment phase, he is only willing to implement the alternative delegation punishment if it is better for him than the centralisation punishment.

For Agent B, on the other hand, the overall effect is unclear, and in particular depends on Δ_B. When Δ_B is positive and large, the reduction in expected adoption costs imply that Agent B similarly prefers the alternative punishment equilibrium under delegation — in this case, the punishment under delegation is less harsh for both agents. However, for sufficiently low values of Δ_B, (25) becomes negative, since the change in expected adoption costs is no longer sufficiently high to outweigh the lower probability of project success; in fact, when $\Delta_B < 0$, both of the aforementioned effects are negative, so that (25) necessarily becomes strictly less than zero. In these cases, Agent B's punishment utility is strictly lower under delegation than centralisation.

Altogether, our analysis suggests the following. The wage paid to Agent A under an organisational design of delegation when $\Delta_A > \widehat{\Delta}_A$ will always be higher than w^{cent}; this is due to the combination of a less harsh punishment path and the additional incentives which may be required in order to ensure the correct production method is implemented (when $\Delta_A > \Delta_A^{crit}$). The wage paid to Agent B, on the other hand, can be either higher or lower than w^{cent}, depending on the relative harshness of the sanctions associated with the two organisational structures. Accordingly, delegation can only become strictly optimal if the reduction in Agent B's wage is sufficient to offset the increase in Agent A's wage.

Figure 4 shows that this can indeed be the case, by illustrating a specific numerical example.[18] The graphic shows the optimal organisational design for the principal as Δ_A and Δ_B vary and is split into two distinct regions. In the top-right of the graph, where both Δ_A and Δ_B are relatively high, the principal's wage costs are minimised by implementing centralisation; in the bottom-left region, where these values are low, delegation becomes optimal. This pattern is consistent with Propositions 1 and 4, since Agent i's wage is increasing in Δ_i under delegation with the alternative punishment strategy, but invariant to such changes in the case of centralisation. The intuition follows directly from the foregoing discussion, since higher values of Δ_i are associated with a less harsh punishment under delegation.

We now derive some results for the general case. Let us define $\alpha_i = w_i^{del} - w^{cent}$ as the difference in wages

[18]For the example studied in Figure 4, the parameters chosen are as follows: $e = 2$, $r = \frac{1}{2}$, $\delta = \frac{3}{4}$, $p_2 = \frac{1}{2}$, $p_1 = \frac{1}{4}$, $p_0 = \frac{3}{16}$; we let $q_l = \kappa p_l$ for $l \in \{0, 1, 2\}$, with $\kappa = \frac{1}{2}$.

for Agent i under the two organisational structures. When α_i is positive, wages are higher under delegation compared to centralisation; the reverse is true for α_i negative. In order for delegation to become optimal, we therefore require that $\alpha_A + \alpha_B \leq 0$.

Proposition 5. *Let $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$. Then:*

(i) α_A is positive, while α_B is positive iff. Δ_B is sufficiently large.

(ii) α_A and α_B are both strictly decreasing in e.

(iii) α_A is strictly decreasing in δ iff. Δ_A is sufficiently large. α_B is strictly decreasing in δ iff. Δ_B is sufficiently small.

(iv) α_A is strictly decreasing in r. α_B is strictly decreasing in r iff. Δ_B is sufficiently large.

(v) In the special case where $q_l = \kappa p_l$ for $l \in \{0, 1, 2\}$, α_A and α_B are both strictly increasing in κ.

The first part of Proposition 5 restates formally the foregoing result, that Agent A's wage is always higher under delegation compared to the case of centralisation, while Agent B's wage can be either higher or lower. The remainder of the proposition outlines how the wage difference depends on the other parameters of the model.

Part *(ii)* implies that $\alpha_A + \alpha_B$ decreases as effort costs e grow. Intuitively, increases in effort costs lead to higher wages, but do not affect the agents' costs of adopting a particular production technology. Since the punishment strategy under delegation involves a strictly lower probability of receiving the bonus wage compared to centralisation, such increases then imply that (25) is reduced, so that the punishment under delegation becomes relatively harsher for both agents.

For part *(iii)*, the key effect of a higher δ is to increase the weight that the agents put on the future; accordingly, the prospect of repeated punishment following shirking plays a greater role in their current-period decision making, allowing for lower wages under either organisational structure as shown by Propositions 1 and 4. The relative size of the decreases in wages therefore naturally depends on the relative sizes of punishment utilities. For Agent B, we know that delegation is associated with a harsher punishment strategy as Δ_B decreases. One can then show that when Δ_B is sufficiently small, α_B becomes decreasing in δ.

For Agent A, we know that the punishment utility cannot be lower under delegation, since $\Delta_A > \widehat{\Delta}_A$. As long as $\Delta_A \leq \Delta_A^{crit}$, it can be shown that α_A is then always increasing in δ. However, there is an additional effect when $\Delta_A > \Delta_A^{crit}$, since in this case the wage must be designed to prevent a profitable deviation to the wrong production method; an increase in δ puts less weight on this deviation, introducing a countervailing effect. For large enough values of Δ_A, this effect becomes sufficiently strong for the direction of the derivative to switch.

Finally, note that by Propositions 1 and 4, the wage payments under delegation are sensitive to changes in the parameters r and κ, whereas wages under centralisation are invariant. Accordingly, variations in these parameters which lead to reductions in w_A^{del} and w_B^{del} make delegation more likely to be adopted, which is reflected by parts *(iv)* and *(v)* of Proposition 5.

Implications for the choice of organisational design. What are the implications of Proposition 5 for the principal's choice between centralisation and delegation? To address this question, we introduce an additional assumption, namely that the principal chooses to implement delegation when the wage costs are identical between the two organisational designs. In practice, there are several reasons that an organisation may prefer to grant a worker decision making authority when aligning the goals of the parties is not too expensive

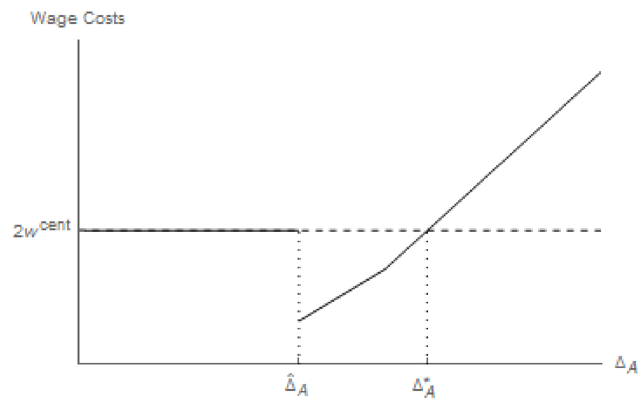

Figure 5: The wage costs associated with centralisation (dashed line) and delegation (solid line) as Δ_A varies.

(Aghion & Tirole 1997). For instance, while not captured by our model, a centralised organisational structure is likely to entail substantial transaction costs associated with the communication of information between different hierarchical levels. Moreover, the management literature emphasises that empowering a worker with decision making authority can be beneficial in itself, due to the resulting boost to employee morale and associated increase in motivation (see for instance Muir 1995).

This additional assumption implies, in particular, that the principal always chooses an organisational structure of delegation when we have $\Delta_A < \widehat{\Delta}_A$, since wage costs are identical to centralisation in this case. We then have the following corollary to Proposition 5.[19]

Corollary 1. *The principal is more likely to choose an organisational design of delegation as:*

1. *Δ_A or Δ_B decrease.*

2. *e increases.*

3. *κ decreases, in the special case where $q_l = \kappa p_l$ for $l \in \{0, 1, 2\}$.*

The corollary follows directly from the results of Proposition 5, along with the fact that both Δ_A^{max} and Δ_B^{max} are increasing in e and decreasing in κ. Part *(i)* is clearly illustrated by Figure 4, since any reduction in Δ_A or Δ_B move us closer to the grey region in which delegation is optimal.[20] The effect of an increase in e, or a decrease in κ, is to expand this grey area, such that delegation is chosen by the principal for a larger set of (Δ_A, Δ_B) combinations.

To complement the foregoing discussion, Figure 5 illustrates how the principal's wage costs change in Δ_A between 0 and Δ_A^{max}. Since the total wages paid to the agents under centralisation, $2w^{cent}$, are invariant to changes in Δ_A and Δ_B, the principal's wage costs are constant and shown by the horizontal dashed line. In Section 4, we showed that for $\Delta_A < \widehat{\Delta}_A$, the costs associated with delegation (illustrated here by the black line) are unchanged from the case of centralisation. However, at the point $\Delta_A = \widehat{\Delta}_A$ there is a discontinuity, as the alternative punishment equilibrium now becomes feasible. As long as Δ_B is sufficiently low, this leads

[19] Since the derivatives of α_A and α_B with respect to δ can have different signs, one can show that it is not possible to make a general statement on the implications of a change in δ for $\alpha_A + \alpha_B$. A similar comment applies for a change in r. In both cases, the sign will vary depending on the values of Δ_A and Δ_B.

[20] Since we have $\Delta_A \geq 0$ by assumption, a lower Δ_A corresponds to the leader having weaker preferences over which production method is to be implemented. This is consistent with Lazear's (2012) empirical finding that those in leadership positions tend to be 'generalists', who are able to employ a wide range of different skills in response to various situations which they may encounter.

to a drop in the wage costs associated with delegation, so that this organisational design becomes strictly optimal. Thereafter, wage costs are increasing in Δ_A, until eventually w_A^{del} becomes sufficiently high such that two curves intersect once more.[21]

Let us denote this intersection point by Δ_A^*. Given the foregoing assumption, the principal prefers to implement delegation for all $\Delta_A \leq \Delta_A^*$, and centralisation for $\Delta_A > \Delta_A^*$. The key effect of a decrease in Δ_B, an increase in e, or a decrease in κ, is to cause a larger drop in the costs associated with delegation to the right of $\widehat{\Delta}_A$; it then follows that Δ_A^* increases, so that delegation is strictly optimal for a larger range of values of Δ_A, holding Δ_B fixed.

Differences in wages across agents. In our framework, the agents receive identical wages under centralisation, and under delegation in the case where $\Delta_A < \widehat{\Delta}_A$. However, for the case where $\Delta_A > \widehat{\Delta}_A$, implementing an organisational structure of delegation requires paying asymmetric wages in the case of project success. To conclude this section, we briefly analyse this wage difference; we then use our results in the subsequent section to discuss which agent should be selected as the team's leader.

First, it is straightforward to verify that if delegation is to be optimal as an organisational structure when $\Delta_A > \widehat{\Delta}_A$, then Agent A must receive a strictly higher wage than Agent B. This follows directly from the foregoing discussion: since we have $\alpha_A > 0$, in order for $\alpha_A + \alpha_B$ to be negative it must be that $\alpha_B < 0$, which in turn implies that $w_A^{del} > w_B^{del}$. Second, Proposition 6 outlines the differences in the marginal impacts of changing the model's parameters on the wage payments to each agent.

Proposition 6. *Let $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$. If delegation is the optimal organisational structure, then the wages w_A^{del} and w_B^{del} have the following properties:*

(i) $\frac{\partial w_A^{del}}{\partial e} \leq \frac{\partial w_B^{del}}{\partial e}$.

(ii) $\frac{\partial w_A^{del}}{\partial \Delta_A} \geq \frac{\partial w_B^{del}}{\partial \Delta_B}$.

(iii) $\frac{\partial w_A^{del}}{\partial r} < \frac{\partial w_B^{del}}{\partial r}$

Parts *(i)* and *(ii)* of Proposition 6 compare the impacts of increases in e and Δ_i on the agents' wages. When $\Delta_A < \Delta_A^{crit}$, w_A^{del} and w_B^{del} have the same structure (i.e. they are identical up to Δ_A and Δ_B) and thus the impact of changes in these parameters is identical across agents. However, this is not the case when $\Delta_A > \Delta_A^{crit}$.

First, with respect to part *(i)*, Agent A's wage can be shown to increase less than Agent B's in response to higher wage costs e. Intuitively, when $\Delta_A > \Delta_A^{crit}$, (13) binds so that Agent A's most profitable deviation in state ω_2 is to shirk *and* implement the incorrect method of production; this reduces the probability of project success to larger extent than Agent B's optimal deviation, which is just to shirk. Accordingly, since Agent A's action can influence the production process to a greater extent, a smaller increase in wages is required in response to higher effort costs.[22]

Second, for part *(ii)*, an increase in Δ_A has a stronger impact on Agent A's wage than an increase in Δ_B does on Agent B's payment. To see why, note that a higher value of Δ_A corresponds to stronger preferences

[21] Clearly, the non-monotonicity of the cost function associated with a delegated organisational structure implies that the principal may benefit from a higher value of Δ_A. Since this parameter represents Agent A's preferences over the two production methods, it is feasible that it could be influenced (at least to some extent) by the principal's choice of various aspects of job design, and thus become endogenous. For instance, a firm may be able to artificially increase the costs associated with moving away from a 'default' method of production by imposing additional requirements upon workers such as paperwork.

[22] This logic is perhaps best illustrated by considering the extreme case in which the influence of an agent's action on the probability of project success tends to zero; in this case, the required adjustment in wages following an increase in effort costs becomes infinite.

for implementing production method γ_1. This implies that Agent A's wage needs to be increased in order to prevent a deviation in which he implements γ_1 in state ω_2; clearly, this effect is absent for Agent B, who does not control the production method.

Finally, part *(iii)* follows immediately from the findings of Proposition 4. When delegation is optimal, it must be the case that Δ_B is relatively small and thus $\frac{\partial w_B^{del}}{\partial r} > 0$, whereas $\frac{\partial w_A^{del}}{\partial r}$ is always negative.

6 Applications to Organisational Design

Much of the literature which studies delegation within organisations focuses on the trade-off inherent in allocating decision rights to a worker who has superior information, but may also have conflicting objectives (see for instance Aghion & Tirole 1997). By assuming that the state of the world is common knowledge to all parties, our analysis abstracts from this aspect of delegation (i.e. the utilisation of decentralised knowledge) in order to isolate the implications for the generation of implicit incentives between team members. This yields a number of novel insights regarding various facets of organisational design, such as the composition of teams, the role of leadership and the effects of transparency between different hierarchical levels of the firm. These complement existing findings, which analyse disparate channels through which a firm's ability to delegate can influence organisational design.

Team Composition & Size. The complexity of modern business environments has increasingly lead to the adoption of multidisciplinary teams, which are made up of workers with diverse backgrounds, expertise and knowledge (Jackson 1996). For example, a successful product design may require the input of many different specialists, from engineers who provide technical expertise, to psychologists who can help construct a practical user interface, to advertising professionals who can advise on the marketability of the product. A question of increasing practical relevance for organisations therefore relates to the various channels through which team diversity might affect overall team performance (Van Knippenberg & Mell 2016).

A number of recent papers have used economic methods to study the implications of diversity for repeated interaction in teams. Glover & Kim (2021) impose an assumption that specialised teams, in which the efforts of team members are substitutes, are inherently most efficient due to productive synergies between workers. Nonetheless, they show that diverse teams, in which efforts are compliments, can become optimal due to the implicit incentives provided by mutual monitoring. Intuitively, this is because complimentary efforts guarantee the existence of a particularly stringent punishment equilibrium in which both agents shirk; in contrast, when efforts are substitutes, agents may still choose to work when their teammate shirks under the optimal wage scheme, undermining the incentives provided by the agents' mutual monitoring. In a related paper, Glover & Kim (2020) show that diversity in the career horizons of agents (captured by variations in the discount factor) can be advantageous for stifling the possibility of collusion between team members.[23]

Our findings compliment these results, similarly showing that team diversity can be useful for the creation of implicit incentives in a decentralised organisational structure, though the mechanism that we consider is different. In our paper, heterogeneity between agents stems from their preferences over alternative production methods, which are absent in the aforementioned papers. We show that disparity in these preferences (i.e. $\Delta_B < 0$) is particularly valuable to the principal due to the tension between the agents' interests. Intuitively, following a breakdown in team cohesion each agent acts in their own self-interest and maximises current-period utility. For Agent A, this involves selecting his favoured production method in each period; if the

[23] See also Flassak & Hofmann (2020), who study the implications of team diversity for the interaction between explicit incentives provided by a principal and the implicit incentives arising from an agent's career concerns.

workers have similar preferences over these methods of production, then this selection is likely to also benefit Agent B, so that punishment following a deviation from the agents' implicit agreement is not particularly harsh. In contrast, when the agents have diverse preferences, Agent A's choice is harmful to Agent B, increasing implicit incentives.[24]

Our model also yields insight regarding the optimal size of teams. In our framework with two agents, an organisational design of delegation can become optimal by reducing the wage payments necessary to motivate the non-leader. In larger teams, there may be many such workers who have no control over the choice of production technology; by the same logic, implementing a decentralised organisational structure may then allow for the reduction of the wages of several agents. The drawback of an organisational design of delegation is the increased wage which must be paid to the team-leader. However, since the team only requires one leader, and since the leader's role is essentially unchanged by the presence of additional agents, this 'cost' associated with delegation does not change as the size of the team grows. Altogether, this suggests that the potential benefit to delegating decision rights to the team leader is greater in larger teams. In practice, however, one would expect that mutual monitoring between agents becomes increasingly difficult as team size grows, placing a limit on the extent to which this effect allows firms to motivate workers.

Leadership in Teams. Leaders in work teams typically play a large number of roles: they make decisions regarding mission and goals, formulate and structure plans, communicate information to subordinates, provide motivation and may even assist with the training and development of other workers within the team (Morgeson et al. 2010, Lazear 2012). Several of these roles have been studied in the economics literature. For instance, Hermalin (1998, 2007) analyses the extent to which a leader can credibly communicate superior information to others within the team, in the static and repeated settings respectively, while Huck & Rey-Biel (2006) show that leadership is valuable (and can emerge endogenously) when team members are conformists. Other studies include Kvaløy & Schöttner (2015), who analyse the role of motivators in the context of team production and the ensuing implications for incentive contracting and Hermalin (2017), who investigates the role of charismatic leadership in communicating information to followers.

Our framework highlights a particular role of leaders within teams, specifically with respect to their authority to decide on how the production process is organised. We show that, importantly, this authority also determines the range of feasible outcomes which can be realised following a breakdown in team cohesion; accordingly, our analysis suggests that through this channel leadership can play a significant role in determining the extent to which mutual monitoring within a team is able to create implicit incentives.

This is particularly obvious in our framework, since the team is only made up of two agents. However, suppose there exists a larger team and consider the informal relationship between Agents i and j, neither of whom are the team leader. Agent i understands that his shirking today might result in Agent j shirking tomorrow, and takes this into account when deciding on his effort provision. But the implications of Agent j's shirking, via its impact on expected project success, are influenced by the leader's choice of production method. In this way, the leader's increased authority also affects the informal relationships between other members of the team, which in turn influences the magnitude of the implicit incentives which are created.

Who should be the leader? Our findings also yield predictions regarding a firm's decision of who to appoint as a team leader when delegation is chosen as the organisational structure. Part *(i)* of Proposition 6 implies that if the agents have asymmetric effort costs, the principal prefers the leader's costs to be higher

[24] Hamilton et al. (2003) present empirical evidence that team heterogeneity in ability is positively correlated with productivity. As a potential explanation for their results, they argue that team heterogeneity has important implications for the informal agreements between team members, similar to the foregoing discussion.

than the subordinates, rather than the other way around. There are several possible interpretations for this result. One is that team leaders should play a significant role in the production progress, or 'lead from the front', when additionally granted authority to undertake decisions on behalf of the team. Many authors in the management literature have discussed the various benefits which stem from an individual *leading* a team by taking an active role, rather than merely *managing* it; see for instance the discussions in Myers et al. (1995), Levasseur (2005) and Morgeson et al. (2010).[25] Our analysis highlights an additional benefit, due to the implications for the implicit incentives between team members.

Alternatively, the result could be interpreted as suggesting that the leader should not necessarily be the most skilled member of the team, contrary to what would seem to be conventional wisdom. Indeed, Hidir & Migrow (2021) highlight a number of empirical studies which document the allocation of decision rights within an organisation to workers who are poorly qualified, or relatively less suited than others to a particular role. They then develop a model which provides a potential explanation for this phenomenon, based on the acquisition of information. Along similar lines, Komai & Stegeman (2010) consider a model with uncertainty regarding project quality in the presence of moral hazard and show that the optimal leader has 'average or unusually high costs of effort'. Our framework generates similar results, due to the complementarity between motivating the correct choice of production method and providing high effort incentives.

Delegation of Decision Rights. Beyond highlighting a potential benefit of allocating decision rights to lower-level workers — the implications for the creation of implicit incentives — our analysis also yields predictions regarding those decisions which we might expect to see frequently delegated. Part *(ii)* of Proposition 5 states that both α_A and α_B are decreasing in e, so that delegation is more likely to be implemented when effort costs are high. In these cases, the private costs and benefits to the agents of a particular production method being implemented (captured by differences in their adoption costs) play a relatively small role in determining their overall welfare. Instead, their utility is much more dependent on whether they choose to work hard and incur large effort costs, and, since these large effort costs are associated with high wages, on whether the project is successful.

Part *(v)* of the same proposition says that, in the case where $q_l = \kappa p_l$ for $l \in \{0, 1, 2\}$, both α_A and α_B are increasing in κ. This implies that delegation is more likely when, following adoption of the incorrect production method, the probability of project success is low. When κ is low, the choice of production method is a large determinant of project success. Since all parties have a common concern for the project being successful, a lower κ works to align the interests of the principal and the agents. From the team leader's point of view, it then becomes more important to choose the correct production method, rather than making the decision which is most favourable for his adoption costs. Our framework therefore predicts that it is those decisions which are crucial for project success, but relatively inconsequential for the agents' private costs and benefits, which we should expect to be allocated to the workers.

Although the mechanism is of course different, these predictions complements similar findings on the allocation of decision rights within organisations by other authors. For instance, Aghion & Tirole (1997) argue that delegation is more likely for decisions which matter little to the principal, or those for which the agent can be trusted due to a high degree of congruency in the preferences of two parties (see Section IV of their paper).[26]

[25] Many papers in the economics literature which study leadership similarly emphasise the importance of the team leader's actions; for instance, Hermalin (1998) studies *leading by example*, whereby an informed leader can undertake high effort in order to credibly communicate information to other members of the team.

[26] In our framework, since all parties benefit from a successful project, one can think of 'congruent preferences' between the principal and team leader as corresponding to situations in which the implications of the decision for the latter's adoption costs are relatively small compared to its implications for project success.

Transparency and the Boundary of the Firm. As discussed in the introduction to this section, several authors have analysed situations in which decision rights are allocated to an informed worker whose preferences are misaligned with those of the firm. In this case, common wisdom suggests that it would be beneficial for the firm to closely monitor the worker, or gather information regarding whether the appropriate decision has been made in a given instance, in order to minimise the scope for misbehaviour. Our results suggest that in the presence of teamwork, there may in fact be benefits to a lack of transparency.

Intuitively, as in Che & Yoo's (2001) model of team production, in our framework the implicit incentives generated by the agents' mutual monitoring stem from the threat of repeated punishment as a response to deviations from the principal's desired play. Clearly, this can only successfully create incentives if both agents believe that this threat will actually be realised. This depends, in particular, on a lack of interference from the principal on the punishment path.

Even when firms cannot credibly commit to not interfere, such an assumption may still be reasonable in Che & Yoo's (2001) model, since in many environments firms likely find it difficult to detect even repeated shirking by workers. However, the assumption is much more questionable if punishments additionally involve changes in the team's production methods, since these are much more likely to be (at least to some degree) observable.

It follows that in order to benefit from delegation and its impact on the agents' mutual monitoring, firms may wish to maintain a certain level of opaqueness, whereby those at a higher hierarchical level within the firm (such as upper-management) have limited information regarding the decision making of lower-level workers. In doing so, they increase the independence of the team and can thus help sustain the agents' beliefs that any sanctions implicitly threatened following a breakdown in team cohesion will indeed become realised.[27]

Taking this argument to an extreme, one way in which a firm could simultaneously guarantee opaqueness and credibly commit to not interfere in the production process would be to outsource the project; in this respect, our findings also have implications for the manner in which organisational designs interact with the boundary of the firm.

Self-Organised Teamwork. Our formal model assumes that under delegation the team is made up of two members: a team leader, who unilaterally decides on the production method to be implemented, and a subordinate who has no input into the decision. While this may be broadly representative of many real world teams, in which one member alone has a significant amount of authority over the production process, in recent years organisations have been increasingly utilising self-organised teams (SOTs). As discussed in the introduction, SOTs are typically made up of diverse workers, who have the collective autonomy to plan, manage and execute tasks interdependently. Authority in SOTs is generally not held by one individual, as in our framework, but is instead distributed among members in a shared leadership model whereby workers have a collective responsibility for team performance.

Nonetheless, our findings still yield insight into the role of implicit incentives within SOTs. Typically, each member of the team will be endowed with authority, either individually or collectively, over various specific aspects of the production process. To maintain team cohesion, workers are expected to make decisions in a way which promotes achievement of the team's goals, rather than in their own self-interest. These could include offering help, motivation or training to others, sharing production-relevant information or taking time to provide other team members with valuable feedback on specific tasks.

Following a breakdown in team cohesion, workers in SOTs may be inclined to instead undertake these

[27] Baker et al. (1999) find that a principal can benefit from a lack of transparency when decision rights can only be informally delegated to an agent, due to the implications for the principal's temptation to renege on the informal agreement.

decisions in a more selfish manner, which in many environments will impose significant costs upon other team members (e.g. the withdrawal of help or important information). In this way, the team's increased authority over the production process generates a positive effect on the implicit incentives associated with the informal relationships between team members, as in our model, despite the fact that authority is dispersed between many workers.

7 Collusion

So far, our formal analysis has focused on the questions of whether the principal's desired outcome can be implemented as a subgame-perfect equilibrium under centralisation and delegation, and, if so, which of these organisational structures allows the principal to utilise a wage structure which minimises her costs. However, we have not considered whether these equilibria are optimal from the agents' point of view. It is possible that for a particular organisational design and wage structure, there exists an alternative pair of strategies for the agents, that form a subgame-perfect equilibrium which allows for a Pareto improvement of the agents' expected payoffs from the game. In this section, we study whether such collusion is possible for the agents, considering each organisational structure in turn.[28]

Under a centralised organisational structure, the agents cannot control the choice of production method and thus can only collude with respect to their effort provision. As shown by Che & Yoo (2001), when this is the case, a wage structure of $\{w^{cent}, w^{cent}\}$ implies that the agents' surplus is maximised in each period when they both undertake high effort.[29] It then follows that no other set of strategies, regardless of whether they form an equilibrium, are able to improve upon repeated play of $(work, work)$ for the agents. Intuitively, implicit incentives are created in the repeated game by the threat of shirking in future periods. For this threat to be effective, it must be the case that the agents are better off — and thus their surplus is higher — when they both work compared to the case where they both shirk. Moreover, the supermodularity of the production function implies that surplus cannot be increased by incorporating outcomes in which only one agent works.

Next, we study collusion when the choice over production methods is delegated to Agent A. Since for $\Delta_A < \widehat{\Delta}_A$ delegation is associated with the same wage costs as centralisation, we shall instead focus on the case of $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ with $\Delta_B < \Delta_B^{max}$ and in particular consider the following question: can the principal continue to achieve strictly lower wage costs through delegation when the equilibrium is additionally required to be collusion proof?

Studying potential collusive outcomes under an organisational design of delegation is significantly more complicated than the case of centralisation for a number of reasons. First and foremost, Agent A's control over the production method implies a wider range of potential outcomes which could be implemented as part of a collusive agreement. For instance, the two agents may agree to always adopt a particular production method, regardless of the state, if they have particularly strong preferences for such an outcome, as well as coordinating on whether to provide high or low effort. Second, the fact that the production method is no longer decided by the principal also implies that collusive strategies will likely become conditional

[28] To be clear, our notion of collusion-proofness requires that there exists no other subgame-perfect equilibrium which is Pareto dominant for the agents, in the sense that it offers a weakly better expected payoff to both and a strictly better payoff to at least one. Other notions of collusion-proofness have been considered in the literature; for instance, Che & Yoo's (2001) concept of 'team equilibrium' requires that an outcome is the subgame-perfect equilibrium which yields the highest total surplus to the agents, while Glover & Kim (2020) primarily study regret-free collusion. As discussed in the model setup, we rule out the possibility of explicit side-contracting between agents à la Itoh (1992, 1993), which can be thought of as an alternative form of collusion.

[29] We omit a formal proof, instead referring the reader to Proposition 4 and the related discussion in their paper.

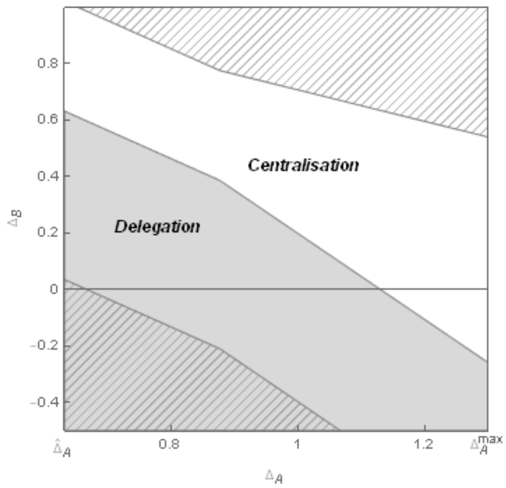

Figure 6: The possibility for collusion.

on the state of the world. Third, from a technical perspective, the asymmetric wage scheme associated with delegation further complicates the problem, since the agents may have differing preferences over the relative attractiveness of working and shirking given a particular production method and state of the world. Nonetheless, we are able to show the following result.

Proposition 7. *There always exists a non-empty set of* (Δ_A, Δ_B) *combinations, where* $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ *and* $\Delta_B < \Delta_B^{max}$*, such that (i) under delegation, the wage scheme* $\{w_A^{del}, w_B^{del}\}$ *implements the principal's desired outcome at a strictly lower cost than centralisation and (ii) the resulting equilibrium is collusion-proof.*

The proof of Proposition 7 shows that there always exists a set of (Δ_A, Δ_B) combinations such that the principal's desired outcome also maximises the surplus of the agents, similar to the case of centralisation. Specifically, we consider the agents' surplus from the principal's desired outcome as a function of $\Delta_{sum} = \Delta_A + \Delta_B$, which can be thought of as capturing their combined preferences over the production methods. We show that there always exists a non-empty interval such that, for all values of Δ_{sum} within this interval, any deviation away from the principal's desired outcome leads to a reduction in the agents' joint expected payoffs, so that they have no incentives to deviate from the equilibrium under delegation. However, we also show that this equilibrium may fail to be collusion proof when Δ_{sum} is either too large, or too small.

Figure 6 reproduces the example studied Section 5 (Figure 4), but now also displaying the implications of changes in Δ_A and Δ_B for the agents' surplus and the possibility of collusion.[30] As before, the white and grey areas respectively show the range of values over which centralisation and delegation are optimal for the principal. The two new shaded areas show the regions where, under an organisational design of delegation, the principal's desired outcome does not maximise the agents' surplus.

First, when Δ_{sum} is very large, the agents have strong (aggregate) preferences for adopting production method γ_1. In this case, agents may be able to increase their joint surplus by agreeing to implement production method γ_1 in both states of the world; in Figure 6, this possibility is illustrated by the shaded region in the top-right of the graphic. However, recall from Proposition 4 that the wage w_i^{del} is increasing in Δ_i for $i = \{A, B\}$; accordingly, loosely speaking, we can say that the total wage payment under delegation is increasing in Δ_{sum}. Moreover, the proof of Proposition 7 shows that in cases where Δ_{sum} is sufficiently large

[30]For the parameters associated with this specific example, see Footnote 18.

such that the aforementioned collusion is a concern, it is always true that the principal faces lower costs under centralisation than delegation. Put differently, as long as delegation is the optimal organisational design for the principal, Δ_{sum} can never become sufficiently high such that collusion is a concern. Corresponding to this result, in Figure 6 the shaded region in the top-right does not intersect with the grey area in which delegation is optimal.

Second, there is a shaded area in the bottom-left of the graphic where Δ_{sum} becomes very low. This is problematic for the principal since, for the aforementioned reason, total wages under delegation will typically be lower than those under centralisation and thus the principal would prefer to adopt an organisational design of delegation. Collusion in this region may occur for two distinct reasons. This first reason is that, as discussed in the foregoing the total wages paid to the agents are decreasing as Δ_{sum} becomes lower. It then follows that for sufficiently small Δ_{sum}, the agents' wages can become so low that working is no longer worthwhile, in which case their joint surplus is maximised when both agents shirk. Moreover, if Δ_A and Δ_B are so low that Δ_{sum} is sufficiently negative, the agents have strong (aggregate) preferences for adopting γ_2 in both states of the world, which may also allow them to increase their surplus by deviating from the principal's desired outcome.

Proposition 7 therefore establishes that there exists the *potential* for collusion; i.e. that there are alternative outcomes which *could* increase the agents' joint surplus. Collusion only becomes feasible if, in addition, these outcomes can be sustained as a sub-game perfect equilibrium, given that the principal selects an organisational structure of delegation and offers the wage scheme $\{w_A^{del}, w_B^{del}\}$. However, the large literature which analyses attainable outcomes in repeated games (in particular by providing Folk Theorems; see for instance Fudenberg & Maskin 1986) emphasises that the set of equilibria will typically be large, so long as the players are sufficiently patient.[31] Accordingly, we can expect there to be many cases in which there exist alternative equilibria which yield a higher joint surplus to the agents than the principal's desired outcome.

For the reasons discussed earlier in this section, characterising the agents' optimal collusive equilibrium for each set of parameters is complex, and is beyond the scope of the paper. Instead, in Appendix II, we present an example of a collusive deviation from the principal's desired outcome under delegation. Specifically, we analyse a situation in which, due to Δ_B being small, Agent B strictly prefers an outcome in which both players shirk in each period to one where both players work; Agent A, on the other hand, prefers high effort from both players. Despite this conflict of interest, they are able to agree to an equilibrium outcome in which Agent A always shirks, while Agent B shirks with probability $\frac{1}{4}$.

We show that this equilibrium is strictly preferable for both players to the principal's desired outcome. For Agent A, shirking in each period leads to a significant decrease in effort costs. Moreover, Agent B exerts high effort with probability $\frac{3}{4}$, which cushions the impact of receiving the wage with a lower probability. For Agent B, shirking $\frac{1}{4}$ of the time also leads to a small decrease in effort costs. While he receives the bonus much less often, since Agent A always shirks, this is not so much of a concern due to the low wages associated with a small value of Δ_B. The example shows that both agents therefore become strictly better off, and that, in addition, neither has an incentive to deviate from their strategies, so that this play forms a subgame-perfect equilibrium.

[31] There is however a small, but important, difference to these setups, since in the current framework the agent's payoffs are sensitive to changes in the discount factor δ.

8 Conclusion

This paper aims to analyse the implications of additional decision making authority for the implicit incentives which are generated in long-lived teams. For that purpose, we consider a repeated game model of team production under moral hazard and compare the principal's costs of providing incentives under organisational designs of centralisation and delegation. We show that while the allocation of decision rights to a team member places additional constraints on the principal's wage scheme, it can nonetheless be strictly optimal due to the implications for the informal relationship between the agents and the ensuing changes to the implicit incentives generated by their repeated interaction. We also provide a comparative static analysis, showing how the wage scheme offered to the agents and the principal's ensuing choice of organisational structure vary with the parameters of the model. We then use these results to discuss various aspects of the design of organisations, such as optimal team composition, transparency and the role played by team leaders. In addition, our findings also provide numerous empirical predictions regarding the utilisation and design of teams with decision making authority. To conclude, we briefly discuss the extent to which the paper's results are consistent with alternative specifications of the model.

Individual Performance Measures. Our framework assumes that the principal has access to a single signal, the outcome of the project, in order to align the incentives of both agents. In practice, however, there are many situations in which firms may be able to instead measure the individual performance of workers within teams. Nonetheless, Che & Yoo (2001) show that firms often wish to use joint performance evaluation, in which the individual signals are aggregated into a team performance measure, in order to provide incentives for mutual monitoring and create implicit incentives. Since delegation can also create additional incentives via the workers' informal relationship, one would expect this result to carry through into our environment. However, this would likely depend on the exact assumptions regarding the implications of the selection of production method for the performance measures of the two workers.

Sequentiality of the Stage Game. Under an organisational structure of delegation, we assume that the production method and the agents' effort choices are selected simultaneously. This is clearly at odds with many of the real world examples discussed in the paper, such as project selection or task allocation, in which the method of production is fixed before workers choose how much effort to exert. However, note that allowing for a sequential structure has no implications when the principal selects a centralised organisational structure. Moreover, a sequential stage game would actually weakly reduce the costs associated with the wage scheme under delegation. Intuitively, in a sequential structure, should Agent A at any point choose to implement the incorrect production method, Agent B would instantaneously observe this deviation from the agents' informal agreement and immediately begin to shirk — i.e. in the very same period. In contrast, in the current framework with a simultaneous stage game, Agent B only begins to punish in the following period, so that Agent A's incentives to deviate are higher in the simultaneous framework. Altogether, this suggests that the principal would find it weakly more beneficial to delegate under a sequential structure, since the costs associated with preventing Agent A's deviation to an incorrect production method are lower when the stage game is sequential.

Verifiability of the Production Method. Throughout, we have assumed that the chosen production method cannot be verified and hence cannot be contracted upon. However, one can imagine many situations in which the production method which is employed is ex post verifiable. Under centralisation, verifiability of the production method makes no difference; this is because, in either state of the world, the agents must be motivated only to undertake high effort conditional on the correct production method being chosen. Since the incentive compatibility constraint (9) is the same in states ω_1 and ω_2 (and hence under production methods γ_1 and γ_2 in these states, respectively), this implies that the principal cannot benefit from this verifiability. In contrast, under a delegated organisational structure, the principal could indeed benefit by conditioning wage payments on the chosen production method. However, since she could also choose to ignore this information, it follows that verifiability of the production method can only lead to a weak decrease in the principal's costs of utilising delegation, so that the principal is more likely to utilise a decentralised organisational structure.

Appendix I

Proof of Proposition 1. Parts *(i)* and *(iii)* are clear. For part *(ii)*, differentiating w^{cent} from (9) yields:

$$\frac{\partial w^{cent}}{\partial \delta} = \frac{-e\,(p_1 - p_0)}{\left[p_2 - (1-\delta)p_1 - \delta p_0\right]^2} < 0 \tag{26}$$

□

Proof of Proposition 2. First, we show that given $\Delta_A < \widehat{\Delta}_A$ and $w_A = w_B = w^{cent}$, (γ_1 *and shirk*; *shirk*) is the unique Nash Equilibrium of the stage game in state ω_1, Γ_1^{del}, and (γ_2 *and shirk*; *shirk*) is the unique Nash Equilibrium of the stage game in state ω_2, Γ_2^{del}. To see this, note that Assumption 1 implies:

$$w^{cent} < \frac{e}{p_2 - p_1} < \frac{e}{q_2 - q_1} \tag{27}$$

and

$$w^{cent} < \frac{e}{p_2 - p_1} < \frac{e}{p_1 - p_0} < \frac{e}{q_1 - q_0} \tag{28}$$

Accordingly, *work* is a strictly dominated strategy for Agent B in both games. In state ω_1, these inequalities also imply that Agent A's unique best response to *shirk* is γ_1 *and shirk*, since $\Delta_A \geq 0$ by assumption. In state ω_2, γ_2 *and shirk* is Agent A's unique best response *iff.* we have:

$$p_0 w^{cent} - c_A^2 > q_0 w^{cent} - c_A^1 \tag{29}$$

$$\iff w^{cent} > \frac{\Delta_A}{p_0 - q_0} \tag{30}$$

$$\iff \widehat{\Delta}_A := \frac{e\,(p_0 - q_0)}{p_2 - (1-\delta)p_1 - \delta p_0} > \Delta_A \tag{31}$$

which is satisfied.

Next, we study the incentive compatibility constraints, using the arguments from the main text. Since the wage and the punishment equilibrium are both identical to the case of centralisation, it is clear that (4) and (5) are satisfied for both agents. For Agent B, these are the only necessary restrictions. For Agent A, the wage w_A must additionally satisfy (10)-(13). We show that none of these additional constraints bind. Since $c_A^2 \geq c_A^1$ by assumption, it is easy to verify that (12) and (13) imply (10) and (11), respectively. Moreover, since $w^{cent} < \frac{e}{q_2 - q_1}$, (13) implies (12). Finally, rearranging (13) and inserting (8) yields:

$$w_A \geq \frac{e + (1-\delta)\Delta_A}{p_2 - (1-\delta)q_1 - \delta p_0} \tag{32}$$

Moreover, we have:

$$w^{cent} \geq \frac{e + (1-\delta)\Delta_A}{p_2 - (1-\delta)q_1 - \delta p_0} \tag{33}$$

$$\iff \frac{e\left[p_1 - q_1\right]}{p_2 - (1-\delta)p_1 - \delta p_0} \geq \Delta_A \tag{34}$$

which is immediately satisfied when $\Delta_A < \widehat{\Delta}_A$ due to Assumption 1. Accordingly, all necessary constraints are satisfied, so that setting $w_A = w_B = w^{cent}$ can sustain the principal's desired outcome as a subgame-perfect equilibrium.

We have shown that $w_A = w_B = w^{cent}$ is the wage scheme which satisfies the foregoing IC constraints at the lowest possible cost to the principal for the given punishment equilibrium. Nonetheless, the principal could in theory design a wage scheme $\{w_A, w_B\}$ which induces an alternative punishment equilibrium, thereby modifying these IC constraints. From the discussions in the main text, the only feasible alternative punishment equilibria are those in which Agent A implements production method γ_1 in state ω_2. However, this requires that γ_1 *and shirk* is Agent A's best response to Agent B's shirking in the stage game Γ_2^{del}; i.e.:

$$w_A \leq \frac{\Delta_A}{p_0 - q_0} \tag{35}$$

Moreover, in the proof of Proposition 3 we shall show that conditional on this alternative punishment equilibrium, the wage payment to Agent A must satisfy the following constraint in order to induce high effort:

$$w_A \geq \frac{e + \delta\,(1-r)\,\Delta_A}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} \tag{36}$$

Clearly, it is only feasible for w_A to satisfy both constraints if:

$$\frac{\Delta_A}{p_0 - q_0} \geq \frac{e + \delta\,(1-r)\,\Delta_A}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} \tag{37}$$

$$\iff \Delta_A \geq \frac{e\,(p_0 - q_0)}{p_2 - (1-\delta)p_1 - \delta p_0} = \widehat{\Delta}_A \tag{38}$$

which violates $\Delta_A < \widehat{\Delta}_A$. Accordingly, the alternative punishment strategy is not implementable, so that setting $w_A = w_B = w^{cent}$ indeed minimises the principal's costs in this case. □

Proof of Proposition 3. First, we show that given $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$, when $w_A = w_A^{del}$ and $w_B = w_B^{del}$, $(\gamma_1 \textit{ and shirk}; \textit{shirk})$ is the unique Nash Equilibrium of the stage games in states ω_1 and ω_2, Γ_1^{del} and Γ_2^{del}. To begin, note that both arguments of the maximum function in (21) are increasing linearly in Δ_A. Setting them equal to one another and rearranging yields the unique intersection point (24). The numerator of this term is clearly positive. The denominator is strictly increasing in δ; at $\delta = 0$ it becomes equal to $p_2 - p_1$ which is strictly positive. Hence, we have $\Delta_A^{crit} > 0$.

Using the definition of w_A^{del} given by (23), it is straightforward to verify that at the point $\Delta_A = \Delta_A^{crit}$ we have $w_A^{del} < \frac{e}{p_2 - p_1}$. Accordingly, since w_A^{del} is strictly increasing in Δ_A, we have:

$$w_A^{del} < \frac{e}{p_2 - p_1}$$

$$\iff \Delta_A < \frac{e\left[p_1 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\,q_0\right]\right]}{(1-\delta r)\,(p_2 - p_1)} = \Delta_A^{max}$$

Similarly, one can verify that for all $\Delta_B < \Delta_B^{max}$, we have $w_B^{del} < \frac{e}{p_2 - p_1}$. Hence, by a similar logic to that of the proof of Proposition 2, *work* is a strictly dominated strategy for Agent B in both games. Moreover, in state ω_1, Agent A's unique best response to *shirk* is γ_1 *and shirk*, since $\Delta_A \geq 0$ by assumption. His unique best response in state ω_2 is also γ_1 *and shirk iff.* $w_A^{del} < \frac{\Delta_A}{p_0 - q_0}$. One can show that:

$$w_A^{del} < \frac{\Delta_A}{p_0 - q_0}$$

$$\iff \Delta_A > \frac{e\,(p_0 - q_0)}{p_2 - (1-\delta)p_1 - \delta p_0} = \widehat{\Delta}_A$$

Accordingly, for all $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$, when $w_A = w_A^{del}$ and $w_B = w_B^{del}$, (γ_1 *and shirk*; *shirk*) is the unique Nash Equilibrium in both states.

All that remains is to show that the necessary IC constraints are satisfied, given that (γ_1 *and shirk*; *shirk*) is played repeatedly in both states during the punishment phase of the game. (4) and (5) are identical; inserting (18) and rearranging yields:

$$w_i \geq \frac{e + \delta\,(1-r)\,\Delta_i}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]}$$

For Agent B, this is the only constraint which needs to be satisfied. For Agent A, we also require that (10)-(13) are satisfied. From the foregoing, we have $w_A^{del} < \frac{e}{p_2 - p_1} < \frac{e}{q_2 - q_1}$ for all $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$; hence, by the discussions in the proof of Proposition 2, it is sufficient for us to show that (13) holds. Inserting (18) into (13) and rearranging yields:

$$w_A \geq \frac{e + (1-\delta r)\,\Delta_A}{p_2 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\,q_0\right]}$$

so that all necessary constraints are satisfied. Altogether, this implies that the wages w_A^{del} and w_B^{del} are able to sustain the principal's desired outcome as a subgame-perfect equilibrium. Moreover, using the foregoing arguments, it is straightforward to verify that $0 < \widehat{\Delta}_A < \Delta_A^{crit} < \Delta_A^{max}$. □

Proof of Proposition 4. Parts *(i)* and *(ii)* are clear from (21) and (22). Substituting $q_l = \kappa p_l$ into these equations, part *(v)* is also clear, since an increase in κ leads to a smaller denominator. For part *(iii)*, from (23) we have:

$$\frac{\partial w_A}{\partial \delta} = \frac{(1-r)\,\Delta_A\,(p_2 - p_1) - e\left[p_1 - \left[rp_0 + (1-r)\,q_0\right]\right]}{\left[p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]\right]^2} \tag{39}$$

for all $\Delta_A \leq \Delta_A^{crit}$. One can show that (39) is strictly negative when evaluated at the point $\Delta_A = \Delta_A^{crit}$; since it is also increasing in Δ_A, it is then strictly negative for all $\Delta_A \leq \Delta_A^{crit}$. Next, for all $\Delta_A > \Delta_A^{crit}$, from (23) we have:

$$\frac{\partial w_A}{\partial \delta} = \frac{-\Delta_A \left[r\left(p_2 - p_0\right) + \left(q_1 - q_0\right)(1-r)\right] - e\left[q_1 - \left[rp_0 + (1-r)\, q_0\right]\right]}{\left[p_2 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\, q_0\right]\right]^2} \tag{40}$$

which is strictly decreasing in Δ_A. Evaluating at the point $\Delta_A = \Delta_A^{crit}$, (40) is strictly negative. Accordingly, (40) is also strictly negative for all $\Delta_A > \Delta_A^{crit}$. From (22) we have:

$$\frac{\partial w_B}{\partial \delta} = \frac{(1-r)\,\Delta_B\left(p_2 - p_1\right) - e\left[p_1 - \left[rp_0 + (1-r)\, q_0\right]\right]}{\left[p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\, q_0\right]\right]^2} \tag{41}$$

which is strictly increasing in Δ_B. Evaluating (41) at the point $\Delta_B = \Delta_B^{max}$ yields zero; accordingly, for all $\Delta_B < \Delta_B^{max}$, (41) is strictly negative.

For part *(iv)*, from (23) we have:

$$\frac{\partial w_A}{\partial r} = \delta \frac{e\left(p_0 - q_0\right) - \Delta_A\left(p_2 - (1-\delta)p_1 - p_0\delta\right)}{\left[p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\, q_0\right]\right]^2} \tag{42}$$

for all $\Delta_A \leq \Delta_A^{crit}$. Evaluating (42) at the point $\Delta_A = \widehat{\Delta}_A$ yields zero. Since (42) is strictly decreasing in Δ_A, it follows that for all $\Delta_A > \widehat{\Delta}_A$, (42) is strictly negative. Next, for all $\Delta_A > \Delta_A^{crit}$, from (23) we have:

$$\frac{\partial w_A}{\partial r} = \delta \frac{e\left(p_0 - q_0\right) - \Delta_A\left[p_2 - (1-\delta)\left(q_1 - q_0\right) - p_0\right]}{\left[p_2 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\, q_0\right]\right]^2} \tag{43}$$

One can show that (43) is decreasing in Δ_A and strictly negative when evaluated at the point $\Delta_A = \Delta_A^{crit}$; accordingly, we have (43) strictly negative for all $\Delta_A > \Delta_A^{crit}$. From (22), we have:

$$\frac{\partial w_B}{\partial r} = \delta \frac{e\left[p_0 - q_0\right] - \left(p_2 - (1-\delta)p_1 - \delta p_0\right)\Delta_B}{\left[p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\, q_0\right]\right]^2} \tag{44}$$

which is strictly decreasing in Δ_B. Note that (44) becomes equal to zero when Δ_B is equal to:

$$\frac{e\left[p_0 - q_0\right]}{p_2 - (1-\delta)p_1 - \delta p_0} \tag{45}$$

which can be shown to be strictly positive, but strictly less than Δ_B^{max}. Accordingly, when $\Delta_B \leq 0$, we have $\frac{\partial w_B}{\partial r} > 0$. For $\Delta_B > 0$, $\frac{\partial w_B}{\partial r} > 0$ *iff.* Δ_B is sufficiently small. □

Proof of Proposition 5. For part *(i)*, from (21), for all $\Delta_A > \widehat{\Delta}_A$, w_A^{del} must satisfy:

$$w_A^{del} > \frac{e + \delta\,(1-r)\,\widehat{\Delta}_A}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\, q_0\right]} \tag{46}$$

$$\iff w_A^{del} > \frac{e}{p_2 - (1-\delta)p_1 - \delta p_0} = w^{cent} \tag{47}$$

Accordingly, $\alpha_A > 0$. w_B^{del} can be higher or lower than w^{cent}: when Δ_B is zero (or negative), we clearly have

$w_B^{del} < w^{cent}$ by comparison of (9) and (22); in contrast, at the point $\Delta_B = \Delta_B^{max}$ we have:

$$w_B^{del} = \frac{e}{p_2 - p_1} > w^{cent} \tag{48}$$

For part *(ii)*, from (9) and (23), for $\Delta_A \leq \Delta_A^{crit}$, we have:

$$\frac{\partial \alpha_A}{\partial e} = \frac{1}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} - \frac{1}{p_2 - (1-\delta)p_1 - \delta p_0} \tag{49}$$

$$\frac{\partial \alpha_A}{\partial e} < 0 \iff rp_0 + (1-r)\,q_0 < p_0 \tag{50}$$

which is clearly satisfied. Similarly, from (9) and (23), for $\Delta_A > \Delta_A^{crit}$, we have:

$$\frac{\partial \alpha_A}{\partial e} = \frac{1}{p_2 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} - \frac{1}{p_2 - (1-\delta)p_1 - \delta p_0} \tag{51}$$

$$\frac{\partial \alpha_A}{\partial e} < 0 \iff 0 < (1-\delta)\,(p_1 - q_1) + \delta\,(1-r)\,(p_0 - q_0) \tag{52}$$

which is also satisfied. By a similar argument we also have $\frac{\partial \alpha_B}{\partial e} < 0$.

For part *(iii)*, using (26) and (39), for $\Delta_A \leq \Delta_A^{crit}$, we have:

$$\frac{\partial \alpha_A}{\partial \delta} = \frac{(1-r)\,\Delta_A\,(p_2 - p_1) - e\left[p_1 - \left[rp_0 + (1-r)\,q_0\right]\right]}{\left[p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]\right]^2} - \frac{-e\,(p_1 - p_0)}{\left[p_2 - (1-\delta)p_1 - \delta p_0\right]^2} \tag{53}$$

Evaluating this at the point $\Delta_A = \widehat{\Delta}_A$ yields:

$$\frac{e\,(p_1 - p_0)}{\left[p_2 - (1-\delta)p_1 - \delta p_0\right]}\left[\frac{1}{\left[p_2 - (1-\delta)p_1 - \delta p_0\right]} - \frac{1}{\left[p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]\right]}\right] \tag{54}$$

which is strictly positive from the foregoing. Moreover, since (53) is increasing in Δ_A, we have $\frac{\partial \alpha_A}{\partial \delta} > 0$ for all $\Delta_A \leq \Delta_A^{crit}$. Next, using (26) and (40), for all $\Delta_A > \Delta_A^{crit}$ we have:

$$\frac{\partial \alpha_A}{\partial \delta} = \frac{-\Delta_A\left[r\,(p_2 - p_0) + (q_1 - q_0)\,(1-r)\right] - e\left[q_1 - \left[rp_0 + (1-r)\,q_0\right]\right]}{\left[p_2 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\,q_0\right]\right]^2} + \frac{e\,(p_1 - p_0)}{\left[p_2 - (1-\delta)p_1 - \delta p_0\right]^2} \tag{55}$$

which is strictly decreasing in Δ_A. Accordingly, for Δ_A sufficiently high $\frac{\partial \alpha_A}{\partial \delta}$ becomes negative; it can be shown that the level required can be greater than or less than Δ_A^{max}, depending on the chosen set of parameters. Finally, using (26) and (42) we have:

$$\frac{\partial \alpha_B}{\partial \delta} = \frac{(1-r)\,\Delta_B\,(p_2 - p_1) - e\left[p_1 - \left[rp_0 + (1-r)\,q_0\right]\right]}{\left[p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]\right]^2} - \frac{-e\,(p_1 - p_0)}{\left[p_2 - (1-\delta)p_1 - \delta p_0\right]^2} \tag{56}$$

which is strictly increasing in Δ_B. Accordingly, $\frac{\partial \alpha_B}{\partial \delta}$ is positive *iff.* Δ_B becomes sufficiently large; it can be shown that the level required can be positive or negative, depending on the chosen set of parameters. □

Proof of Proposition 6. (i) From (23) and (22), for the case where $\Delta_A < \Delta_A^{crit}$ the derivatives are identical.

When $\Delta_A > \Delta_A^{crit}$, we have $\frac{\partial w_A}{\partial e} < \frac{\partial w_B}{\partial e}$ iff.:

$$\frac{1}{p_2 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} < \frac{1}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} \tag{57}$$

$$\iff \quad 0 < (1-\delta)\,(p_1 - q_1) \tag{58}$$

which is always satisfied. *(ii)* Similarly, when $\Delta_A < \Delta_A^{crit}$ the two derivatives are identical. When $\Delta_A > \Delta_A^{crit}$, we have $\frac{\partial w_A}{\partial \Delta_A} > \frac{\partial w_B}{\partial \Delta_B}$ iff.:

$$\frac{(1-\delta r)}{p_2 - (1-\delta)q_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} > \frac{\delta\,(1-r)}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} \tag{59}$$

$$\iff \quad p_2 - (1-\delta r)\,p_1 - \delta r p_0 + \delta\,(1-r)\,(q_1 - q_0) > 0 \tag{60}$$

which is also always satisfied. For part *(iii)*, note from the discussions in the main text that a necessary condition for delegation to be strictly optimal is $\alpha_B < 0$, or $w_B^{del} < w^{cent}$. This requires that:

$$\frac{e + \delta\,(1-r)\,\Delta_B}{p_2 - (1-\delta)p_1 - \delta\left[rp_0 + (1-r)\,q_0\right]} < \frac{e}{p_2 - (1-\delta)p_1 - \delta p_0} \tag{61}$$

$$\iff \quad \Delta_B < \frac{e\left[p_0 - q_0\right]}{p_2 - (1-\delta)p_1 - \delta p_0} \tag{62}$$

From the Proof of Proposition 4, this implies that $\frac{\partial w_B}{\partial r} > 0$. Moreover, Proposition 4 states that $\frac{\partial w_A^{del}}{\partial r} < 0$, completing the proof. $\square$

Proof of Proposition 7. We initially restrict attention to cases where $\Delta_A \leq \Delta_A^{crit}$ and $\Delta_{sum} = \Delta_A + \Delta_B \geq 0$; the alternative cases are discussed at the end of the proof. We begin by deriving conditions under which the principal's desired outcome, whereby both agents work and the correct method of production is implemented in each period, maximises the agents' surplus amongst all possible outcomes in both states, regardless of whether they can be implemented as an equilibrium. It is straightforward to verify that this is the case *iff.* the following set of constraints holds:

$$(p_2 - p_1)(w_A + w_B) \geq e \tag{63}$$

$$(p_2 - p_0)(w_A + w_B) \geq 2e \tag{64}$$

$$(p_2 - q_2)(w_A + w_B) \geq -\Delta_{sum} \tag{65}$$

$$(p_2 - q_1)(w_A + w_B) \geq e - \Delta_{sum} \tag{66}$$

$$(p_2 - q_0)(w_A + w_B) \geq 2e - \Delta_{sum} \tag{67}$$

$$(p_2 - q_2)(w_A + w_B) \geq \Delta_{sum} \tag{68}$$

$$(p_2 - q_1)(w_A + w_B) \geq e + \Delta_{sum} \tag{69}$$

$$(p_2 - q_0)(w_A + w_B) \geq 2e + \Delta_{sum} \tag{70}$$

Claim 1. Let $\Delta_{sum} \geq 0$. If (64), (68) and (70) hold, all eight constraints (63)-(70) are satisfied.

Proof. If (64) holds, then:

$$(p_2 - p_1)(w_A + w_B) + (p_1 - p_0)(w_A + w_B) \geq 2e \tag{71}$$

which implies that (63) is satisfied since $p_2 - p_1 > p_1 - p_0$. Next, summing (68) and (70) and rearranging yields:

$$(2p_2 - (q_2 + q_0))(w_A + w_B) \geq 2e + 2\Delta_{sum} \tag{72}$$

Since $q_2 - q_1 > q_1 - q_0 \iff q_2 + q_0 > 2q_1$, we therefore have:

$$(2p_2 - 2q_1)(w_A + w_B) \geq 2e + 2\Delta_{sum} \tag{73}$$

which implies (69). Finally, note that since we are restricting attention to the case where $\Delta_{sum} \geq 0$, (68), (69) and (70) immediately imply (65), (66) and (67), respectively. $\square$

Since we are restricting attention to cases where $\Delta_A \leq \Delta_A^{crit}$, using (23) and (22), conditions (64), (68) and (70) can be rewritten as:

$$\Delta_{sum} \geq \frac{2e\left[p_0-(1-\delta)p_1-\delta\left[rp_0+(1-r)\,q_0\right]\right]}{(p_2-p_0)\,\delta\,(1-r)} =: \widetilde{\Delta}^1_{sum} \tag{74}$$

$$\Delta_{sum} \leq \frac{(p_2-q_2)\,2e}{p_2-(1-\delta)p_1-\delta\left[rp_0+(1-r)\,q_0\right]-(p_2-q_2)\,\delta\,(1-r)} =: \widetilde{\Delta}^2_{sum} \tag{75}$$

$$\Delta_{sum} \leq \frac{2e\left[(1-\delta)p_1+\delta\left[rp_0+(1-r)\,q_0\right]-q_0\right]}{p_2-p_2\delta\,(1-r)-(1-\delta)p_1-\delta rp_0} =: \widetilde{\Delta}^3_{sum} \tag{76}$$

respectively. Moreover, the principal will choose to implement delegation *iff.* the following condition holds:

$$\frac{2e+\delta\,(1-r)\,\Delta_{sum}}{p_2-(1-\delta)p_1-\delta\left[rp_0+(1-r)\,q_0\right]} < \frac{2e}{p_2-(1-\delta)p_1-\delta p_0} \tag{77}$$

$$\iff \Delta_{sum} < \frac{2e\,(p_0-q_0)}{p_2-(1-\delta)p_1-\delta p_0} =: \Delta^*_{sum} \tag{78}$$

That is, if total wages under delegation are strictly less than total wages under centralisation. We have the following:

$$\Delta^*_{sum} > \widetilde{\Delta}^1_{sum} \tag{79}$$

$$\iff (1-\delta)\,(p_1-p_0)\left[p_2-(1-\delta)p_1-\delta\left[rp_0-(1-r)\,q_0\right]\right] > 0 \tag{80}$$

$$\widetilde{\Delta}^2_{sum} > \Delta^*_{sum} \tag{81}$$

$$\iff \left[(p_2-p_0)-(q_2-q_0)\right]\left[p_2-(1-\delta)p_1-\delta\left[rp_0+(1-r)\,q_0\right]\right] > 0 \tag{82}$$

$$\widetilde{\Delta}^3_{sum} > \Delta^*_{sum} \tag{83}$$

$$\iff (1-\delta)\left[p_1-p_0\right]\left[p_2-(1-\delta)p_1-\delta\left[rp_0+(1-r)\,q_0\right]\right] > 0 \tag{84}$$

all three of which are always satisfied. In addition, we have $\Delta^*_{sum} > 0$.

Altogether, we have the following. One can always choose a combination of $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{crit}\right]$ and $\Delta_B < \Delta_B^{max}$ which yield a particular value of $\Delta_{sum} \in \left[0, \widehat{\Delta}_A + \Delta_B^{max}\right)$. Moreover, it is straightforward to verify that $\Delta_A^{crit} + \Delta_B^{max} > \Delta^*_{sum}$; accordingly, the intersection between the intervals $\left[0, \Delta_A^{crit} + \Delta_B^{max}\right)$ and $\left[\widetilde{\Delta}^1_{sum}, \Delta^*_{sum}\right)$ is non-empty.

Select any combination of $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{crit}\right]$ and $\Delta_B < \Delta_B^{max}$ such that the resulting Δ_{sum} is in this intersection of intervals. By Proposition 3, delegation which induces the alternative punishment strategy to centralisation is implementable and since we have $\Delta_{sum} < \Delta^*_{sum}$, the total wage costs are strictly lower than those under centralisation. Moreover, since $\Delta_{sum} \geq \widetilde{\Delta}^1_{sum}$ and $\Delta_{sum} < \Delta^*_{sum} < \widetilde{\Delta}^2_{sum}, \widetilde{\Delta}^3_{sum}$, by Claim 1 the principal's desired outcome maximises the agents' surplus.

A formal examination of the cases where $\Delta_A > \Delta_A^{crit}$, or where $\Delta_{sum} < 0$ is not necessary for the proof. Nonetheless, one can show that in all situations where $\Delta_A > \Delta_A^{crit}$, $\Delta_{sum} > 0$ and delegation is optimal as an organisational structure, Δ_A and Δ_B can never become so high that colluding to implement production method γ_1 in both states of the world increases the joint surplus, as argued in the main text. However, the proof is significantly lengthier in this case, and is therefore omitted. One implication of this result is that, by continuity, there will always exist a non-empty set of (Δ_A, Δ_B) pairs, with $\Delta_A > \Delta_A^{crit}$, such that delegation is both optimal and collusion proof. □

Appendix II

We provide a numerical example of a situation in which the equilibrium induced by delegation is not collusion proof; that is, there exists an alternative outcome which *(i)* is weakly better for both players and strictly better for one and *(ii)* can be sustained as an equilibrium.

Consider the following set of parameters: $e = 1$, $r = \frac{1}{2}$, $\delta = \frac{1}{2}$; we let $q_l = \kappa p_l$ for $l = 0, 1, 2$, with $p_2 = \frac{1}{2}$, $p_1 = \frac{502}{1024}$, $p_0 = \frac{497}{1024}$ and $\kappa = \frac{1}{2}$. We set $c_A^1 = 0$ and $c_A^2 = \frac{2000}{95}$ so that $\Delta_A = \frac{2000}{95} \approx 21.05$. We also set $c_B^1 = c_B^2 = 1$, so that $\Delta_B = 0$. Note that with these parameters, we have $\widehat{\Delta}_A = 19.88$, $\Delta_A^{max} = 25.35$ and $\Delta_B^{max} = 25.85$; accordingly, we have $\Delta_A \in \left(\widehat{\Delta}_A, \Delta_A^{max}\right)$ and $\Delta_B < \Delta_B^{max}$ as required by Proposition 3. In addition, $\Delta_A^{crit} = 21.14$ so that we also have $\Delta_A < \Delta_A^{crit}$.

Using (9), the total wage payment under centralisation is given by $2w^{cent} = 163.84$. Moreover, the expected utility of each agent is as follows:

$$EU_A^{cent} = p_2 w^{cent} - e - rc_A^1 - (1-r)c_A^2 = 29.43 \tag{85}$$

$$EU_B^{cent} = p_2 w^{cent} - e - rc_B^1 - (1-r)c_B^2 = 38.96 \tag{86}$$

so that both agents prefer to participate.

From (23) and (22), keeping in mind that $\Delta_A < \Delta_A^{crit}$, the sum of the wages under delegation is $w_A^{del} + w_B^{del} = 99.66$. Their respective expected utilities are given by:

$$EU_A^{HT} = p_2 w_A^{del} - e - rc_A^1 - (1-r)c_A^2 = 31.45 \tag{87}$$

$$EU_B^{HT} = p_2 w_B^{del} - e - rc_B^1 - (1-r)c_B^2 = 4.86 \tag{88}$$

so that again, both agents prefer to participate. It follows that, in the absence of collusion, the principal strictly prefers to implement delegation.

We now show that if the principal chooses to implement delegation, there exists a collusive strategy which yields strictly higher utility to both players. Suppose that in state ω_1, Agent A always plays γ_1 *and shirk* and in state ω_2, always plays γ_2 *and shirk*. That is, Agent A always implements the correct method of production and shirks. Suppose that Agent B, regardless of the state of the world, plays *work* with probability $\frac{3}{4}$ and *shirk* with probability $\frac{1}{4}$. This yields the following expected utility to each agent:

$$EU_A^{col} = r\left[\left(\frac{3}{4}p_1 + \frac{1}{4}p_0\right)w_A - c_A^1\right] + (1-r)\left[\left(\frac{3}{4}p_1 + \frac{1}{4}p_0\right)w_A - c_A^2\right] = 31.5 \tag{89}$$

$$EU_B^{col} = r\left[\frac{3}{4}(p_1 w_B - e) + \frac{1}{4}p_0 w_B - c_B^1\right] + (1-r)\left[\frac{3}{4}(p_1 w_B - e) + \frac{1}{4}p_0 w_B - c_B^2\right] = 4.96 \tag{90}$$

Since $EU_A^{col} > EU_A^{HT}$ and $EU_B^{col} > EU_B^{HT}$, both players are strictly better off under these sets of strategies compared to the principal's desired equilibrium. It remains to show that these strategies form an equilibrium of the dynamic game, given the threat of punishment: (γ_1 *and shirk*; *shirk*) being played in every period and in either state.[32]

We first consider Agent A's incentive to deviate. In state ω_1, by the proof of Proposition 3, Agent A's most profitable deviation is γ_1 *and shirk*, which is his prescribed action; hence, Agent A has no short term incentive to deviate from this strategy in state ω_1. In state ω_2, if Agent B shirks, by the proof of Proposition 3 Agent A's most profitable deviation is γ_1 *and shirk*; if Agent B works, Agent A's most profitable deviation is either γ_1 *and shirk* or γ_2 *and shirk*, the latter of which is his prescribed action.[33] Altogether, for Agent A we require the following constraints to hold:

$$(1-\delta)(p_0 w_A - c_A^2) + \delta EU_A^{col} \geq (1-\delta)\left(q_0 w_A - c_A^1\right) + \delta\left[rp_0 w_A + (1-r)\,q_0 w_A - c_A^1\right] \tag{91}$$

$$(1-\delta)(p_1 w_A - c_A^2) + \delta EU_A^{col} \geq (1-\delta)\left(q_1 w_A - c_A^1\right) + \delta\left[rp_0 w_A + (1-r)\,q_0 w_A - c_A^1\right] \tag{92}$$

Note that since $p_0 - q_0 < p_1 - q_1$, if (91) holds then (92) does automatically. Numerically, (91) reduces to:

$$26.08 \geq 26.07$$

which is satisfied.

Next, we consider Agent B's incentive to deviate. In either state, by the proof of Proposition 3, Agent B's most profitable short-term deviation is to *shirk*. Thus, we must check that Agent B would not shirk (in either state) when his prescribed action is to work; we therefore require:

$$(1-\delta)(p_1 w_B - e - c_B^1) + \delta EU_B^{coll} \geq (1-\delta)\left(p_0 w_B - c_B^1\right) + \delta\left[rp_0 w_B + (1-r)\,q_0 w_B - c_B^1\right] \tag{93}$$

$$(1-\delta)(p_1 w_B - e - c_B^2) + \delta EU_B^{coll} \geq (1-\delta)\left(p_0 w_B - c_B^2\right) + \delta\left[rp_0 w_B + (1-r)\,q_0 w_B - c_B^1\right] \tag{94}$$

Note that these constraints are equivalent. Numerically, they both reduce to:

$$4.84 \geq 4.83$$

which is also satisfied. Hence, neither player has an incentive to deviate and thus the collusive strategy outlined forms an equilibrium in which both players are strictly better off than the principal's desired equi-

[32] By the proof of Proposition 3, this is the unique pure strategy Nash Equilibrium of both underlying stage games.

[33] By the proof of Proposition 3, Agent A's most profitable deviation in the short term is always to shirk.

librium.[34]

References

Adrian, N. & Möller, M. (2020), 'Self-managed work teams: An efficiency-rationale for pay compression', *Journal of Economics & Management Strategy* **29**(2), 315–334.

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

Au, P. H. (2020), 'Pay to quit and team incentives', *Journal of the European Economic Association* **18**(2), 964–1008.

Baker, G., Gibbons, R. & Murphy, K. J. (1999), 'Informal authority in organizations', *Journal of Law, Economics, and Organization* **15**(1), 56–73.

Baldenius, T., Glover, J. & Xue, H. (2016), 'Relational contracts with and between agents', *Journal of Accounting and Economics* **61**(2-3), 369–390.

Boning, B., Ichniowski, C. & Shaw, K. (2007), 'Opportunity counts: Teams and the effectiveness of production incentives', *Journal of Labor Economics* **25**(4), 613–650.

Bresnahan, T. F., Brynjolfsson, E. & Hitt, L. M. (2002), 'Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence', *Quarterly Journal of Economics* **117**(1), 339–376.

Chalioti, E. (2016), 'Team production, endogenous learning about abilities and career concerns', *European Economic Review* **85**, 229–244.

Che, Y.-K. & Yoo, S.-W. (2001), 'Optimal incentives for teams', *American Economic Review* **91**(3), 525–541.

Douglas, C. & Gardner, W. L. (2004), 'Transition to self-directed work teams: implications of transition time and self-monitoring for managers' use of influence tactics', *Journal of Organizational Behavior* **25**(1), 47–65.

Flassak, K. & Hofmann, C. (2020), 'A note on the effect of environmental uncertainty on team composition', *Working paper* .

Fudenberg, D. & Maskin, E. (1986), 'The folk theorem in repeated games with discounting or with incomplete information', *Econometrica* **54**(3), 533–554.

Glover, J. & Kim, E. (2020), 'Teams, career horizon diversity, and tacit collusion', *Working paper* .

Glover, J. & Kim, E. (2021), 'Optimal team composition: Diversity to foster implicit team incentives', *Management Science* .

Hamilton, B. H., Nickerson, J. A. & Owan, H. (2003), 'Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation', *Journal of Political Economy* **111**(3), 465–497.

[34] Using the notation of the Proof of Proposition 7, note that we have $\Delta_{sum} = 21.05 < \widetilde{\Delta}^1_{sum} = 31.8$. Accordingly, both players shirking yields a strictly higher surplus than both players working; this is what opens the door to collusion in this example.

Hermalin, B. E. (1998), 'Toward an economic theory of leadership: Leading by example', *American Economic Review* **88**(5), 1188–1206.

Hermalin, B. E. (2007), 'Leading for the long term', *Journal of Economic Behavior & Organization* **62**(1), 1–19.

Hermalin, B. E. (2017), 'At the helm, kirk or spock? the pros and cons of charismatic leadership', *Working paper* .

Hidir, S. & Migrow, D. (2021), 'Credibility, efficiency and the structure of authority', *Working paper* .

Holmström, B. (1982), 'Moral hazard in teams', *Bell Journal of Economics* **13**(2), 324–340.

Huck, S. & Rey-Biel, P. (2006), 'Endogenous leadership in teams', *Journal of Institutional and Theoretical Economics* **162**(2), 253–261.

Ishihara, A. (2017), 'Relational contracting and endogenous formation of teamwork', *The RAND Journal of Economics* **48**(2), 335–357.

Ishihara, A. & Muramoto, A. (2021), 'Cooperation or collusion? rents in relational contracts for teams', *Working paper* .

Itoh, H. (1992), 'Cooperation in hierarchical organizations: An incentive perspective', *Journal of Law, Economics, & Organization* **8**(2), 321–345.

Itoh, H. (1993), 'Coalitions, incentives, and risk sharing', *Journal of Economic Theory* **60**(2), 410–427.

Jackson, S. E. (1996), *Handbook of Work Group Psychology*, John Wiley & Sons Ltd., chapter The consequences of diversity in multidisciplinary work teams, pp. 53–75.

Komai, M. & Stegeman, M. (2010), 'Leadership based on asymmetric information', *The RAND Journal of Economics* **41**(1), 35–63.

Kräkel, M. (2017), 'Self-organizing teams', *Economics Letters* **159**, 195–197.

Kvaløy, O. & Olsen, T. E. (2006), 'Team incentives in relational employment contracts', *Journal of Labor Economics* **24**(1), 139–169.

Kvaløy, O. & Schöttner, A. (2015), 'Incentives to motivate', *Journal of Economic Behavior & Organization* **116**, 26–42.

Lazear, E. P. (2012), 'Leadership: A personnel economics approach', *Labour Economics* **19**(1), 92–101.

Lazear, E. P. & Oyer, P. (2012), *The Handbook of Organizational Economics*, Princeton University Press, Princeton, NJ, chapter Personnel Economics, pp. 479–519.

Lazear, E. P. & Shaw, K. L. (2007), 'Personnel economics: The economist's view of human resources', *Journal of Economic Perspectives* **21**(4), 91–114.

Levasseur, R. E. (2005), 'People skills: Change management tools — leading teams', *Interfaces* **35**(2), 179–180.

Magpili, N. C. & Pazos, P. (2018), ‘Self-managing team performance: A systematic review of multilevel input factors’, *Small Group Research* **49**(1), 3–33.

Morgeson, F. P., DeRue, D. S. & Karam, E. P. (2010), ‘Leadership in teams: A functional approach to understanding leadership structures and processes’, *Journal of Management* **36**(1), 5–39.

Muir, J. (1995), ‘Effective management through delegation’, *Work Study* **44**(7), 6–7.

Myers, A., Kakabadse, A., McMahon, T. & Spony, G. (1995), ‘Top management styles in Europe: Implications for business and cross-national teams’, *European Business Journal* **7**(1), 17–28.

Rajan, R. G. & Wulf, J. (2006), ‘The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies’, *Review of Economics and Statistics* **88**(4), 759–773.

Rohlfing-Bastian, A. & Schöttner, A. (2020), ‘Delegation of job design and the value of communication when performance measurement is imperfect’, *Working paper* .

Van Knippenberg, D. & Mell, J. N. (2016), ‘Past, present, and potential future of team diversity research: From compositional diversity to emergent diversity’, *Organizational Behavior and Human Decision Processes* **136**, 135–145.

Villas-Boas, J. M. (2020), ‘Repeated interaction in teams: Tenure and performance’, *Management Science* **66**(3), 1496–1507.