UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede Amministrativa Università Degli Studi Di Padova
Dipartimento di Scienze Economiche e Aziendali "Marco Fanno"
SCUOLA DI DOTTORATO DI RICERCA IN
ECONOMIA E MANAGEMENT
CICLO XXIII

THREE ESSAYS ON INCENTIVES IN SMALL GROUPS

Direttore della Scuola: Ch.mo Prof. Guglielmo Weber

Supervisore: Ch.mo Prof. Antonio Nicolò

Dottoranda: Regine Oexl

31 Gennaio 2012

A tutti quelli che mi vogliono bene.

Vorrei ringraziare innanzitutto il mio supervisore Antonio Nicolò, guida e modello non soltanto scientifico. Poi vorrei ringraziare i miei colleghi che hanno reso piacevole la vita durante il dottorato, in particolare: Veronica, Betty, Devis, Diana, Francesca, Francesco, Gabriel, Michele F., Natalia, Pablo, Paolo R., Paolo S., e Valentina. Ringrazio anche i Professori della scuola per quello che mi hanno insegnato e per avermi fatto stare bene. Grazie anche ai miei coautori, Zachary Grossman e Natalia Montinari.

Un ringraziamento anche ai miei compagni di appartamento: Manfred, Michele, Enrico, Gianluca, ed ai vecchi compagni, Nick e Andrea, senza i quali la mia vita qui a Padova non sarebbe stata così meravigliosa. Vorrei anche ringraziare Gonzalo, Magdalena, Santi e Thomas per la compagnia.

Anche per loro colpa amo tanto Padova e dintorni: Giulia Beltrame, Ale, Aronne, Beppe, Cespo, Giorgio, Elisa B., Elisa S., Erika, Laura C., Luca, Mary, Nicolino, Oscar, Pietro T., Stefi, e la compagnia del Pollaio, in particolare Alice, Benry, Coma, Diego, Erica, Maria, Mattia, il Moro, Paride, Pietro V., Steno, e Nicolò. Vorrei anche ringraziare a Sam e Marcus per i tempi divertenti a Boston e sulla neve, e Raffaella.

Grazie anche alla "Cassa di Risparmio di Padova e Rovigo" per il sostegno finanziario e per avermi fatto conoscere tanta bella gente nelle diverse residenze studentesche dove mi ha fatto vivere.

In fine ma non per ultimi vorrei ringraziare ovviamente la mia sorella gemella Tina per tutto, i miei genitori e la mia famiglia per aver avuto fiducia in me.

Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

List of Figures

Introduction: Incentives in Small Groups

This PhD thesis is about incentives in the presence of agents who interact in small groups. It is composed by three chapters, each corresponding to a self-contained paper, applying different methodologies (theoretical and empirical) and different perspectives. I refer to groups not as collection of independently acting individuals, but, in line with Arrow, McGrath, and Berdahl (2000), as "adaptive, dynamic systems that are driven by interactions both among group members and between the group and its embedding contexts". Following such point of view, I focus on groups in which individuals' each actions revoke responses by another individuals. While the first chapter considers purely rational individuals, abstracting from other-regarding preferences such as reciprocity, inequity aversion, and fairness considerations (Dufwenberg, Heidhues, Kirchsteiger, Riedel, and Sobel, 2008; Sobel, 2005; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002), in the other two chapters I include such behavioral aspects, which lead to the emergence of new strategic behaviors with respect to the case in which such preferences are not taken into account.

Hence, the aim of this thesis is not only to understand the working and reasoning behind incentivized decision making, to be able to explain observable phenomena in economic and social interaction, but also to study mechanisms that may restore incentives in cases in which efficiency needs to be increased, and, when possible, establish optimal incentives. While the first chapter contains a theoretical model dealing with an optimal mechanism, considering the interaction among several agents or firms on the market, the second chapter treats with a mechanism dealing with agents inside an organization, applying experimental methodology. Instead of suggesting a new mechanism, as the first chapter does, the second chapter tries to understand the mechanisms behind the observable phenomena of having mediocre agents in organizations. Finally, the last chapter tries to explain the reasoning behind the delegating of decisions, which may play a role either inside a firm (when a manager delegates to his subordinate) or outside (when two firms negotiate).

More in detail, the first chapter deals with the aspect of incentivizing a firm to take the optimal investment decision when simple complete contracts cannot be written as to induce the desired outcome, or, more precisely, the hold-up problem is present (Hart and Moore, 1999). The principal interacts with two or more agents, where agent may be understood as an individual player or as well as firm. More specifically, the case I consider is a seller vending identical products to two buyers that have a common interest in inducing the seller to make a quality enhancing investment. I show that a trilateral contract may provide the correct incentives to restore efficiency, in a setting in which two bilateral contracts would have failed to provide such incentives. Because there is more than one buyer, the purchasing decision can be transformed from a contemporaneous to a sequential problem. This allows to condition the exchange of an extra payment on verifiable previous transactions. The contract induces a coalition proof Nash equilibrium and holds under complete as well as incomplete information. The presence of two or more agents is important, since it is the sequentiality with which payments of the two agents are exchanged with the principal that resolve the issue of incentive compatibility. With one agent, this sequentiality could not be made use of. The number of agents may be easily extended to more than two buyers.

The second chapter applies economic experimental methodology to study the incentives behind employing mediocre agents in organizations, and its focus is individual decision taking. Me and my coauthors Natalia Montinari and Antonio Nicolò run a decision making experiment in which a principal selects one agent to perform a task for a fixed compensation. Agents' productivity depends on two components, a non contractible ex-post effort and an exogenous ex-ante productivity. The principal chooses between two agents with different ex-ante productivity, in a setting with common information. Once the principal has selected one agent, this agent chooses a level of costly non-contractible effort. We run three different treatments: one treatment with communication, in which the principal can send a message to the selected agent; one treatment without communication, and a control treatment where the agent is selected by a random device. Our results are the following: in treatments where the principal selects the agent, a significant share of principals employ the agent who has the ex-ante lowest productivity. In the communication treatment, the selected agents with lowest ex-ante productivity exert a significantly higher effort compared to the agents with ex-ante highest productivity. The higher effort overcompensates the low ex-ante productivity, and principals who have chosen low ex-ante productive agents gain on average 40% more than principals who hired high ex-ante productive agents. In the no-communication treatment and in the control treatment the average effort exerted by the ex-ante low productive and ex-ante high productive agents does not differ significantly. Our experiment provides a rationale for the selection of mediocre (ex-ante less productive) workers in organizations. Agents who are ex-ante less entitled to fill a position reciprocate more than individuals who are ex-ante more competent, and exert a higher non contractible effort ex-post when principals are able to inspire mediocre agents to feel indebted towards them. It follows that principals may find it profitable to hire agents who ex-ante can be viewed as less qualified. We suggest that this is a prominent characteristic in organizations, like civil service, where i) reciprocity cannot be induced by means of higher wages and ii) non-contractible effort is a relevant component of the employer's production function.

The third chapter presents again a laboratory experiment, with which me and my coauthor Zachary Grossman investigate the incentives behind the delegation of a decision. Beyond the classical reasons of efficiency, commitment, the distribution of information, or incentive provision, a person may also delegate decision rights so as to avoid blame for an unpopular or immoral decision. Extending the results of Bartling and Fischbacher (2011), we show that by delegating to an intermediary, a dictator facing an allocation decision can effectively shift moral responsibility onto the delegee *even when doing so necessarily eliminates the possibility of a fair outcome*. Dictators choosing selfishly via an intermediary are punished less and earn greater profits than those who do so directly. Despite being powerless to influence the fairness of the outcome, an intermediary given the choice between two unfair outcomes is punished more than when the dictator chooses one directly. This is not the case when the intermediary merely can initiate the random selection of one of the outcomes. Our findings reinforce and clarify the usefulness of agency as a tool to evade perceived culpability, addresses the limits of blame shifting, and makes an interesting contribution to the emerging literature on the delegation of responsibility.

Arrow, H., J. E. McGrath, and J. L. Berdahl (2000): *Small Groups as Complex Systems: Formation, Coordination, Development, and Adaptation*. Sage Publications, Inc, 1 edn.

Bartling, B., and U. Fischbacher (2011): "Shifting the Blame: On Delegation and Responsibility", forthcoming in: Review of Economic Studies 32, Thurgauer Wirtschaftsinstitut, Universitaet Konstanz.

Bolton, G. E., and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition", American Economic Review, 90(1), 166 - 193.

Charness, G., and M. Rabin (2002): "Understanding Social Preferences With Simple Tests", The Quarterly Journal of Economics, 117(3), 817 - 869.

Dixit, A. (2003): "On Modes of Economic Governance", Econometrica, 71(2), 449 - 481.

Dufwenberg, M., P. Heidhues, G. Kirchsteiger, F. Riedel, and J. Sobel (2008): "Other-Regarding Preferences in General Equilibrium", CEPR Discussion Papers 6815.

Fehr, E., and K. M. Schmidt (1999): "A Theory Of Fairness, Competition, And Cooperation", The Quarterly Journal of Economics, 114(3), 817 - 868.

Hart, O., and J. Moore (1999): "Foundations of Incomplete Contracts", The Review of Economic Studies, 66(1), 115 - 138.

Sobel, J. (2005): "Interdependent Preferences and Reciprocity", Journal of Economic Literature, 43(2), 392 - 436.

Abstract [In Italian]: Incentivi in Gruppi Piccoli

Questa tesi analizza gli incentivi in presenza di agenti che interagiscono tra loro in piccoli gruppi. Questo lavoro è un compendio di tre articoli tra loro indipendenti che applicano metodologie diverse, sia teoriche che empiriche, e impiegano diverse prospettive. Arrow, McGrath e Berdahl (2000) descrivono i gruppi come sistemi adattivi e dinamici, che sono guidati da interazioni sia tra i componenti dei gruppi, sia tra i gruppi stessi e il loro contesto - e non come un insieme di individui che si comportano in modo indipendente gli uni dagli altri. Partendo da questo punto di vista, in questa tesi mi focalizzo su piccoli gruppi in cui gli individui interagiscono tra di loro e ciascuna azione dei singoli provoca una risposta degli altri componenti del gruppo. Nel primo capitolo vengono analizzati i comportamenti di individui puramente razionali mentre nei capitoli successivi si considerano anche aspetti comportamentali tra i quali la reciprocità, avversione alla disuguaglianza, il senso di giustizia (Dufwenberg, Heidhues, Kirchsteiger, Riedel, e Sobel, 2008; Sobel, 2005; Fehr e Schmidt, 1999; Bolton e Ockenfels, 2000; Charness e Rabin, 2002), che inducono comportamenti strategici diversi rispetto a quando non si considerano le preferenze sociali.

L'obiettivo di questo lavoro è, quindi, non solo di analizzare i meccanismi di funzionamento dei processi decisionali, ma anche di studiare come sia possibile generare degli incentivi in grado di migliorare il livello di efficienza, che è lo scopo ultimo dello sviluppo economico. Il primo e il secondo capitolo si concentrano sulla definizione di schemi di incentivi in contesti specifici. In particolare, il primo capitolo presenta un modello teorico che analizza gli incentivi ottimali considerando l'interazione tra diversi agenti o aziende presenti nel mercato. Mentre il primo capitolo assume che gli aspetti comportamentali non siano rilevanti, gli ultimi due capitoli introducono aspetti comportamentali nell'interazione degli agenti all'interno di un'organizzazione. Il primo capitolo definisce un nuovo sistema di incentivi , mentre nel secondo si analizzano degli aspetti comportamentali che spiegano la presenza di lavoratori mediocri nelle organizzazioni. Infine, l'ultimo capitolo si pone l'obiettivo di spiegare un meccanismo alla base del processo di delega decisionale, che si rivela di notevole importanza sia all'interno dell'azienda (quando, ad esempio, un manager delega la decisione ad un subordinato), sia tra aziende (quando due aziende negoziano tra loro).

Più in dettaglio, il primo capitolo affronta attraverso un modello teorico, come

incentivare un'organizzazione a prendere delle decisioni ottimali quando non possono essere stipulati dei semplici contratti bilaterali ottimali (Hart e Moore, 1999), o, più precisamente, quando sia presente il problema di "hold-up". Il principale interagisce con due (o più) agenti, dove per agente si intende sia un individuo singolo o un'azienda. Più specificamente, il caso considerato è quello di un'impresa che vende dei prodotti identici a due diversi acquirenti, che hanno un interesse comune ad indurre il venditore a fare un investimento che aumenta la possibilità che il prodotto sia di alta qualità. Attraverso un'analisi teorica si dimostra che un contratto trilaterale può fornire gli incentivi adatti a ripristinare l'efficienza, in un ambiente in cui due contratti bilaterali falliscono in questo compito. Data l'esistenza di una molteplicità di acquirenti, la decisione di acquisto può essere trasformata da una decisione simultanea ad un problema sequenziale. La sequenzialità permette di subordinare lo scambio di un pagamento addizionale alla verifica delle transazioni precedenti. Il contratto induce un equilibrio "coalition proof Nash" di decisione sequenziale, sia sotto l'ipotesi di informazione completa che incompleta. La presenza di due agenti è rilevante, dato che la sequenzialità con la quale i pagamenti dei due agenti sono stati versati al principale risolve il problema della compatibilità degli incentivi. Con un solo agente, infatti, non si potrebbe far uso di questa sequenzialità. A partire da questo risultato, il numero di agenti, può essere esteso facilmente a più di due.

Il secondo capitolo applica la metodologia sperimentale per studiare gli incentivi che motivano l'impiego di agenti mediocri nelle organizzazione, focalizzandosi sul meccanismo decisionale degli individui. Insieme a Natalia Montinari e Antonio Nicolò abbiamo realizzato un esperimento di laboratorio, in cui incentiviamo gli individui a prendere una decisione in cui un principale sceglie un agente per eseguire un compito, pagandogli un compenso fisso. La produttività dell'agente dipende da due componenti: uno sforzo non contrattabile a posteriori, ed una produttività esogena a priori. Il principale sceglie tra due agenti che hanno diversi livelli di produttività a priori. Una volta scelto un agente, questo sceglie un livello di sforzo costoso non contrattabile. Eseguiamo, quindi, tre trattamenti diversi: uno con comunicazione, in cui un principale può mandare un messaggio all'agente selezionato; uno senza comunicazione, ed infine un trattamento di controllo in cui l'agente viene selezionato in modo casuale. I risultati dell'esperimento sono i seguenti. Nei trattamenti nei quali il principale seleziona l'agente, una quota significativa dei principali impiega l'agente che ha a priori la produttività più bassa. Nel trattamento con comunicazione, gli agenti con la produttività a priori più bassa esercitano uno sforzo significativamente più alto rispetto agli agenti con la produttività a priori più alta. Lo sforzo più alto compensa in maniera più che proporzionale la produttività più bassa e i principali che scelgono l'agente a priori meno produttivo guadagnano in media il 40% di più rispetto ai principali che assumono un agente a priori più produttivo. Nel trattamento senza comunicazione e nel trattamento di controllo, non c'è evidenza dell'esistenza di uno sforzo più alto da parte dei dipendenti di produttività bassa. Dunque, il nostro esperimento fornisce una spiegazione razionale per la selezione di individui mediocri (cioè quelli che sono meno produttivi a priori) nelle organizzazioni. Coloro che sono scelti nonostante abbiano la consapevolezza di avere minori capacità reciprocano di più rispetto agli individui che sono più competenti a priori ed esercitano uno sforzo non-contrattabile più alto quando i principali sono in grado di far percepire agli agenti mediocri la loro condizione di inferiorità. Pertanto, i principali possono trovare remunerativo: assumere un agente che a priori può essere considerato meno qualificato. Suggeriamo che questa è una caratteristica importante specie nelle organizzazioni dove i) la reciprocità non può essere indotta tramite salari più alti, e ii) lo sforzo non contrattabile è una componente rilevante nella funzione di produttività del datore di lavoro (si pensi, per esempio, alle organizzazioni della pubblica amministrazione).

Anche il terzo capitolo, scritto a due mani con Zachary Grossman, presenta un esperimento di laboratorio. L'obiettivo del capitolo è di indagare gli incentivi che sottostanno alla delega di una decisione. Oltre ai tradizionali incentivi alla scelta di delegare una decisione identificati in letteratura, come l'efficienza, l'impegno, la distribuzione di informazione o la trasmissione di incentivi, studi recenti hanno suggerito che anche la volontà di evitare di addossarsi la colpa per una decisione impopolare o immorale possa motivare un agente a delegare. Bartling e Fischbacher (2011) hanno dimostrato che, tramite la delega di una decisione ad un intermediario, un dittatore che è di fronte a una decisione di allocazione può effettivamente spostare la responsabilità morale sulla persona che poi prende la decisione (il delegato). Grazie al nostro esperimento si dimostra che questo meccanismo di trasferimento della responsabilità funziona anche se la scelta di delegare comporta necessariamente l'implementazione di una scelta ingiusta. I dittatori che scelgono l'allocazione ingiusta sono puniti di meno e guadagnano di più se utilizzano un intermediario rispetto a quelli che prendono questa decisione senza delegare. Pur non essendo in grado di influenzare il risultato, un intermediario che ha la scelta tra due allocazioni inique viene punito di più rispetto a quando quella stessa scelta sia presa direttamente dal dittatore. Questo non avviene quando l'intermediario può soltanto iniziare il processo di selezione casuale tra le due allocazioni, ma non scegliere direttamente tra queste due allocazioni inique. I nostri risultati rafforzano e chiariscono l'utilità di utilizzare un intermediario come mezzo per deviare l'attribuzione di responsabilità. Grazie al confronto tra il caso in cui l'intermediario può scegliere rispetto a quando può solo iniziare il processo di scelta casuale, il nostro lavoro contribuisce alla crescente letteratura sulla delegazione della responsabilità.

Chapter 1

Trilateral Contracts and the Hold-up Problem

1.1 Introduction

The hold-up problem has been extensively analyzed by the economic literature in the last decades. In its classical version, this problem applies when two parties, for instance a manufacturer and a customer, or, more generally, a seller and a buyer, are unable to extract all the surplus from their interaction. Typically, the party that should make a quality enhancing relation-specific investment is unable to receive all the benefits that accrue from this investment, as future (re)negotiation may confer parts of the benefit from the customized investment to the party with higher bargaining power. When neither the investment nor the induced quality can be verified by a third party, the contract cannot be contingent on them: investment will be below the social optimum (Williamson, 1985; Hart and Moore, 1999). We consider purely a cooperative investment, i.e. an investment that generates a direct benefit to the trading partner (the buyer), offering no accompanying direct benefits to the investor (the seller) (see Che and Hausch, 1999). Precisely, in our setting investment increases the probability that the good is of high quality. A contract with a fixed price - in which the seller receives a fixed payment for the product, independent on the level of quality - would give the seller no incentive to invest. Similarly, a contract in which the buyer has the option to buy the product depending on the quality gives the seller no incentive to invest, since the seller anticipates that the buyer may renegotiate the terms of the contract once the investment is sunk (Hart and Moore, 1999).

In what follows, we present a new approach to remedy the hold-up problem, applicable in a setting where there is more than one buyer. We model a situation where a seller produces identical products for two noncompeting buyers that have a common interest in inducing the seller to make a quality enhancing investment. A trilateral contract may provide the correct incentives to lower the hold-up problem and restore efficiency.

The reason why the trilateral contract solves the hold-up problem is straightforward. If trade between the seller and each of the buyers is sequential, then it is possible to make payments contingent on verifiable previous exchanges. More specifically, before any investment is done, the three parties sign a contract which stipulates that when the quality of the good is high, the first buyer can purchase the product at a price equal to the price of the low quality product (or equal to the market price). In case the first buyer purchases the product, the second buyer has to pay a premium. Thereafter, he can also buy the product at market price. This premium thus induces the seller to invest efficiently.

Hence, even though an option contract with fixed prices is signed, the seller has sufficient incentive to invest in quality enhancement, and the induced level of investment is as high as the social optimum. The contract is self-enforcing, which means that neither party has incentive to renege (Baker, Gibbons, and Murphy, 2002). It is coalition renegotiation proof, hence not even a subgroup of agents has incentive to renege jointly (Bernheim, Peleg, and Whinston, 1987). By additionally specifying payments among the firms upon contracting, we ensure that all parties have incentive to participate in the contract.

Our trilateral contract constitutes a one shot cooperative project, in which investment is incurred a single time. In this sense, it differs from solving the problem by vertically integrating or restructuring firm boundaries and asset ownership, as suggested by Baker, Gibbons, and Murphy (2002); Grossman and Hart (1986); Klein, Crawford, and Alchian (1978). Neither does the contract rely on repeated interaction with the same agent or within a group, where incentives arise based on reputational effects (Radner, 1981; Kandori, 1992; Dixit, 2003). Also, it does not require any additional agent like an intermediary or arbitrator, cases considered in Dixit (2004) and Laffont and Martimort (1997). By contrast, in our contract, all agents participating may benefit directly from the contract; the interaction among the agents *involved in the transaction* suffices to induce efficient incentives.

There are numerous examples of situations in which this kind of contract could be of use. Since we consider no competition among the buyers, the buyers might either be end-consumers, or firms operating in different markets or industries, or companies requiring basic research for processes and products to further developing different (end-) products. Hence, our contract may apply to any such one shot scenario regarding the hold-up problem. More specific examples include software provision for firms operating in different industries, research on cells/genes used for developing diverse medical cures for different pharmaceutical companies, or research on earthquake technology used by different countries.

Despite the direct application of our model, the method of solving can also be applied to other problems. The innovative part of this paper is the transformation of a contemporaneous setting to a sequential one. This way, we create the possibility to condition actions on previous occurrences. The sequentiality of the contract does not lead to an increase in information about quality, this is observable at any time. Yet, the fact that the first buyer is buying gives the possibility to create a new, verifiable variable on which the additional payment can be conditioned on, which is perfectly correlated with quality.

Furthermore, we require different or less assumptions to be fulfilled than other papers considering contractual solutions to the hold-up problem. Like us, Rogerson (1992) considers multi-agent settings. His' major conclusion is that first-best contractual solutions to the hold-up problem exist, if each agent's investment directly affects only his own type(on the contrary, as specified above we consider cooperative investment), and he inhibits renegotiation. We do not need to inhibit renegotiation, since our contract is renegotiation proof. Also Aghion, Dewatripont, and Rey (1990) do not inhibit renegotiation. Yet, they require to establish default options in case renegotiation breaks down, and allocate bargaining power to one of the two contracting parties. We do not require either of these. Similarly Noeldeke and Schmidt (1995) differ in the assumptions they use to solve the hold-up problem. While they look at bilateral trade, in their setting the court can observe if the seller delivered or not the right good, *or* if the buyer refused to buy. By making the price difference between delivering and not delivering sufficiently high, they induce the seller to invest. In our setting, the court needs not to observe whether the seller delivered or not, or whether it was the buyer that refused to buy, since everything is incentive compatible. The only observation that needs to be verifiable is the transaction in itself, *if* the good is exchanged - this we need to observe to be able to make the extra payment contingent on it. Che and Hausch (1999) provide a contractual solution to a cooperative hold-up problem. In their setting, in a range of bargaining shares contracting does not offer the parties an advantage over ex-post negotiation. This is not true in our case - beside the fact that contracting offers an advantage even if bargaining is purely selfish, different to Che and Hausch we even manage to re-establish the first best level of investment.

The remaining part of the paper is organized as follows. Section 1.2 introduces the basic model of a trilateral contract, concentrating on the case with only two buyers. After presenting the benchmark and several verifiability issues in section 1.2.1, section 1.2.2 presents the model under complete information, taking into account the fact that the two buyers might have different valuations for the product. We show that efficiency can be restored. In section 1.3, we comment on joint renegotiation, showing that a modification of the contract is coalition renegotiation proof: also jointly parties do not have incentive to deviate. Section 1.4 considers the case of asymmetric information: when the valuation of each buyer is private information, we show that there exists a modification of the multilateral contract that induces truthful revelation and restores the optimal level of investment. The extension of the model to more than two buyers is straightforward and exposed in section 1.5; section 3.4 concludes.

1.2 Model

Consider three players: one upstream firm A and two downstream firms P_i, $i \in \{1, 2\}$. The two downstream firms are not competing with each other. A produces two goods, and chooses the level of investment $e \geq 0$; investment is costly, with $c(e)$ an increasing convex function[1]. The two goods are of the same level of quality, which may be high or low. More precisely, the probability that the goods are of high quality is $\pi(e) \in [0, 1]$, with $\pi(e)$ an increasing quasiconcave function. The monetary value of the low quality good is normalized to zero[2]. The high quality goods can be sold to the market at a price $m > 0$, each[3]. The downstream firms attach a value $v_i = \beta_i m$, $i = 1, 2$ to a high quality product, where $\beta_i \geq 1$.

[1]Introducing an additional fixed marginal cost per good does not change anything.

[2]Alternatively, one might think about the case in which the downstream firms value the low quality product ϵ, while the buyers on the market value it zero. In this case, the contract still works, as long as $\epsilon \leq m$; it is sufficient that equation (1.5) is fulfilled.

[3]The contract works also for m = 0, setting the valuations of the downstream firms equal to $v_i = \beta_i$, with $\beta_i > 0$.

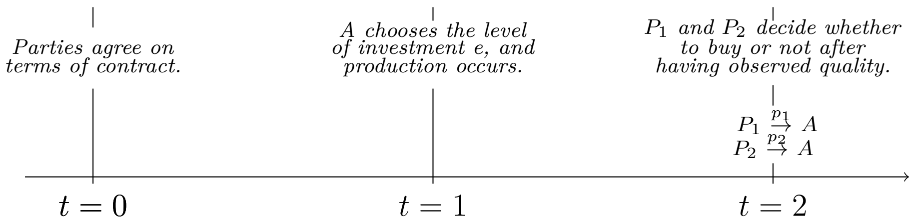


Figure 1.1: Timeline

In order to derive a closed-form solution for the model, we employ specific functional forms for the probability and cost functions. Namely, we assume that $\pi(e) = min\{\eta e, 1\}$, with $\eta > 0$, and $c(e) = \frac{\alpha}{2}e^2$, with $\alpha > 0$. Moreover, we assume that α is sufficiently high and η sufficiently small to prevent A to choose such a large investment level as to induce $\pi(e) = 1$. All players are risk neutral with standard utility functions. There are no transaction costs.

Timing is as follows (see figure 1.1): at time $t = 0$, partners decide upon the terms of the contract. After agreeing on the contract, at time $t = 1$, A chooses the level of investment, and production occurs. At $t = 2$, P_1 and P_2 decide whether to buy or not, after having observed the quality of the good.

1.2.1 Benchmark and Verifiability

To identify the efficient level of investment, we consider the case of a social planner that chooses the amount of investment e to maximize welfare:

$$\begin{aligned} & \max_e \pi(e)(\beta_1 + \beta_2)m - c(e) \\ = \; & \max_e \eta e(\beta_1 + \beta_2)m - \frac{\alpha}{2}e^2. \end{aligned} \tag{1.1}$$

From the first order condition we can easily derive the optimal level of investment

$$e^{FB} \equiv \frac{\eta}{\alpha}(\beta_1 + \beta_2)m.$$

If *investment is verifiable, but quality is not*, this first best level of investment is still obtainable. A contract which specifies $e = e^{FB}$ and a fixed price $p_i \in [\frac{1}{2}c(e^{FB}), \beta_i m\pi(e^{FB})]$ for the product, independently of the realized quality, induces an efficient outcome and guarantees to each party profits at least as big as no trading.

Also if *investment is not verifiable, but quality is*, there exist incentive compatible contracts which induce efficient outcomes. Any contract which specifies a pair of prices (p_{h_i}, p_{l_i}) such that $p_{h_i} - p_{l_i} = \beta_i m$ and lump sum transfers $\tau \in R_+$ to distribute profits induces an efficient outcome. The most intuitive case is $p_i = \beta_i m$ for the high quality product and $p_i = 0$ for the low quality product, with $\tau = 0$.

Now suppose that *neither quality nor investment is verifiable.* On the one hand, a contract that specifies a fixed price for the good independently of the realized level of quality does not provide any incentive to invest to the upstream firm A. On the other hand, a contract where each downstream firm has the option to buy the good at time $t = 1$ at a price $p_i > m$ may be subject to renegotiation at time $t = 1$. If the downstream firm refuses to buy when the quality is high, the upstream firm can sell the products to the market just at a price m. Following Hart and Moore (1999), we assume that in the renegotiation stage the downstream firms have all bargaining power. Hence, anticipating the renegotiation, A invests

$$\begin{aligned} & \max_{e|p_i=m} U_A \\ = & \max_{e|p_i=m} \pi(e)(2m) - c(e) \\ = & \max_{e|p_i=m} \eta e 2m - \frac{\alpha}{2}e^2. \end{aligned} \tag{1.2}$$

From the first order conditions we can easily derive the induced investment level

$$e^{IC} \equiv \frac{\eta}{\alpha}2m,$$

which is henceforth called the incentive compatible investment level. It is the highest level of investment that can be induced when neither quality nor investment is verifiable.

1.2.2 Multilateral Contract

We now turn to the case of non-verifiable quality and non-verifiable investment. The following section shows that signing a complex contract involving all parties can lead to a more efficient solution than negotiating independent bilateral contracts. We focus on option contracts, i.e. contracts where P_1 and P_2 have the option to buy once they have observed the quality of the good. As we have shown, P_1 and P_2 will only exert the option if it costs less than the market price, m. Yet, parties may specify additional payments contingent on the fact that the other downstream firm buys the good. This way, parties can increase the surplus the upstream firm obtains in case of having produced a high-quality good. The contract specifies the following: P_1 has the option to buy the good at price m; if P_1 buys, then P_2 is required to pay ρ to A; once P_1 has taken its decision, P_2 has the option to buy the good at price m. The timing is summarized in figure 1.2 and 1.3.

When the goods are of high quality, then both P_1 and P_2 exert the option and firm A obtains a payoff of $(2m + \rho)$. By fixing $\rho = (\beta_1 + \beta_2 - 2)m$, the overall payment A obtains for two high quality goods is equal to $(\beta_1 + \beta_2)m$, the social optimum, and therefore A has incentives to choose the efficient level of investment. Given that P_2 is required to pay ρ contingent on P_1 having bought the good, P_2 needs to be compensated to fulfill the participation constraints. We denote x_0 and x_1 the payments that A and P_1 make in favor of P_2 upon signing the contract[4].

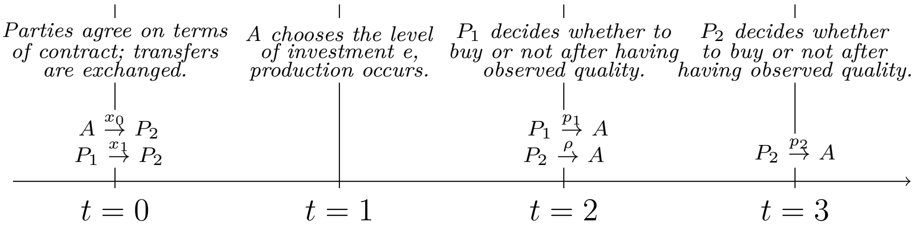


Figure 1.2: Timeline Trilateral Contract

While it may seem that some of the payments would cancel out -A pays x_0 to P_2, and P_2 pays ρ and m to A- this is not the case, seeing as the payments are conditional upon different events. The timing of the game is crucial: only by making payments conditional on previous payments is incentive compatibility assured. Note that the reason why sequential purchases work is not because it provides additional information -the firms observe the level of quality at all times- but because it generates a observable variable (i.e. whether P_1 bought the good or not), which is perfectly correlated with quality.

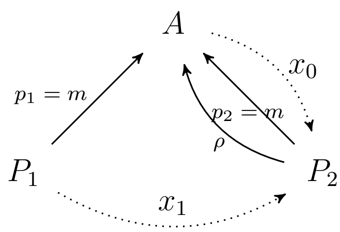


Figure 1.3: Trilateral Contract

[4]It can be shown that the contract works without monetary transfers from the upstream firm A, as well as without any transfers at all exchanged unconditional on the quality of the product. Yet, the range of valuations of the downstream firms that lead to a higher than the incentive compatible level of investment is smaller than in case the transfers are paid. The reason is that the unconditional transfers serve to relax the participation constraint of the downstream firm P_2. If A does not provide this transfer, the amount x_1 that P_2 receives is limited by the participation constraint of P_1.

If instead there are no unconditional transfers exchanged at all, the participation constraint of P_2 is even more binding.

Proposition 1 *The trilateral contract* $\{p_1 = p_2 = m, \rho = (\beta_1 + \beta_2 - 2)m,\ x_0 = \frac{\eta^2}{\alpha} 2m^2(\beta_1 + \beta_2 - 2)$ *and* $x_1 = \frac{\eta^2}{\alpha} m^2(\beta_1 - 1)(\beta_1 + \beta_2 - 2)\}$ *is self-enforcing and induces the optimal level of investment.*

Proof Suppose the contract has been signed. Knowing that A gets the payments $p_1 = m$, $p_2 = m$ and ρ in case the product is of high quality, and since the payment x_0 is paid *before* the level of investment is chosen[5], A maximizes

$$\begin{aligned} & \max_{e \left| \left\{ \substack{p_i = m, \forall i \in \{1,2\} \\ \rho} \right\} \right.} U_A \\ = & \max_{e \left| \left\{ \substack{p_i = m, \forall i \in \{1,2\} \\ \rho} \right\} \right.} \pi(e)(\sum_{i=1}^{2} p_i + \rho) - c(e) \\ = & \max_{e \left| \left\{ \substack{p_i = m, \forall i \in \{1,2\} \\ \rho} \right\} \right.} \eta e(2m + \rho) - \frac{\alpha}{2} e^2. \end{aligned} \tag{1.3}$$

From the first order conditions follows that

$$\tilde{e} \equiv \frac{\eta}{\alpha}(2m + \rho)$$

is the optimal level of investment for A, given the contract has been signed. It is strictly increasing in ρ. For any $\rho > 0$, $\tilde{e}$ is greater than the incentive compatible level of investment e^{ic}; for $\rho = (\beta_1 + \beta_2 - 2)m$, $\tilde{e}$ equals the optimal level of investment e^{FB}.

We now study the conditions under which this contract is renegotiation-proof. Suppose the quality of the good is high. Both downstream firms have the possibility to refuse buying from the upstream firm, which is then forced to sell the products to the market at a price m. Since we assumed zero transaction costs, P_1 and P_2 can then buy at a price m. Once the quality is known, P_i will not renegotiate if

$$\beta_i m - p_i \geq \beta_i m - m \quad \forall i \in \{1, 2\}. \tag{1.4}$$

It will not buy a good of low quality if

$$-p_i \leq 0 \quad \forall i \in \{1, 2\}. \tag{1.5}$$

[5]Even if x_0 was paid *after* choosing the level of investment, A would incur the same level of investment as specified below, as long as x_0 and ρ are specified as in (1.3).

Both inequalities are clearly satisfied.

Is the contract individual rational? If the contract is not signed, A will invest e^{ic}, and the downstream firms pay m at the market. The expected value for the downstream firms computes to $\pi(e^{ic})(\beta_1 m - m)$. Therefore, having to pay x_1 upon signing the contract, P_1 is willing to participate in the contract if

$$-x_1 + \pi(\tilde{e})(\beta_1 m - p_1) \geq \pi(e^{ic})(\beta_1 m - m); \tag{1.6}$$

and P_2 will sign the contract if

$$x_0 + x_1 + \pi(\tilde{e})(\beta_2 m - p_2) - \pi(\tilde{e})\rho \geq \pi(e^{ic})(\beta_2 m - m). \tag{1.7}$$

A partakes when

$$\pi(\tilde{e})(2m + \rho) - c(\tilde{e}) - x_0 \geq \pi(e^{ic})(2m) - c(e^{ic}). \tag{1.8}$$

Assume the participation constraints (1.6) and (1.7) of the downstream firms to be binding. Replacing the resulting $x_0 = \pi(e^{ic})(\beta_1 + \beta_2 - 2)m - \pi(\tilde{e})(\beta_1 m + \beta_2 m - 2m - \rho)$, $\pi(\cdot)$, $c(\cdot)$ and the levels of investment e^{ic} and $\tilde{e}$ and choosing $\rho = (\beta_1 + \beta_2 - 2)m$, the participation constraints of A, P_1 and P_2 are fulfilled. The resulting x's are:

$$\begin{aligned} x_0 &= \frac{\eta^2}{\alpha} 2m^2(\beta_1 + \beta_2 - 2), \text{ and} \\ x_1 &= \frac{\eta^2}{\alpha} m^2(\beta_1 - 1)(\beta_1 + \beta_2 - 2). \square \end{aligned}$$

The trilateral contract that induces the efficient level of investment is not unique. However, they will all fix $p_1 = m, p_2 = m$ and $\rho = (\beta_1 + \beta_2 - 2)m$, and differ just in the payments at time $t = 0$, the x's. The contract that is specified above is such that the extra profits - compared to the incentive compatible case - are completely skimmed by the upstream firm.

Which of the two downstream firms P_1 and P_2 buys first is decided randomly. Since the expected payoff of the respective downstream firm is equal to the incentive compatible payoff they are indifferent to being the first or the second buyer[6].

[6]This makes sense as long as we assume that both transactions occur in a relatively short period of time. Obviously, it is also possible to change unconditional transfers such that the second downstream firm captures a higher share of the profit generated. Similarly, there exists a symmetric case in which both firms are paying and receiving the exact same transfers; see section 1.D in the appendix.

1.3 Joint Deviation Under Complete Information

In the previous section we have shown that neither of the two downstream firms has incentive to renegotiate the contract. Yet, we have not ruled out the possibility that a subset of the participants coordinate its actions in a mutually beneficial way. Ignoring collusion would be an important oversight, especially since the participants can communicate at any stage of the contract.

For example, after signing the contract, P_1 and P_2 might agree not to exert the option and buy the good from the market at a price m. This would result in a joint payment for the two units of the good for P_1 and P_2 of $2m$ rather than $(2m + \rho)$. Hence, P_1 and P_2 have an incentive to deviate from the specified contract. In a similar vein, P_1 and A might want to defraud ρ from P_2 even in case the good is of low quality. In either case, to make deviation incentive compatible, parties need to agree upon the exchange of side payments. By including two additional clauses in the contract, the agents cannot credibly commit to fulfill the specifications of side-agreements. This way, we can make these deviations infeasible.

The two additional clauses we specify are the following. a) a clause inhibiting participating firms from making side contracts *conditional on the asserted quality*; and b) a clause specifying the exchange of a payment $0 < m_c < m$, payed from the first downstream firm P_1 to the second downstream firm P_2, in case P_1 claims low quality and P_2 claims high quality[7].

A contract is defined as coalition proof if it induces a Coalition-Proof Nash equilibrium. Being a Coalition-Proof Nash equilibrium means that no subcoalition of the agents taking part in the contract has incentive to deviate from the specified equilibrium (Bernheim, Peleg, and Whinston, 1987). Making use of the fact that a deviation is not self-enforcing, we show that

Proposition 2 *There exists a contract that is coalition deviation proof.*

There are three possible coalitions: $\{A, P_1\}$, $\{A, P_2\}$, and $\{P_1, P_2\}$. The coalition $\{A, P_2\}$ cannot gain anything by jointly deviating. Since the exchange of the payment ρ depends only on what the first downstream firm, P_1, reports, any possible joint deviation wanting to extract this amount has to include P_1. It remains to show that neither $\{A, P_1\}$ nor $\{P_1, P_2\}$ have incentive to deviate.

Proof See appendix, section 1.A. □

[7]Also with this clause the incentive compatibility constraint for P_2 when the good is of low quality is still satisfied: P_2 does not have incentive to buy a low quality product (claiming it to be high quality), since$-m + m_c < 0$.

The intuition behind the proof is the following. The coalition $\{A, P_1\}$ cannot gain anything in case the good is of high quality: when claiming low quality, P_1 would reduce the overall amount their coalition receives by ρ. If the good is of low quality, on the other hand, P_1 can increase the overall amount its coalition receives by ρ, claiming it to be of high quality. A and P_1 may agree upon a payment $\epsilon_{01} \in (m, \rho)$ as compensation for P_1 reporting falsely in such case. Yet, since the two parties are not allowed to make side contracts, the exchange of this payment is not incentive compatible. If ϵ_{01} is exchanged before P_1 reports the quality, P_1 does not have incentive to report falsely; if ϵ_{01} is supposed to be exchanged after P_1 has reported, A does not have incentive to pay, which in turn P_1 will anticipate.

The coalition $\{P_1, P_2\}$ cannot gain anything when the good is of low quality, since by falsely claiming high quality they have to pay for products their valuation is actually zero. If however the good is of high quality, by falsely claiming low quality, P_1 and P_2 might gain ρ. In this case, both clauses are needed to prevent a profitable joint deviation. Assume P_1 and P_2 agree upon a payment ϵ_{21} when P_1 claims low quality. If ϵ_{21} is exchanged *before* P_1 reports the quality, P_1 does not have incentive to falsely claim low quality. If on the other hand ϵ_{21} is supposed to be exchanged *after* P_1 claims the good to be of low quality, then P_2 is strictly better of by not paying it and reporting high quality. Anticipating this, P_1 will claim the true quality.

In conclusion, appending the subclauses above to the contract will prevent any coalition of the firms from reneging on the contract or falsely reporting the quality of the good.

1.4 Asymmetric Information

Up to now we considered the value the downstream firms attach to a high quality good, β_i to be common knowledge. But does the contract hold as well under information asymmetries? In this section we show that the contract specified in section 1.2.2 can be extended to situations in which the valuation of the good to each buyer is private information.

Again, we consider the case of two downstream firms. Assume that, before the contract is made, each downstream firm privately observes its type $\beta^k, k \in \{H, L\}$, where, as before, $\beta^H \geq \beta^L \geq 1$. The types β^H and β^L are identically and independently distributed, with $\Pr\{\beta_i = \beta^H\} = p \in [0, 1]$, the distribution being common knowledge. Let $\hat{\beta}_i$ be the reported type.

Proposition 3 *By specifying the extra-payment ρ and the transfers conditional on the reported values $(\hat{\beta}_1, \hat{\beta}_2)$, there exists a contract that induces truthful revelation and the optimal level of investment.*

Proof See section 1.B in the appendix. □

The intuition behind the proof is the following. The parameters $\{\rho(\hat{\beta}_1, \hat{\beta}_2), x_0(\hat{\beta}_1, \hat{\beta}_2), x_1(\hat{\beta}_1, \hat{\beta}_2)\}$ can be specified for each possible state - both firms reporting high valuations, both firms reporting low, as well as the two cases when they report differently. With this set we can show that truthful revelation holds in dominant strategies: neither of the downstream firms has incentive to misreport its type, independently of being of high or low type, independently on what the other downstream firm reports. Assuming that each firm can decide whether to participate or not *after* each downstream firm has revealed its type, we show that the ex-post participation constraints are fulfilled[8] all agents have incentive to participate in the contract.

Unlike the case of complete information, the downstream firms now capture some of the payoff when being of high type, even when transfers are specified as to maximize the upstream firm A's payoff. This can be explained by the existence of informational rents for the downstream firms. If they are of the low type , their participation constraints are binding, while when they are of the high type, their constraints on truthful reporting are. Since the x_1 transferred in the case of reporting (β^L, β^H) is already the biggest P_1 can provide (the participation constraint is binding), to keep truthful reporting a (weakly) dominant strategy, also the x_1 exchanged in case of reporting (β^H, β^H) cannot be decreased. This results in informational rents for the downstream firms. A similar reasoning holds for the x_2's received by P_2.

Again, as in the case of complete information, the payoffs for the downstream firms do not depend on the order in which they are placed. Whether they are in the first or the second downstream firm, each firm receives the same amount depending on its type.

1.5 More Than Two Downstream Firms

When extending the model to more than two downstream firms, several modifications to the trilateral contract come to mind. We work out one possible non-symmetric case more than two downstream firms, assuming complete information. For a symmetric case, see section 1.D in the appendix. We show that the optimal level of investment can still be induced.

The setting is very similar to the one considered in the previous sections. There are n downstream firms, each buying the product at a price equal to the market price $p_i = m$ in case the quality is high. Conditional on the first downstream

[8]Since as ex-post constraints are stronger than interim and ex-ante participation constraints, the latter two will also be fulfilled.

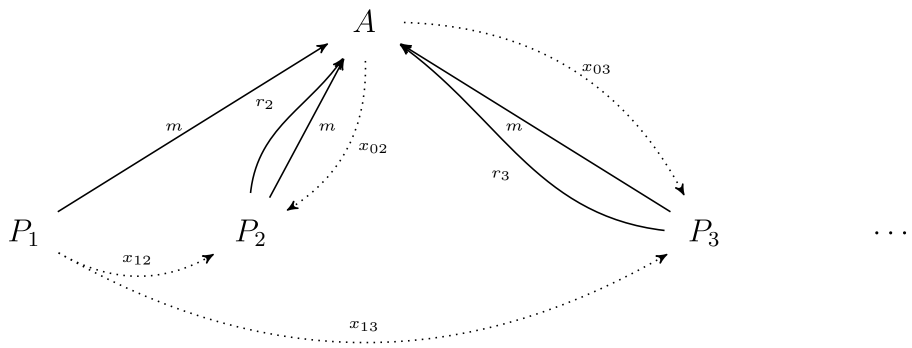


Figure 1.4: Multilateral Contract with More than Two Downstream Firms

firm, P_1, buying, the other downstream firms, P_i $i \in \{2, \ldots, k\}$, pay a share $r_i = \beta_i \frac{\sum_{i=1}^{k} \beta_i m - km}{\sum_{i=1}^{k} \beta_i}(1 + \frac{\beta_1}{\sum_{i=2}^{k} \beta_i})$ to the upstream firm.

Proposition 4 *A modification of the trilateral contract holds for $n \geq 2$ downstream firms. It is self-enforcing and induces the optimal level of investment.*

Proof See section 1.C in the appendix. □

The intuition of the proof follows the reasoning for the trilateral contract presented in section 1.2. When there are more than two downstream firms involved in the contract, one might rethink the distribution of bargaining power: now it seems plausible that the downstream firms receive a higher share of the generated profit. Yet, as mentioned above, through changes in the unconditional transfers, profit can be easily distributed in a different way.

1.6 Conclusion

In extending the literature on the hold-up problem, we have shown that such dilemma can be solved when there is more than one buyer involved in the transaction. By introducing sequentiality, we create the possibility to make transfers conditional on observed payments, thereby restoring efficiency. The result holds both under complete and under asymmetric information, and in the latter case induces truthful revelation of types. The contract is coalition renegotiation proof and extendable to more than two downstream firms.

The trilateral contract can induce the first best level of investment and satisfies the participation constraints of all agents. Yet, it is not the unique possible implementation of the contract. What is crucial is the exchange of payments *conditional* on another party buying the product. Depending on how the bargaining power is distributed, the surplus generated by the trilateral contract may be divided differently among the upstream firm and the downstream firms. While in the base-model we assumed the upstream firm to capture all the extra surplus, it is also possible to specify the transfers such that the surplus is divided differently. This might be of particular importance when considering more than two downstream firms.

Up to now, we have assumed that the high quality products can be sold at a price $m > 0$ to the market. Relaxing this assumption, the contract has to be changed slightly. If $m = 0$, we cannot condition the exchange of the extra payment upon the exchange of payments anymore. As before, the first downstream firm will only want to buy the good when it is of high quality. Yet, now it will not have to pay anything to get it. Hence, in this case the contract has to specify the exchange of an extra payment conditional on the *exchange of the product*. Apart from this modification, the reasoning on why the contract works stays the same.

Depending on the real cost-function of the investor, the gains from such a contract may be considerable. It is easily extendable to more than two downstream firms. For example, it could be used by the biggest players in the pharmaceutical industry to jointly invest in fundamental research, developing necessary basic processes and products to further develop different medicines. An interesting extension to our model would be to see what happens when we introduce competition among the downstream firms.

References

Aghion, P., M. Dewatripont, and P. Rey (1990): "On renegotiation design," *European Economic Review*, 34(2-3), 322–329.

Baker, G., R. Gibbons, and K. J. Murphy (2002): "Relational Contracts And The Theory Of The Firm," *The Quarterly Journal of Economics*, 117(1), 39–84.

Bernheim, B., B. Peleg, and M. D. Whinston (1987): "Coalition-Proof Equilibria I. Concepts," *Journal of Economic Theory*, 42, 1–12.

Che, Y.-K., and D. B. Hausch (1999): "Cooperative Investments and the Value of Contracting," *American Economic Review*, 89(1), 125–147.

Dixit, A. (2003): "On Modes of Economic Governance," *Econometrica*, 71(2), 449–481.

Dixit, A. K. (2004): *Lawlessness and Economics: Alternative Modes of Governance*. Princeton University Press.

Grossman, S. J., and O. D. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94(4), 691–719.

Hart, O., and J. Moore (1999): "Foundations of Incomplete Contracts," *The Review of Economic Studies*, 66(1), 115–138.

Kandori, M. (1992): "Social Norms and Community Enforcement," *Review of Economic Studies*, 59(1), 63–80.

Klein, B., R. G. Crawford, and A. A. Alchian (1978): "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21(2), 297–326.

Laffont, J.-J., and D. Martimort (1997): "Collusion Under Asymmetric Information," *Econometrica*, 65(4), 875–911.

Noeldeke, G., and K. M. Schmidt (1995): "Option Contracts and Renegotiation: A Solution to the Hold-up Problem," *The RAND Journal of Economics*, 26(2), pp. 163–179.

Radner, R. (1981): "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship," *Econometrica*, 49(5), 1127–1148.

Rogerson, W. P. (1992): "Contractual Solutions to the Hold-Up Problem," *Review of Economic Studies*, 59(4), 777–93.

Williamson, O. E. (1985): *The Economic Institutions of Capitalism.* New York: Free Press.

Appendix “Trilateral Contracts and the Hold-up Problem”

1.A Proof of Proposition 2

Proof We will consider the possible deviations in turn.

a) Coalition $\{A, P_1\}$

Suppose the good is of high quality. A and P_1 cannot benefit when P_1 falsely reports low quality, since this reduces the amount the upstream firm A receives by ρ.

Now assume the good being of low quality. In this case, A and P_1 may benefit when P_1 announces it to be of good quality instead. While P_1 then pays $p_1 = m$ for the product that has no monetary value, A receives $\rho = (\beta_1 + \beta_2 - 2)m$ from P_2. In case $\beta_1 + \beta_2 \geq 3$, A may promise P_1 a payment of $\epsilon_{01} \in (m, \rho)$ for claiming the good to be of high quality, when instead it is low quality. This payment ϵ_{01} may be agreed upon after the quality is realized[9].

ϵ_{01} can be exchanged either before or after the downstream firm P_1 claims high quality. If A and P_1 agree upon exchanging it *after* P_1 has claimed high quality, since the payment is not legally enforceable, A is strictly better of by not paying ϵ_{01}, since

$$p_1 + \rho > p_1 + \rho - \epsilon_{01}.$$

Hence, P_1 wants to receive ϵ_{01} *before* claiming high quality. However, once P_1 has received ϵ_{01}, P_1 is strictly better of by not respecting the side-agreement with A, since

$$\epsilon_{01} - p_1 < \epsilon_{01}.$$

Therefore, this side-agreement is not deviation-proof, and hence not self-enforcing.

b) Coalition $\{P_2, P_1\}$

Assume the good is of low quality, or in other words had no monetary value for the two downstream firms. This means that they do not have incentive to jointly

[9] Before the quality is realized, the upstream firm would prefer not to agree to any side-payments, since it would gain strictly less when the quality is high after all. In any case, the reasoning also holds if the payment is agreed upon before quality is realized.

claim high quality, as they would have to pay an amount greater zero for receiving the good. Hence, at least one of the two firms is strictly worse of compared to reporting truthfully[10].

Now suppose the product is of high quality. When telling the truth, the downstream firms pay $(2m + \rho)$ for two high quality products. If instead they deviate by agreeing on P_1 claiming low quality, they only pay $2m$ to the upstream firm. The "profit" ρ can be shared in such a way that both parties are strictly better of, specifying shares $\{\epsilon_{21}, \rho - \epsilon_{21}\}$ for P_1 and P_2 respectively, with $\epsilon_{21} \in (0, \rho)$. ϵ_{21} may be exchanged before or after P_1 claims low quality. If ϵ_{21} is exchanged *after* P_1 has claimed low quality, since the side-agreement is not legally enforceable, P_2 is strictly better of by deviating and not paying ϵ_{21}, since

$$\beta_2 m - p_2 + m_c > \beta_2 m - p_2 + m_c - \epsilon_{21}.$$

Now assume ϵ_{21} is exchanged *before* P_1 reports low quality[11]. Recall that m_c has to be paid by P_1 in case P_2 reports high quality *after* P_1 *has reported low quality.* Then, if P_1 has claimed the good to be of low quality, given that the quality is high, P_2 strictly prefers to claim high quality: since it already has paid ϵ_{21}, it can gain m_c when reporting a different level of quality than P_1. Hence, reporting high quality is preferred:

$$\beta_2 m - p_1 - \epsilon_{21} + m_c > \beta_2 m - m - \epsilon_{21}.$$

Anticipating this, P_1 will not claim low quality in the first place. He has already received ϵ_{21}, and reporting truthfully high quality, he does not have to pay m_c:

$$\beta_1 m - p_2 + \epsilon_{21} > \beta_1 m - m + \epsilon_{21} - m_c.$$

Therefore, also this deviation is not self-enforcing. It follows that our contract is coalition deviation proof.□

1.B Proof of Proposition 3

Proof Incentive compatibility in (weakly) dominant strategies requires that there exists a strategy $\hat{\beta}_i = \beta_i^k$, $\forall\, i \in \{1, 2\}$, such that

$$U_i(\hat{\beta}_i, \hat{\beta}_{-i}|\beta_i) \;\; \geq U_i(\hat{\beta}'_i, \beta_{-i}|\beta_i) \;,\; \forall \hat{\beta}_i \text{ and all } \hat{\beta}'_i. \tag{1.9}$$

To find an equilibrium in dominant strategies, we need to specify the three functions $\{\rho(\hat{\beta}_1, \hat{\beta}_2),\, x_0(\hat{\beta}_1, \hat{\beta}_2), x_1(\hat{\beta}_1, \hat{\beta}_2)\}$ such that condition (1.9) is fulfilled. In

[10] The two downstream firms can both claim the good to being of high quality, in which case the first firm pays $-p_1 < 0$ and the second firm pays overall $-\rho - p_2 < 0$. Alternatively, they could agree upon P_1 claiming low and P_2 claiming high quality, in which case P_2 is strictly worse of ($-p_2 < 0$) and P_1 is indifferent; or they can agree upon P_1 claiming high and P_2 claiming low quality, in which both P_1 and P_2 are worse of, with $-p_1 < 0$ and $-\rho < 0$ respectively.

[11] W.r.t. specifying the exchanged payment before or after quality is realized, the same reasoning holds as with the previous coalition.

addition, the participation constraints of the downstream firms (equations (1.24) - (1.31)) and of the upstream firm (equations (1.32) - (1.35)) have to be satisfied. Define $x_2(\hat{\beta}_1, \hat{\beta}_2) \equiv x_0(\hat{\beta}_1, \hat{\beta}_2) + x_1(\hat{\beta}_1, \hat{\beta}_2)$ and to simplify notation, replace $e(\hat{\beta}_H, \hat{\beta}_H)$ by e^{HH}, and similarly for e^{HL}, e^{LH}, and e^{LL}. Specify the x's as follows[12]:

$$\begin{aligned}
\tilde{x}_1(\hat{\beta}_1^H, \hat{\beta}_2^H) &= \pi(e^{HH})(\beta^H - 1)m - \pi(e^{LH})(\beta^H - \beta^L)m \\
&\quad -\pi(e^{ic})(\beta^L - 1)m, && (1.10)\\
\tilde{x}_2(\hat{\beta}_1^H, \hat{\beta}_2^H) &= \pi(e^{ic})(\beta^L - 1)m - \pi(e^{HH})(\beta^H m - m - \rho^{HH}) \\
&\quad +\pi(e^{HL})(\beta^H - \beta^L)m, && (1.11)\\
\tilde{x}_1(\hat{\beta}_1^L, \hat{\beta}_2^H) &= [\pi(e^{LH}) - \pi(e^{ic})](\beta^L - 1)m, && (1.12)\\
\tilde{x}_2(\hat{\beta}_1^L, \hat{\beta}_2^H) &= \pi(e^{ic})(\beta^L - 1)m - \pi(e^{LH})(\beta^H m - m - \rho^{LH}) \\
&\quad +\pi(e^{LL})(\beta^H - \beta^L)m, && (1.13)\\
\tilde{x}_1(\hat{\beta}_1^H, \hat{\beta}_2^L) &= \pi(e^{HL})(\beta^H - 1)m - \pi(e^{LL})(\beta^H - \beta^L)m \\
&\quad -\pi(e^{ic})(\beta^L - 1)m, && (1.14)\\
\tilde{x}_2(\hat{\beta}_1^H, \hat{\beta}_2^L) &= \pi(e^{ic})(\beta^L - 1)m - \pi(e^{HL})(\beta^L m - m - \rho^{HL}), && (1.15)\\
\tilde{x}_1(\hat{\beta}_1^L, \hat{\beta}_2^L) &= [\pi(e^{LL}) - \pi(e^{ic})](\beta^L - 1)m, && (1.16)\\
\tilde{x}_2(\hat{\beta}_1^L, \hat{\beta}_2^L) &= \pi(e^{ic})(\beta^L - 1)m - \pi(e^{LL})(\beta^L m - m - \rho^{LL}). && (1.17)
\end{aligned}$$

First, we show that firms will prefer to report their evaluations truthfully, after which we show that the participation constraints of the downstream firms are satisfied and that the participation constraint of the upstream firm is fulfilled.

1.B.1 Truthful Revelation

The payoffs of the downstream firms are respectively:

$$\begin{aligned}
U_1(\hat{\beta}_1, \hat{\beta}_2|\beta_1) &= \pi(\tilde{e}(\hat{\beta}_1, \hat{\beta}_2))\,[\beta_1 m - p_1] - x_1(\hat{\beta}_1, \hat{\beta}_2), && (1.18)\\
U_2(\hat{\beta}_2, \hat{\beta}_1|\beta_2) &= \pi(\tilde{e}(\hat{\beta}_1, \hat{\beta}_2))\left[\beta_2 m - p_2 - \rho(\hat{\beta}_1, \hat{\beta}_2)\right] && (1.19)\\
&\quad +x_0(\hat{\beta}_1, \hat{\beta}_2) + x_1(\hat{\beta}_1, \hat{\beta}_2).
\end{aligned}$$

Set $p_1 = p_2 = m$, and consider the downstream firms in turn.

a) Downstream firm P_1

We show that P_1 has no incentive to misreport its type, regardless of being of high or low type, and independently on what the second downstream firm P_2 reports.

[12]The $x's$ are chosen according to the following strategy: to minimize the amount A has to provide, we choose x_1 such that P_1 pays the biggest amount possible satisfying its incentive compatibility and participation constraints. Similarly, x_2 is chosen such that P_2 receives the smallest amount possible such that its incentive compatibility and participation constraints are fulfilled.

Suppose P_2 reports β^H, and suppose P_1 *is* of type β^H. Having specified the x's as above, when truthfully *reporting* being of type β^H, P_1 receives a payoff of

$$\pi(e^{LH})\left[\beta^H - \beta^L\right]m + \pi(e^{ic})\left[\beta^L - 1\right]m;$$

when reporting to be of type β^L, it receives

$$\pi(e^{LH})\left[\beta^H - \beta^L\right]m + \pi(e^{ic})\left[\beta^L - 1\right]m.$$

Seeing as both payoffs are equal, reporting the true type weakly dominates non-truthful reporting. Now suppose the dowstream firm P_1 is of type β^L, while the downstream firm P_2 still reports being of high type. When P_1 reports to be of type β^H, it gets

$$\left[\pi(e^{LH}) - \pi(e^{HH})\right]\left[\beta^H - \beta^L\right]m + \pi(e^{ic})\left[\beta^L - 1\right]m, \tag{1.20}$$

while when truthfully reporting β^L, it receives

$$\pi(e^{ic})\left[\beta^L - 1\right]m. \tag{1.21}$$

x_0 is exchanged *before* the level of investment is incurred, hence, it can be shown that the payoff in (1.20) is smaller than the payoff in (1.21), since

$$\left[\pi(e^{LH}) - \pi(e^{HH})\right]\left[\beta^H m - \beta^L m\right] \leq 0.$$

Let us turn to the level of investment. Knowing that for a high quality good A receives an overall payment of $2m + \rho(\hat{\beta}_1, \hat{\beta}_2)$, it will try to maximize:

$$\begin{aligned}
& \max_{e|\left\{\substack{p_i = m, \forall i \in \{1,2\} \\ (\hat{\beta}_1, \hat{\beta}_2)}\right\}} U_A \\
= & \max_{e|\left\{\substack{p_i = m, \forall i \in \{1,2\} \\ (\hat{\beta}_1, \hat{\beta}_2)}\right\}} \pi(e(\hat{\beta}_1, \hat{\beta}_2))[\sum_{i=1}^{2} p_i + \rho(\hat{\beta}_1, \hat{\beta}_2)] - c(e(\hat{\beta}_1, \hat{\beta}_2)) \\
= & \max_{e|\left\{\substack{p_i = m, \forall i \in \{1,2\} \\ (\hat{\beta}_1, \hat{\beta}_2)}\right\}} \eta e(\hat{\beta}_1, \hat{\beta}_2)[2m + \rho(\hat{\beta}_1, \hat{\beta}_2)] - \frac{\alpha e(\hat{\beta}_1, \hat{\beta}_2)^2}{2}.
\end{aligned}$$

This results in $e(\hat{\beta}_1, \hat{\beta}_2) = \frac{\eta}{\alpha}(2m + \rho(\hat{\beta}_1, \hat{\beta}_2))$. Setting $\rho(\hat{\beta}_1, \hat{\beta}_2) = (\hat{\beta}_1 + \hat{\beta}_2 - 2)m$ will once more induce the optimal level of investment. Then, $e(\hat{\beta}_1, \hat{\beta}_2) = \frac{\eta}{\alpha}[\hat{\beta}_1 + \hat{\beta}_2]m$, and therefore, for $\beta^H \geq \beta^L \geq 1$, $e^{HH} \geq e^{HL}$. Hence,
$\left[\pi(e^{LH}) - \pi(e^{HH})\right]\left[\beta^H - \beta^L\right]m \leq 0$.
Now suppose the downstream firm P_2 reports being of type β^L. Suppose P_1 is of type β^H. When reporting being of type β^H, it receives a payoff of

$$\pi(e^{LL})\left[\beta^H - \beta^L\right]m + \pi(e^{ic})\left[\beta^L - 1\right]m,$$

while when reporting being of type β^L, it receives

$$\pi(e^{LL})\left[\beta^H - \beta^L\right] + \pi(e^{ic})\left[\beta^L - 1\right]m.$$

Again, the two payoffs are equal, and reporting the true type weakly dominates non-truthfully reporting. Now suppose P_1 is of type β^L, with P_2 still reporting being of type β^L. When reporting being of type β^H, P_1 receives a payoff of

$$\left[\pi(e^{LL}) - \pi(e^{HL})\right]\left[\beta^H - \beta^L\right] m + \pi(e^{ic})\left[\beta^L - 1\right] m,$$

while when truthfully reporting β^L, it receives

$$\pi(e^{ic})\left[\beta^L - 1\right] m.$$

Applying the reasoning above and taking into account that $e^{HH} \geq e^{LH}$, reporting the truth dominates non-truthful reporting. Since, in expected terms, final payoffs of the downstream firms are equal - given that they are of the same type - a symmetric reasoning holds for the truthful reporting of P_2.

In conclusion, with the $x's$ specified as in (1.10) - (1.17), the downstream firms have incentive to truthfully reveal their types. But do they also want to join the trilateral contract? In the following sectino, we show that the participation constraints are fulfilled.

1.B.2 Participation Constraints

a) Participation Constraints of Downstream Firms

The participation constraints of P_1 and P_2 are respectively:

$$\begin{aligned} \pi(\tilde{e}(\hat{\beta}_1, \hat{\beta}_2))(\beta_1 m - p_1) - x_1(\hat{\beta}_1, \hat{\beta}_2) &\geq \pi(e^{ic})(\beta_1 m - m), \\ \pi(\tilde{e}(\hat{\beta}_1, \hat{\beta}_2))[\beta_2 m - p_2 - \rho(\hat{\beta}_1, \hat{\beta}_2)] + x_2(\hat{\beta}_1, \hat{\beta}_2) &\geq \pi(e^{ic})(\beta_2 m - m), \end{aligned}$$

which is

$$\begin{aligned} x_1(\hat{\beta}_1, \hat{\beta}_2) &\leq \pi(\tilde{e}(\hat{\beta}_1, \hat{\beta}_2))(\beta_1 m - m) - \pi(e^{ic})(\beta_1 m - m), && (1.22) \\ x_2(\hat{\beta}_1, \hat{\beta}_2) &\geq \pi(e^{ic})(\beta_2 m - m) && (1.23) \\ &\quad -\pi(\tilde{e}(\hat{\beta}_1, \hat{\beta}_2))[\beta_2 m - m - \rho(\hat{\beta}_1, \hat{\beta}_2)]. \end{aligned}$$

Assuming truthful reporting, for the respective values of $\hat{\beta}_i$ and β_i, 1.22 and 1.23 become

$$\begin{aligned} x_1(\hat{\beta}_1^H, \hat{\beta}_2^H) &\leq [\pi(e^{HH}) - \pi(e^{ic})](\beta^H m - m), && (1.24) \\ x_2(\hat{\beta}_1^H, \hat{\beta}_2^H) &\geq \pi(e^{ic})(\beta^H m - m) - \pi(e^{HH})(\beta^H m - m - \rho^{HH}), && (1.25) \\ x_1(\hat{\beta}_1^L, \hat{\beta}_2^H) &\leq [\pi(e^{LH}) - \pi(e^{ic})](\beta^L m - m), && (1.26) \\ x_2(\hat{\beta}_1^L, \hat{\beta}_2^H) &\geq \pi(e^{ic})(\beta^H m - m) - \pi(e^{LH})(\beta^H m - m - \rho^{LH}), && (1.27) \\ x_1(\hat{\beta}_1^H, \hat{\beta}_2^L) &\leq [\pi(e^{HL}) - \pi(e^{ic})](\beta^H m - m), && (1.28) \\ x_2(\hat{\beta}_1^H, \hat{\beta}_2^L) &\geq \pi(e^{ic})(\beta^L m - m) - \pi(e^{HL})(\beta^L m - m - \rho^{HL}), && (1.29) \\ x_1(\hat{\beta}_1^L, \hat{\beta}_2^L) &\leq [\pi(e^{LL}) - \pi(e^{ic})](\beta^L m - m), && (1.30) \\ x_2(\hat{\beta}_1^L, \hat{\beta}_2^L) &\geq \pi(e^{ic})(\beta^L m - m) - \pi(e^{LL})(\beta^L m - m - \rho^{LL}). && (1.31) \end{aligned}$$

By substituting $\tilde{x}_1(\hat{\beta}_1, \hat{\beta}_2)$ and $\tilde{x}_2(\hat{\beta}_1, \hat{\beta}_2)$ in the participation constraints (1.24) - (1.31), it can be seen that equations (1.26), (1.29), (1.30), and (1.31) are binding. Equations (1.24), (1.25), (1.27), and (1.28) can be simplified to, respectively

$$\begin{aligned}
(\pi(e^{LH}) - \pi(e^{ic}))(\beta^H m - \beta^L m) &\geq 0, \\
(\pi(e^{HL}) - \pi(e^{ic}))(\beta^H m - \beta^L m) &\geq 0, \\
(\pi(e^{LL}) - \pi(e^{ic}))(\beta^H m - \beta^L m) &\geq 0, \\
(\pi(e^{LL}) - \pi(e^{ic}))(\beta^H m - \beta^L m) &\geq 0,
\end{aligned}$$

which are all clearly satisfied for $\beta^H \geq \beta^L \geq 1$. So, the specified $\tilde{x}$'s also satisfy the participation constraints of P_1 and P_2 .

It remains to check if A can provide $x_0(\hat{\beta}_1, \hat{\beta}_2) = x_2(\hat{\beta}_1, \hat{\beta}_2) - x_1(\hat{\beta}_1, \hat{\beta}_2)$, which is in each case:

$$\begin{aligned}
\tilde{x}_0(\hat{\beta}_1^H, \hat{\beta}_2^H) &= \pi(e^{ic})(2\beta^L - 2)m - \pi(e^{HH})(2\beta^H m - 2m - \rho^{HH}) \\
&\quad + [\pi(e^{HL}) + \pi(e^{LH})](\beta^H - \beta^L)m \\
\tilde{x}_0(\hat{\beta}_1^L, \hat{\beta}_2^H) &= \pi(e^{ic})(2\beta^L - 2)m - \pi(e^{LH})(\beta^H m + \beta^L m - 2m - \rho^{LH}) \\
&\quad + \pi(e^{LL})(\beta^H - \beta^L)m \\
\tilde{x}_0(\hat{\beta}_1^H, \hat{\beta}_2^L) &= \pi(e^{ic})(2\beta^L - 2)m - \pi(e^{HL})(\beta^H m + \beta^L m - 2m - \rho^{HL}) \\
&\quad + \pi(e^{LL})(\beta^H - \beta^L)m \\
\tilde{x}_0(\hat{\beta}_1^L, \hat{\beta}_2^L) &= \pi(e^{ic})(2\beta^L - 2)m - \pi(e^{LL})(2\beta^L m - 2m - \rho^{LL}).
\end{aligned}$$

b) Participation Constraint of Upstream Firm

The ex-post participation constraint of A is

$$\pi(e(\hat{\beta}_1, \hat{\beta}_2))[2m + \rho(\hat{\beta}_1, \hat{\beta}_2)] - c(e(\hat{\beta}_1, \hat{\beta}_2)) - x_0(\hat{\beta}_1, \hat{\beta}_2) \geq \pi(e^{ic})(2m) - c(e^{ic}),$$

which for each case results in:

$$x_0(\hat{\beta}_1^H, \hat{\beta}_2^H) \leq \pi(e^{HH})(2m + \rho^{HH}) - \pi(e^{ic})(2m) - c(e^{HH}) + c(e^{ic}) \tag{1.32}$$

$$x_0(\hat{\beta}_1^L, \hat{\beta}_2^H) \leq \pi(e^{LH})(2m + \rho^{LH}) - \pi(e^{ic})(2m) - c(e^{LH}) + c(e^{ic}) \tag{1.33}$$

$$x_0(\hat{\beta}_1^H, \hat{\beta}_2^L) \leq \pi(e^{HL})(2m + \rho^{HL}) - \pi(e^{ic})(2m) - c(e^{HL}) + c(e^{ic}) \tag{1.34}$$

$$x_0(\hat{\beta}_1^L, \hat{\beta}_2^L) \leq \pi(e^{LL})(2m + \rho^{LL}) - \pi(e^{ic})(2m) - c(e^{LL}) + c(e^{ic}) \tag{1.35}$$

Inserting the respective values for $\rho(\cdot)$, $e(\cdot)$, $\pi(\cdot)$, $c(\cdot)$, and $\tilde{x}_0(\cdot)$, equations (1.32) - (1.35) become

$$2(\beta^L)^2 - 4\beta^L + 2 \geq 0,$$
$$\frac{1}{2}(\beta^H)^2 + \frac{5}{2}(\beta^L)^2 - \beta^H\beta^L - 4\beta^L + 2 \geq 0,$$
$$\frac{1}{2}(\beta^H)^2 + \frac{5}{2}(\beta^L)^2 - \beta^H\beta^L - 4\beta^L + 2 \geq 0,$$
$$2(\beta^L)^2 - 4\beta^L + 2 \geq 0.$$

It can be checked that all four equations hold for $\beta^H \geq \beta^L \geq 1$. So we have shown that there exist $\tilde{x}$'s that induce truthful revelation and fulfill the participation constraints of each firm. □

1.C Proof of Proposition 4

Proof When there are k downstream firms, A receives an overall payment of $\sum_{i=1}^{k} p_i + \sum_{i=2}^{k} r_i$. It will try to maximize:

$$\begin{aligned}
&\max_{e|\left\{\substack{p_i = m, \forall i \in \{1,\dots k\} \\ \rho}\right\}} U_A \\
= &\max_{e|\left\{\substack{p_i = m, \forall i \in \{1,\dots k\} \\ \rho}\right\}} \pi(e)(\sum_{i=1}^{k} p_i + \sum_{i=2}^{k} r_i) - c(e) \\
= &\max_{e|\left\{\substack{p_i = m, \forall i \in \{1,\dots k\} \\ \rho}\right\}} \eta e(km + \rho) - \frac{\alpha e^2}{2},
\end{aligned}$$

which results in

$$\bar{e} = \frac{\eta}{\alpha}(km + \rho), \quad \text{with} \ \ \rho = \sum_{i=2}^{k} r_i.$$

Summing the payments r_i over the $(k-1)$ downstream firms, it can be seen that ρ is equal to the amount required to induce the optimal level of investment: $\sum_{i=2}^{k} r_i = \rho = \sum_{i=1}^{k} \beta_i m - km$. The incentive compatible level of investment on the other hand equals $e^{ic} = \frac{\eta}{\alpha}(km)$, which comes from maximizing A's $\pi(e)(km) - c(e)$. The incentives for renegotiation are similar to the trilateral contract - since no downstream firm pays more *for the product* than the market price, no one has incentive to renegotiate once the contract has been signed. Additionally, none of the downstream firms has an incentive to buy a low quality product. The participation constraint of P_1 is

$$\pi(\bar{e})(\beta_1 - 1)m - \sum_{i=2}^{k} x_{1i} \geq \pi(e^{ic})(\beta_1 - 1)m.$$

For the remaining $(k-1)$ downstream firms $P_i,\ \forall i \in \{2, \dots, k\}$, it equals

$$\pi(\bar{e})(\beta_i m - m - r_i) + x_{0i} + x_{1i} \geq \pi(e^{ic})(\beta_i - 1)m.$$

Summing up over all downstream firms, this results in

$$\sum_{i=2}^{k} x_{0i} + \sum_{i=2}^{k} x_{1i} \geq \pi(e^{ic})(\sum_{i=2}^{k} \beta_i - (k-1))m - \pi(\bar{e})(\sum_{i=2}^{k} \beta_i m - (k-1)m - \rho).$$

The participation constraint of the upstream firm is

$$\pi(\bar{e})(km + \rho) - c(\bar{e}) - \sum_{i=2}^{k} x_{0i} \geq \pi(e^{ic})(km) - c(e^{ic}). \tag{1.36}$$

Taking the participation constraints of all downstream firms as binding and inserting the respective values for $\sum_{i=2}^{k} x_{0i}$ and $\sum_{i=2}^{k} x_{1i}$, equation (1.36) becomes $\sum_{i=1}^{k} \beta_i \geq k$. For $\beta_i \geq 1, \forall i \in \{1, \dots, k\}$, this condition is fulfilled. The contract is implementable. □

1.D More Than Two Downstream Firms: Symmetric Case

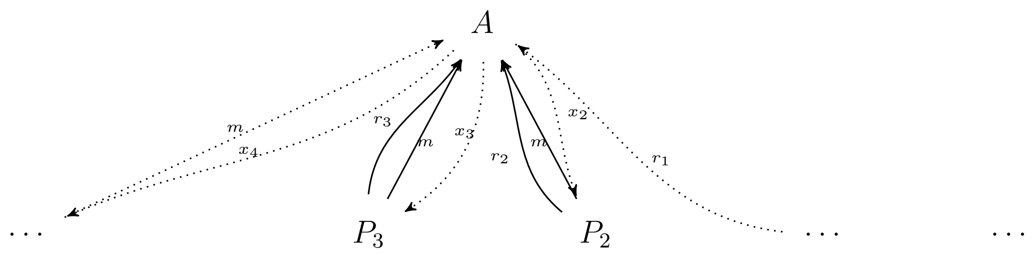


Figure 1.5: Multilateral Contract with More than Two Downstream Firms, Symmetric Case

Proposition 5 *There exists a self-enforcing multilateral contract $\tilde{c}$ that induces the optimal level of investment, and therefore increases overall welfare.*

Each downstream firm, when buying the product from A, pays a price p_i equal to the market price m. Conditional on buying, another downstream firm pays a transfer r_i to the upstream firm (see figure 1.5).

Proof Knowing that for producing high quality products, A receives a payment of $km + \rho$, it will try to maximize

$$\begin{aligned}
& \max_{e|\left\{\substack{p_i = m, \forall i \in \{1, \dots k\} \\ \rho}\right\}} U_A \\
= & \max_{e|\left\{\substack{p_i = m, \forall i \in \{1, \dots k\} \\ \rho}\right\}} \pi(e)(\sum_{i=1}^{k} p_i + \sum_{i=1}^{k} r_i) - c(e) \\
= & \max_{e|\left\{\substack{p_i = m, \forall i \in \{1, \dots k\} \\ \rho}\right\}} \eta e(km + \rho) - \frac{\alpha e^2}{2},
\end{aligned}$$

which results in

$$\bar{e} = \frac{\eta}{\alpha}(km + \rho), \ \text{with} \rho = \sum_{i=1}^{k} r_i.$$

The incentive compatible level of investment with k downstream firms is $e^{ic} = \frac{\eta}{\alpha}(km)$, resulting from A's maximization of $\pi(e)(km) - c(e)$. The ρ required to induce the optimal level of investment is $\rho = \sum_{i=1}^{k} \beta_i m - km$. The incentives for renegotiation are similar to the trilateral contract - since no downstream firm pays more *for the product* than the market price, no one has incentive to renegotiate once the contract has been signed. As well, no downstream firm has incentive to buy a low quality product. The participation constraints of the k downstream firms P_i, $i \in \{1, \dots, k\}$, are respectively

$$\pi(\bar{e})(\beta_i m - m - r_i) + x_i \geq \pi(e^{ic})(\beta_i m - m), \tag{1.37}$$

which results, for all downstream firms together, in

$$\sum_{i=1}^{k} x_i \geq \pi(e^{ic})(\sum_{i=1}^{k} \beta_i - k)m - \pi(\bar{e})(\sum_{i=1}^{k} \beta_i m - km - \rho).$$

The participation constraint of the upstream firm A is

$$\pi(\bar{e})(km + \rho) - c(\bar{e}) - \sum_{i=1}^{k} x_i \geq \pi(e^{ic})(km) - c(e^{ic}), \tag{1.38}$$

which, taking (1.37) as binding and replacing $\sum_{i=1}^{k} x_i$ can be simplified to

$$[\pi(\bar{e}) - \pi(e^{ic})] \sum_{i=1}^{k} \beta_i m \geq c(\bar{e}) - c(e^{ic}). \tag{1.39}$$

Replacing the functional forms, this results in $\sum_{i=1}^{k} \beta_i \geq k$. This condition is fulfilled for $\beta_i \geq 1 \ \forall i \in \{1, \dots, k\}$, and hence, the participation constraints are fulfilled. The contract is implementable. □

Chapter 2

Does Reciprocity Foster Mediocrity?

– When Reciprocity Can Not be Induced Through a Higher Wage –

"5. Non ti consiglio però di scegliere persone infamate per la scostumatezza e che fossero screditate presso del popolo; ma ti consiglio di scegliere uomini mediocri, timidi, incerti ne' principj, macchiati di vizi bensì ma con riserva, e tali che facciano eseguire quanto gli si ordina senza ambiguità alcuna."
Pietro Verri, Caligola

"In short, if one is entitled to everything, then one is thankful for nothing".
Emmons and Shelton, 2002

2.1 Introduction

In open competitions it may happen that participants complain about the choice of the selection committee because it was chosen a mediocre candidate and not the one who seems (or actually is) ex-ante the more qualified. This may simply be due to the fact that (almost) each candidate feels to be the most qualified to fill the position. Nevertheless, sometimes this complain appears justified: a candidate who is not the top one is chosen, even when there are objective criteria that suggest a commonly accepted ranking among candidates. This paper proposes an answer to the following puzzling question: in a job selection, if candidates' ex-ante productivity is public information, why should the employer choose an individual who is not the highest ranked, according to this criterion?

Our argument is the following: the agent who is *ex-ante* the most productive not necessarily is also the one that is *ex-post* the most productive. Ex-ante and ex-post productivity may differ due to the presence of non-contractible effort in performing the required task. If a mediocre (less ex-ante productive agent) exerts a higher non-contractible effort ex-post, this may overcompensate the principal for the loss in ex-ante productivity. Hence, selecting a mediocre agent may be overall the most profitable choice for the principal. If such is the case and the principals correctly anticipate this, then we observe (ex-ante) mediocre agent to be selected for a job.

Why should a mediocre agent exert a higher non-contractible effort than the top-ranked individual? Similar to the mechanisms underlying the working of the fair-wage hypothesis (Akerlof, 1982)[1], we believe that reciprocity (see for example Sobel (2005); Rabin (1993); Dufwenberg and Kirchsteiger (2004)) is the behavioral explanation of this phenomenon. Specifically, in our setting, agents who feel to be less *entitled* to fill a position (Fahr and Irlenbusch, 2000)[2] may reciprocate more than individuals who feel to deserve it, and exert a higher non contractible effort ex-post. To test our conjecture, we conduct incentivized decision making experiments in the laboratory.

We model a one-shot game between a principal and two agents competing for a position.[3] The principal has to select one agent to perform a task for a fixed compensation. Agents' productivity, and consequently the principal's payoff, depends on two components: a non contractible ex-post effort and an exogenous ex-ante productivity. The principal chooses between two agents with different ex-ante productivity. Each agent's ex-ante productivity is public information, which

[1]The fair-wage-hypothesis states that employers may pay a higher than the incentive compatible wage, anticipating correctly that this 'gift' is reciprocated by the employees. This hypothesis has been confirmed numerous times experimentally, both in the laboratory and in the field; see Gächter and Thöni (2010); Gneezy (2004); Fehr and Gächter (1998); Fehr, Gachter, and Kirchsteiger (1997); Charness (2004), and Hennig-Schmidt, Rockenbach, and Sadrieh (2010) for an overview over recent evidence.

[2]See as well Deresiewicz (2008) about the entitlement feeling of elite students.

[3]In the following, we will refer to the principal as she, while we refer to the two agents both as he.

means that it is known not only by the principal but by the other agent, too. Once the principal has selected one agent, this agent chooses a level of costly non-contractible effort. We run three different treatments. In the first treatment, the *communication* treatment, we allow the principal to send a message to the selected agent. In the second treatment, *no-communication* treatment, the principal could only select one agent, without having the possibility to communicate with him. Finally, in a control treatment, named *random device*, instead of the principal, a random device selects the agent who, as in the other two treatment, once hired chooses the level of non-contractible effort to exert.

The results are the following. In the *communication* treatment our conjecture is fully corroborate. Around 30% of principals select the agent who has the lowest ex-ante productivity. Selected agents with low ex-ante productivity exert on average an effort more than 50% higher than the agents with high ex-ante productivity. The higher effort overcompensates the ex-ante lower productivity, and principals who have chosen low productivity agents gain on average 40% more than principals who hired a high ex-ante productive agents. In the *no-communication* treatment, the fraction of principals who choose the lowest ex-ante productive agents is almost the same as in the *communication* treatment. However, there is no evidence of a higher effort provision by part of the low productivity employees: the average effort exerted by ex-ante high and low productive agents who are selected does not differ significantly. Finally, in accordance with our expectations, we find that in the *random device* treatment, agents with different ex-ante productivity exert on average the same amount of effort, and its level is the lowest among the treatments. Summing up, we find evidence of a reciprocity component induced by mediocrity, but only in the treatment with communication. We find this result interesting for two reasons. Firstly, we believe that the communication treatment better describes most of the real-world situations in which we observe the choice of a mediocre candidate. Usually, the selection occurs after a job interview, and in any case, before the agent performs the tasks, the principals always have the opportunity to tell the agents the reason why they have chosen them. Secondly, our experiment clarifies that the effect of the decision of selecting a mediocre candidate depends on the ability of the principals in inducing the mediocre agent to feel indebted towards them.

We believe that the emergence of *mediocrity* in candidate selection is particularly important for the civil service, and more in general, for all contracts where i) the employer cannot use the wage *level* to motivate the agents - either by linking the wage to the performance of the agent (a typical moral hazard setting) - or through a gift-exchange; ii) non-contractible effort is a relevant component of the employees' production function. Theoretical and experimental research on the gift-exchange (Fehr, Kirchsteiger, and Riedl, 1993), or, more specifically, the 'fair-wage hypothesis' has shown that by paying a 'fair' -wage, the problem of moral hazard can be overcome by making use of reciprocity. However, in our setting, wage is fixed. Hence, our experiment is an important extension of the study of reciprocity, providing evidence about its meaning in cases in which the wage level cannot be used to motivate workers. The novelty of our contribution is to link the gift exchange literature to the selection problem. Our theory predicts that the choice of the agent with lower productivity may be an important way to motivate the agent in such setting.

The motives for mediocre agents being in organizations has been investigated theoretically by several authors, among others, Gambetta and Origgi (2009); Kräkel (2009); Levine, Weinschelbaum, and Zurita (2010); Bramoulle and Goyal (20011) and Prendergast and Topel (1996). Bramoulle and Goyal examine the economic origins and the consequences of favoritism in groups. Kräkel shows that individuals that have the lowest fall-back positions have the highest incentives to succeed in career contests, hence mediocrity resulted out of a lack of incentive by part of the best agents. On the other hand, Prendergast and Topel suggest subjective evaluations as leading to favoritism, in settings in which performance is not objectively measurable, resulting in mediocrity. We, however, consider a setting in which performance *is* objectively measurable; and the more productive agents *is* ranked first; and, since we make use of a ranking which is not influenced by strategic concerns towards the latter jobs, incentives are not distorted. Gambetta and Origgi (2009) claim that norms agreeing upon "low quality" interactions may favor mediocrity, and give a descriptive example of favoring academic procedures in Italy. Furthermore, Levine, Weinschelbaum, and Zurita (2010) explain the presence of less than competent workers and overemployment by nepotism, nepotism being defined in the widest sense as a favoring of family members, friends, or any person from whose gratitude they could benefit. They identify factors when this leads to efficiency or inefficiency, making reference to extraneous reasons for which a firm may be forced to pay wages above the worker's reservation wage. The theoretical evidence is backed up by empirical evidence. Kramarz and Thesmar (2007) comment on poor performance of corporate executives due to favoritism of school mates in elite French schools; Anderson (2011) and Pande (2003) investigate why caste members tend to favor each other in political and economic interactions. Bertrand and Thesmar (2007) comment on favor exchange between corporate executives and politicians, and Durante, Labartino, and Perotti (2011) show that family connections in academic department faculty are correlated with lower performance in Italy.[4]

The rest of the paper is organized as follows. Section 2.2 describes our experimental design, including a discussion of the theoretical background. Section 2.3 presents our main results, and section 2.4 concludes.

[4]For a more detailed overview on literature regarding favoritism, see Bramoulle and Goyal (20011).

2.2 Experimental Design

2.2.1 Theoretical Background

We use a modification of the gift exchange game for our experiment. Gift exchange experiments are the most common approach to investigate reciprocity. The game was introduced by Fehr, Kirchsteiger, and Riedl (1993) and continued to be used in numerous laboratory and field experiments (Charness, Frechette, and Kagel, 2004; Gneezy and List, 2006; Charness, 2004; Hennig-Schmidt, Rockenbach, and Sadrieh, 2010; Kube, Marechal, and Puppe, 2010; Hannan, Kagel, and Moser, 2002).

We use a three-player game of complete information with one principal and two agents.[5] Agents have randomly assigned different level of productivity (each agent's productivity is observed by the principal and the competing agent). At stage one, the principal chooses one of the two agents. At stage two, the selected agent decides whether to refuse the offer ($e = 0$) or to accept it, choosing a level of non contractible effort $e > 0$.

The principal's monetary payoff function is given by

$$\begin{aligned} P(e_j, \theta_j) = & \quad 0 & \text{if } e_j = 0 \\ = & \quad (v_0 - w) \cdot (e_j + \theta_{jk}) & \text{if } e_j > 0, \end{aligned} \tag{2.1}$$

where v_0 represents, as in Fehr, Kirchler, Weichbold, and Gaechter, an exogenously given redemption value. The principals payoff is strictly increasing in the effort e_j and the productivity θ_{jk} of the chosen agent j, $k =$ high, low.

[5] Other papers that investigate reciprocity with two agents include Charness and Kuhn (2007); Gaechter, Nosenzo, and Sefton (2010); Gächter and Thöni (2010); Abeler, Altmann, Kube, and Wibral (2009) and Güth, Königstein, Kovács, and o (2001). Güth, Königstein, Kovács, and o report an experiment in which a principal can pay two agents with deterministic but unequal productivity equal or unequal wages. Their treatment variable is the information that one agent has about the other agent's contract offer. They find that principals offer less asymmetric contracts when work contracts are observable than when contracts are not observable, and conclude that horizontal fairness matters. Similarly, Gächter and Thöni find that experimental workers who face disadvantageous wage discrimination significantly reduce their effort relative to a situation with equal wages. And also Charness and Kuhn examine the workers' responsiveness to coworkers' wages. In their setting, one principal is paired with one low ability and one high ability worker, and they also vary whether wages are public or private. Principals can determine the wages of both workers, choosing equal or different wages. They find that effort is not affected by coworkers' wage. Similarly, Gaechter, Nosenzo, and Sefton find that pay comparison information does not affect reciprocity in a three-person gift-exchange game where an employer pays two employees. Abeler, Altmann, Kube, and Wibral consider a setting in which the principal pays equal wages in one treatment and can set individual wages in the other. They find that the use of equal wages elicits lower efforts, concluding that this difference seems to be driven by the fact that the norm of equity is violated far more frequently in the equal wage treatment. Yet, all those settings differ from ours in the sense that our principal cannot choose whether to offer equal wages or not. We are not interested in the effort of the disadvantaged agent, and our game is of complete information.

The selected agent's monetary payoff function is simply the difference between the wage w, the incurred effort costs $c(e_j)$ and a fixed cost c_0:

$$\begin{aligned} p_j = &\ 0 && \text{if } e_j = 0 \\ = &\ w - c(e_j) - c_0 && \text{if } e_j > 0 \end{aligned} \tag{2.2}$$

In our experiment, we set $v_0 = 120$, $c_0 = 20$, $\theta_H = 0.05$ and $\theta_L = 0.00$, $w = 70$ and the available effort levels and their corresponding effort costs are the ones that are depicted in Table 2.1. The agent that is not selected receives an unemployment benefit of $w_0 = 10$.

Effort level	0	1	2	3	4	5	6	7	8	9	10
Cost of effort	0	0	1	2	4	6	8	10	12	15	18

Table 2.1: Effort Levels and Costs of Effort

A payoff maximizing agent accepts the offer and chooses a level of effort $e = 1$, independently of his productivity. Hence, assuming that the principal does not play weakly dominated strategies, in equilibrium a principal who expects a payoff maximizing selects the high productivity agent.In the same way, a principal expecting the same level of effort by both agents, irrespectively from their ex-ante productivity, will choose the ex-ante high productive agent.

Now what if we include reciprocity? Let β be a dummy, $\beta = 1$ if the chosen agent is the low productivity agent and $\beta = 0$ if he is the high productivity agent, and $\rho_i \in [0,1]$ a reciprocity parameter. Let $\rho_j = 0$ if j is not reciprocal at all and $\rho_i > 0$ if j is reciprocal; in addition, we include a kindness parameter $\lambda_j \in [-1,1]$ that depends on β:

$$\begin{aligned} \lambda_i(\beta) &> 0, && \text{if } j \text{ beliefs the principal is kind} \\ &= 0, && \text{if } j \text{ is does not belief that the principal is kind.} \end{aligned} \tag{2.3}$$

Selecting the low productivity agent is profitable for the principal if

$$P(e, \theta | e(\rho_j, \lambda_{j,\beta=1})) \geq P(e, \theta | e(\rho_j, \lambda_{j,\beta=0})). \tag{2.4}$$

Since the employers payoff function is increasing in the effort provided by the employee, the employees payoff function, including reciprocity, become :

$$\begin{aligned} p_j = &\ 0 && \text{if } e_j = 0 \\ = &\ w - c(e_j) - c_0 + \lambda_j(\beta)\rho_j e_j && \text{if } e_j > 0. \end{aligned} \tag{2.5}$$

Hence, if $\rho = 0$ or $\lambda(\beta) = 0$, FOCs result in $c'(e) = 0$ and the optimal effort for a money maximixing individual remains $e = 1$. If however $\rho > 0$ and $\lambda > 0$, choosing $c_j = \frac{\alpha}{2} e_j^2$ with $\alpha > 0$ leads to an optimal effort provision of

$$e_j = \frac{\lambda_j(\beta)\rho_j}{\alpha}, \tag{2.6}$$

which is increasing in ρ and in λ. Therefore, if $\lambda(\beta = 1) > \lambda(\beta = 0)$, the choice of effort is higher when the selected agent is the low productivity agent

Replacing this optimal effort for the agent back into the principals' payoff function and assuming the payoff function still to be as in equation 2.2, it results that employing the second best agent is beneficial if

$$\lambda_{j,\beta=1} - \lambda_{j,\beta=0} > (\theta^H - \theta^L)\frac{\alpha}{\rho_j}. \tag{2.7}$$

This means that, depending on the parameters α and ρ, it is beneficial to employ the second best agent if the induced increment in reciprocity is higher than the difference in abilities. Note that the way how we have structured the cost function of effort up to now, it is beneficial for the principal to employ the second productive agent if reciprocity leads to an increase in effort of one level.

What about the maximization of the payoffs' sum? Without reciprocity, the effort that maximizes the sum of the principals' and the employers' payoff is

$$\frac{v - w}{\alpha} = e^{FB}. \tag{2.8}$$

With reciprocity, however, this effort becomes

$$\frac{v - w + \lambda\rho}{\alpha} = e^{FB}_{rec}. \tag{2.9}$$

Replacing this back into the sum of the payoff functions, we find that overall efficiency is maximized when the second productive agent is employed, if

$$\alpha(\theta^H - \theta^L) \leq (\lambda_{j,\beta=1} - \lambda_{j,\beta=0})\rho + \rho^2\frac{\lambda^2_{j,\beta=1} - \lambda^2_{j,\beta=0}}{v - w}. \tag{2.10}$$

2.2.2 Procedures

The experiment was computerized with the software Z-tree (Fischbacher, 2007) and conducted in the experimental laboratory of the strategic interaction group at the Max Planck Institute of Economics in Jena (MPIEJ), Germany. Subjects were mainly student from the MPIEJ Experimental and Behavioral Economics Laboratory subject pool. They were recruited via the ORSEE software (Greiner, 2003). We conducted 16 sessions, featuring 156 groups with a total of 468 subjects, in November and December 2011. The sessions last about 50 minutes. Average payment was 9.88 Euros, including the show-up fee.[6]

Upon arriving at the experiment, participants were randomly allocated to computer terminals, were given a paper copy of the instructions that were read aloud

[6]Our employees were confronted with a payoff table instead of a payoff function (see also Charness, Frechette, and Kagel, 2004). The experiment was conducted in German. In the Appendix an English version of the instruction is reproduced. Full instructions and the software are available from the authors.

Treatment	N firms	N participants
Communication	58	174
No-communication	58	174
Random Device	40	120
Total	156	468

Table 2.2: Participants and Treatments

to establish common knowledge. Before starting the experiment, subjects had to answer a control questionnaire which tested their comprehension of the rules; after the experiment they could answer a short unincentivized post-experimental questionnaire.[7] Our experiment has two parts. In part one, we conducted a one-shot gift exchange, after which in part two we played the same game several rounds repeatedly- 7 rounds in the treatment *no-communication* and 5 rounds in the *communication* and the *random device* treatment.[8]

Participants were told that the experiment is composed by two parts. They first received instruction about phase one, then, at the end of part one, they received the instruction about the content of the second part. In part one, each participant was matched with two other participants to form a group of three. We referred to each group as a firm, and to the group members as employer, employee A and employee B. Privately, each subject got randomly assigned a role. Note that we assign the level of productivity randomly; we wanted there to be no strategic thinking involved in this step. Within the firm, there is one job to be offered.

At this point, the three treatments differ. In the *no-communication* treatment, first, the employer chooses one employee to hire. Then the selected employee learns being the selected, and chooses a level of effort. In the *communication* treatment, when choosing the employee to hire, the employer can simultaneously send a message to him. The treatment is described more in detail in section 2.2.2. The selected employee learns being selected, can read the message and chooses a level of effort. In the *random device* treatment, instead of an employer choosing one employee, a random device selects one of the two employees - see also section 2.2.2. Thereafter the employee learned to be selected, and chose a level of effort.

In part two, while the procedures and subjects' roles stay the same as in part one, the subjects are rematched with other two participants in each round such that a perfect strange matching is implemented.

[7]We categorized the content of the messages following Koukoumelis, Levati, and Weisser (2009): first two researchers examined independently a sample of the messages and establish their own distinct set of preliminary categories. Each category represented one or more arguments that the communicator wants to transmit, and each message may pertain to one or more categories. Then the two sets of categories were compared and the two researchers agreed on which to use. Thereafter, other two student assistants from the University of Jena classified each message according to the categories.

[8]In the treatment *communication*, in one session the repetitions were just four due to a problem with the software.

Communication

When including communication, the principal could use a text box to type in her message; she had a maximum of three minutes to compose the message and the only restrictions regarding the content were that she could neither identify herself, nor indicate something to happen after the experiment had ended (threaten the other group members, promise a side payment or similar things). Messages were screened before being sent, and if a subject did not comply with these restrictions, she was not paid. This actually never happened in our sessions. It was common knowledge that the messages were cheap talk (i.e., non-binding) and costless in the common sense.

Different to cheap talk usually used in the literature[9], the exchange of messages in our setting neither led to an increase of information (everything is common information before messages were sent) nor does it include non-binding announcements about intended decisions, since the messages are received by the employees at the same time as they are informed about the action taken by the employer.

Yet, beside such functioning, it is still possible that the messages increase the closeness among participants. In addition, it gives our employers the possibility to explain their expectations. Hence, beside assuring the employers that the employee knows the motivation for choosing him, it might also lead to increase in the reciprocal behavior. In short, including communication should lead not only to an overall increased reciprocity due to an increase in closeness among the participants, but should also lead to an increase in reciprocity by part of the low productivity employee due to an increment in the shared reasoning about the strategy to play in this game.

Random Device Treatment

The aim of this treatment was to investigate what happens if the employees are selected by a random device instead of being chosen by an employer. We still expect them to reciprocate more than the minimum effort, since, as in the previous two treatments, agents may be motivated as well by general context dependend non-standard other regarding preferences, such as altruism, inequality aversion, or max-min preferences (Dufwenberg, Heidhues, Kirchsteiger, Riedel, and Sobel, 2008; Sobel, 2005; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002) and, again, the surprise effect. Yet, we expect there to be no difference in effort provision by part of the high and low productivity workers that are selected.

[9]For a nice overview of survey of experiments on communication and cheap talk as well as an overview over the models introduced by Crawford and Sobel (1982) and Green and Stokey (1980), see Crawford (1998). Survey that addresses the theoretical conditions under which augmenting a game with cheap talk leads to achieving efficient outcomes: Farrell and Rabin (1996).

Beliefs

In the *random device* treatment, the *communication* treatment and two sessions of the *no-communication* treatment we elicited incentivezed beliefs from the employers about the level of effort he expects the employee to provide. The reason is that the principal might select the second productive not only because he expects a higher effort from him than from the agent ranked first, but also to give him "a chance" , out of compassion, randomly, or out of curiosity ("what happens if?"). On the contrary, he might also select the *more productive* agent ex-ante because, for some reason, he expects him to reciprocate more than the agent with lower ex-ante productivity, or because he expects this to maximize overall sum of payoff of the firm. By asking the level of effort the principal expects from the employer she selects, we want to collect incentivized answers regarding their motivations in selecting the agent they choose, adding information to the unincentivized answers the subjects give in the post-experimental questionnaires. If their guess was exact, they received 10 ECUs extra; if it was in a range of $+/-$ 1 point, they received an additional 5 ECUs, and if it was in a range of $+/-$ 2 points, they received an additional 2 ECUs. We did not ask to principals the productivity of the agents who were not selected due to the impossibility of incentivizing their beliefs. We asks beliefs after the choice was made and we believe that at this point it is worthless to ask the expected productivity of the non selected agents, because principal could express ex-post self-confirming beliefs.

Furthermore, we elicited incentivized beliefs from the employees about the number of principals they expect to choose the high productivity employee, and we also asked the subjects whether they expected to be chosen by themselves or not (even though the latter question was unincentivized). Since it is possible that our employees did not expect the employer to choose the low productivity employee, it is possible that we only observe the effort choice of "the high productivity employees that expect to be chosen" and "low productivity employees that do *not* expect to be chosen". If a subject should not expect to being chosen, conditional on being selected he might provide a *higher* effort than a subject who expects to be chosen, due to a positive surprise. Hence, this could drive our results in the same direction as what we want to observe, and we wanted to control for such.

2.3 Results

In the following, we will focus on the results of the *first part* of the experiment, the one shot game our subjects were playing without knowing about the second part of the experiment. With respect to the second part of the experiment we will comment on the results when they are significantly different to the results in the first part. Note however that our subjects did not receive any information about their payoff or the strategies of the other players, during the whole experiment. Except the beliefs of the employees about the probability of being chosen (out of their own experience of being (not) chosen), we can exclude any updating of beliefs. We proceed as follow: first we discuss the choice made by the principal, then we focus on the agents' effort choices.

Our first result refers to the choice of the principals:

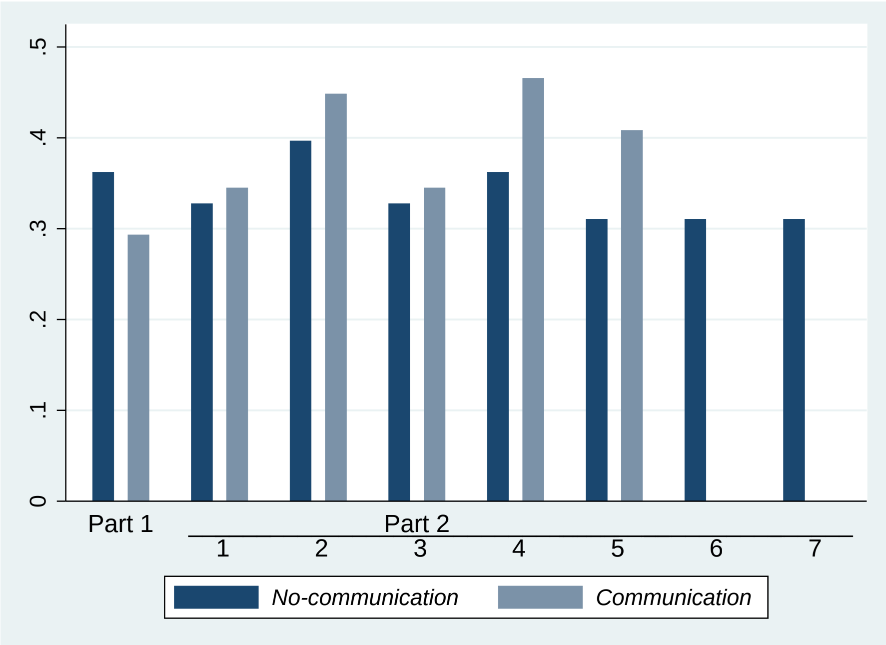


Figure 2.1: The Choice of the Employers in the *Communication* and *No-Communication* Treatment

Result 1 *A significant share of our principals choose the low productivity worker.*

Support for result 1 can be found in Figure 2.1 and Table 2.3. Although the majority of employers choose to employ the high productive workers, in the *communication* treatment 29.31% of the employers chose the low productivity agent, and in the *no-communication* treatment 36.21% of the employers chose the low productivity agent. The two proportions are not significantly different from each other (two sample test of proportions: $z = 0.79$ $p = 0.43$). Moreover, the two proportions are significantly different from 50%. Hence, it follows that we can reject the hypothesis that principals randomly choose the agents.[10] Do the principals take their decision out of the motivation we are investigating?

Our second results refers to the effort exerted by the selected agents. We first concentrate on all the selected agents irrespectively from their ex-ante productivity.

[10]While these two percentages are significantly lower than 50% (Binominal probability test, $p = 0.05$ and $p < 0.01$, respectively), they are also significantly higher than 10% - an amount that might be reasonable to be accepted as people "making mistakes" (Binominal probability test, $p < 0.01$ and $p < 0.01$, respectively).

	Part 1	**Part 2**						
		Round 1	2	3	4	5	6	7
Communication treatment								
%	29.31	34.48	44.83	34.48	46.55	40.82		
n	17	20	26	20	27	20		
N	58	58	58	58	58	49		
Binomial probability test:		Probability that this is equal to 50%						
p-value	< 0.01	0.03	0.51	0.02	0.70	0.25		
Binomial probability test:		Probability that this is equal to 10%						
p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01		
No-communication treatment								
%	36.21	32.76	39.66	32.76	36.21	31.03	31.03	31.03
n	21	19	23	19	21	18	18	18
N	58	58	58	58	58	58	58	58
Binomial probability test:		Probability that this is equal to 50%						
p-value	0.05	0.01	0.15	0.01	0.05	< 0.01	< 0.01	< 0.01
Binomial probability test:		Probability that this is equal to 10%						
p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Random device treatment								
%	57.50	52.50	52.50	55.00	65.00	45.00		
n	23	21	21	22	26	18		
N	40	40	40	40	40	40		
Test of proportions:		Difference in the *comm.* / *no-comm.* treatment						
z	0.79	0.20	0.56	0.20	1.13	1.05		
p-value	0.43	0.84	0.57	0.84	0.26	0.29		

Table 2.3: Fraction of Low Productive Workers Chosen

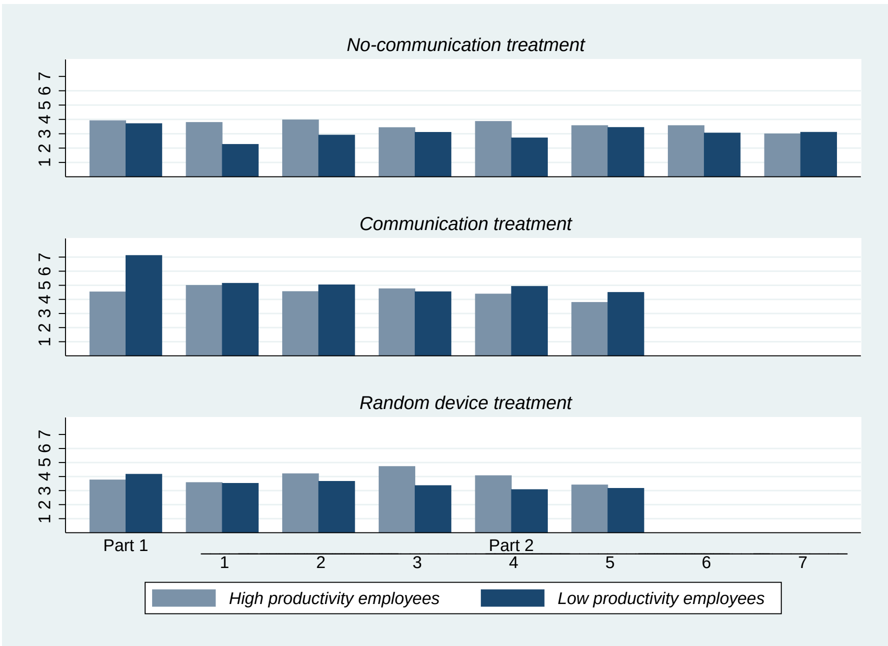


Figure 2.2: Effort Provided by the High and the Low Productivity Employees

Result 2 *On average, the agents exert a higher than the minimum enforceable effort. Agents in the* communication *treatment exert on average a significantly higher effort than agents in the* no-communication *treatment. The average effort exerted in the* random device *treatment is not significantly different from the one exerted in the* no-communication *treatment.*

The effort chosen by the employees is summarized in Figure 2.2 and Table 2.4. In the *communication* treatment, they exert on average an effort of 5.29, in the *no-communication* treatment, the selected agents exert on average an effort of 3.84, and in the *random device* treatment, they exert on average an effort of 4.00. Hence, the agents exert a significantly higher effort than the minimum enforceable effort of 1, which is what would be expected from pure money maximizers (*no-communication* treatment, Mann Whitney test, $p < 0.01$, $z = 8.10$, *communication* treatment, $p < 0.01$, $z = 8.24$, and *random device* treatment, $p < 0.01$, $z = 6.17$)[11]. The difference in the levels of effort provided in the *communication* and *no-communication* treatment is statistically significant (Mann Whitney test, $z = 2.88, p < 0.01$). The difference between the *random device* and *no-communication* treatment is not significant ($z = 0.08, p = 0.94$). Hence, in accordance with the literature, we find that communication has a positive effect on reciprocity.[12]

This choice of effort may be due to reciprocity as well as to different other regarding preferences like inequity aversion or altruism. In the *communication* and *no-communication* treatment an higher effort can be also due to the "I want YOU" effect.[13] However, it is only the reciprocity effect induced through the entitlement, or the effect of being chosen expecting such / not expecting such should lead to the two agents providing *different* level of effort, conditional on being selected. Therefore it is important to consider the difference in the effort exerted by the two types of agents. When considering this difference, we find that:

[11] This result is anticipated by the principals. When asking incentivized beliefs, we find that the employers expect on average a higher effort than the minimum enforceable effort: in the *no-communication* treatment, the employers on average expect an effort of 3.72 from the employees. This is statistically different from one, the profit maximizing effort for the employees (Mann Whitney test, $z = 8.08$, $p < 0.01$). Similarly, in the *communication* treatment, on average the employers expect an effort of 5.67, which again is statistically different from one ($z = 9.43$, $p < 0.01$), and the same holds for the *random device* treatment - in the latter treatment, employers were informed about they employee they got assigned and had then to state the effort they expected from him/her - (average expected effort: 3.45, $z = 5.86$, $p < 0.01$).

[12] In line with the literature is also our result regarding the second part of the experiment: in the *no-communication* treatment, significance levels regarding the difference in the effort provided between the two groups of agents are similar over all rounds except two. In the *communication* treatment, the difference is not significant in any round. This indicates that the effect of communication does not persist, even though our principals could send a new message in each round (see for example Abeler, Altmann, Kube, and Wibral (2009)).

[13] Brandts, Guth, and Stiehler (2006) show that knowingly selected allocators keep less for themselves than randomly selected ones. Since in our setting in any case only *one* agent can be chosen, we should find in any case that the one that is chosen is happy and hence reciprocates.

Result 3 *In the* communication *treatment, the low productivity agents on average exert a significantly higher effort than the high productivity agents. There are no differences in average effort exertion by part of the high and low productivity agents in the other two treatments.*

In the *communication* treatment, the high productivity agents on average exert an effort of 4.54 when being selected, while the low productivity agents on average exert a significantly higher effort of 7.12 (Mann Whitney test, $z = 3.69$, $p < 0.01$). In the *no-communication* treatment, the high productivity agents on average exert an effort of 3.92, while the low productivity agents on average exert an effort of 3.71; the difference is not significant ($z = 0.09$, $p = 0.93$). In the *random device* treatment, the high productivity agents on average exert an effort of 3.76 when being selected, while the low productivity agents on average exert an effort of 4.17; the difference is again not significant ($z = 0.38$, $p = 0.71$).[14]

The difference in effort provision by part of the chosen agent leads to the following:

Result 4 *In the* communication *treatment, choosing the low productivity agent is profit maximizing for the principal. It is not significantly more profitable to choose the low instead of the high productivity agent in the* no-communication *treatment.*

In the *communication* treatment, average profit for the principals that choose the high productivity employer is 25.18 ECUs. On the contrary, average profit for the principals that choose the low productivity employer is 35.59 ECUs. The difference is statistically significant (Mann Whitney test, p=0.01, z=2.55). On average, this means that in *communication* treatment the principals who chose a low productive gained 43% more. In the *no-communication* treatment, on the other hand, average profit for the principals that choose the high productivity employer is 22.09 ECUs and average profit for the principals that choose the low

[14]Also this is in line with the expectations of the employers. In the *no-communication* treatment, the employers that choose the low productivity employee on average expect a higher effort choice than the employers that choose the high productivity agents. In the *communication* and in the *random device* treatment, the expectations of the employers that choose the low productivity employee and the employers that choose the high productivity agents do not differ. In Table 2.5 and Figure 2.3 we report the average employers' expected effort. In the *no-communication* treatment, when selecting an high productivity agent, the employers expect an effort of 2.95 from the those agent. On the contrary, when selecting an low productivity agent the employers expect on average an effort of 5.10 from the those agent. The difference is significant (Mann Whitney test, $p < 0.01, z = 3.63$). In the *communication* treatment, the principals expect on average an effort of 5.70 from the high productivity agent when selecting such, and when selecting an low productivity agent the employers expect on average an effort of 5.59 from him. The difference is not significant ($p < 0.92, z = 0.11$). In the *random device* treatment, the employers that got assigned a high productivity agent expect on average an effort of 3.88 from those agents. When getting assigned a low productivity agent the employers expect on average an effort of 3.13 from him. Again, the difference is not significant ($p = 0.68, z = 0.41$).

productivity employer is 18.57 ECUs. The difference is not statistically significant (p=0.27, z=1.10).[15]

	Part 1	**Part 2**						
		Round 1	2	3	4	5	6	7
Communication treatment								
Av. effort by A^H	4.54	5.00	4.56	4.67	4.39	3.79		
Av. effort by A^L	7.12	5.15	5.04	4.55	4.32	4.50		
MW test: z	3.69	0.50	0.89	0.02	1.00	1.28		
p-value	< 0.01	0.62	0.37	0.99	0.29	0.20		
No-communication treatment								
Av. effort by A^H	3.92	3.79	3.97	3.44	3.87	3.58	3.58	3.00
Av. effort by A^L	3.71	2.26	2.91	3.11	2.71	3.44	3.1	3.1
MW test: z	0.09	2.15	1.68	0.65	1.65	0.17	0.79	0.01
p-value	0.93	0.03	0.09	0.51	0.10	0.86	0.43	0.99
Random device treatment								
Av. effort by A^H	3.76	3.58	4.21	4.72	4.07	3.41		
Av. effort by A^L	4.17	3.52	3.67	3.63	3.08	3.17		
MW test: z	0.38	0.07	0.58	1.47	0.99	0.55		
p-value	0.71	0.94	0.56	0.14	0.32	0.58		
Does the average effort differ in the *communication* and *no-communication* treatment?								
Diff. A^H:MW-test: z	0.95	1.95	0.96	2.35	0.76	0.22		
p-value	0.34	0.05	0.34	0.02	0.45	0.83		
Diff. A^L: MW-test: z	3.83	3.33	2.52	1.64	2.36	1.56		
p-value	< 0.01	< 0.01	0.01	0.10	0.02	0.25		
Does the average effort differ in the *random device* and *no-communication* treatment?								
Diff. A^H:MW-test: z	0.25	0.20	0.33	1.54	0.21	0.05		
p-value	0.80	0.84	0.74	0.12	0.84	0.96		
Diff. A^L: MW-test: z	0.28	1.89	1.10	0.42	0.53	0.42		
p-value	0.78	0.06	0.27	0.68	0.60	0.68		
Does the average effort differ in the *communication* and *random device* treatment?								
Diff. A^H:MW-test: z	0.96	1.74	0.56	0.09	0.51	0.12		
p-value	0.34	0.08	0.58	0.93	0.61	0.91		
Diff. A^L: MW-test: z	3.10	2.08	1.69	1.41	2.21	1.51		
p-value	< 0.01	0.04	0.09	0.16	0.03	0.13		

Table 2.4: Effort Provided by the High and the Low Productive Employees that are Selected

[15]It is out from the goal of this experimental design to investigate whether hiring a mediocre agent is beneficial or not for the *overall organization* in terms of efficiency. In our design, indeed, the non-contractible effort exerted by the agents can be thought both as an effort which is productive for the organization and as an effort which only produces benefits for the principal. We leave a more detailed discussion on this issue on the conclusion.

	Part 1	**Part 2**						
		Round 1	2	3	4	5	6	7
Communication treatment								
Av. exp. effort by A^H	5.70	5.37	5.31	4.79	4.81	4.79		
Av. exp. effort by A^L	5.59	5.35	5.23	5.90	5.33	5.35		
MW test: z	0.11	0.08	0.23	1.85	0.91	0.88		
p-value	0.92	0.93	0.81	0.06	0.36	0.38		
No-communication treatment								
Av. exp. effort by A^H	2.95	3.54	2.94	3.18	2.92	2.85	2.95	2.58
Av. exp. effort by A^L	5.10	4.16	4.83	4.84	5.24	4.72	4	4.56
MW test: z	3.63	0.95	3.24	2.50	3.48	3.21	2.01	3.66
p-value	< 0.01	0.34	< 0.01	0.01	< 0.01	< 0.01	0.04	< 0.01
Random device treatment								
Av. exp. effort by A^H	3.88	3.47	3.16	3.83	2.50	4.31		
Av. exp. effort by A^L	3.13	3.61	4.29	3.68	4.50	3.11		
MW test: z	0.41	0.31	1.09	0.14	2.31	1.34		
p-value	0.68	0.76	0.28	0.89	0.02	0.16		
Does the expected effort differ between the *communication* and *no-communication* treatment?								
Diff. A^H: MW-test: z	5.13	2.99	3.81	2.82	2.90	3.23		
p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01		
Diff. A^L: MW-test: z	0.76	1.72	0.96	1.94	0.50	1.38		
p-value	0.45	0.09	0.34	0.05	0.62	0.17		
Does the expected effort differ between the *communication* and *random device* treatment?								
Diff. A^H: MW-test: z	2.13	2.56	3.10	1.54	2.66	0.55		
p-value	0.03	0.01	< 0.01	0.12	< 0.01	0.58		
Diff. A^L: MW-test: z	2.89	2.41	1.55	.2.82	1.11	2.63		
p-value	< 0.01	0.02	0.12	< 0.01	0.27	< 0.01		
Does the expected effort differ between the *no- communication* and *random device* treatment?								
Diff. A^H: MW-test: z	0.56	0.25	0.57	1.02	0.58	2.11		
p-value	0.58	0.80	0.57	0.31	0.56	0.03		
Diff. A^L: MW-test: z	2.85	0.85	1.00	1.27	0.92	2.47		
p-value	< 0.01	0.39	0.32	0.20	0.36	0.01		

Table 2.5: Effort Employers Expect From the Employee They Select

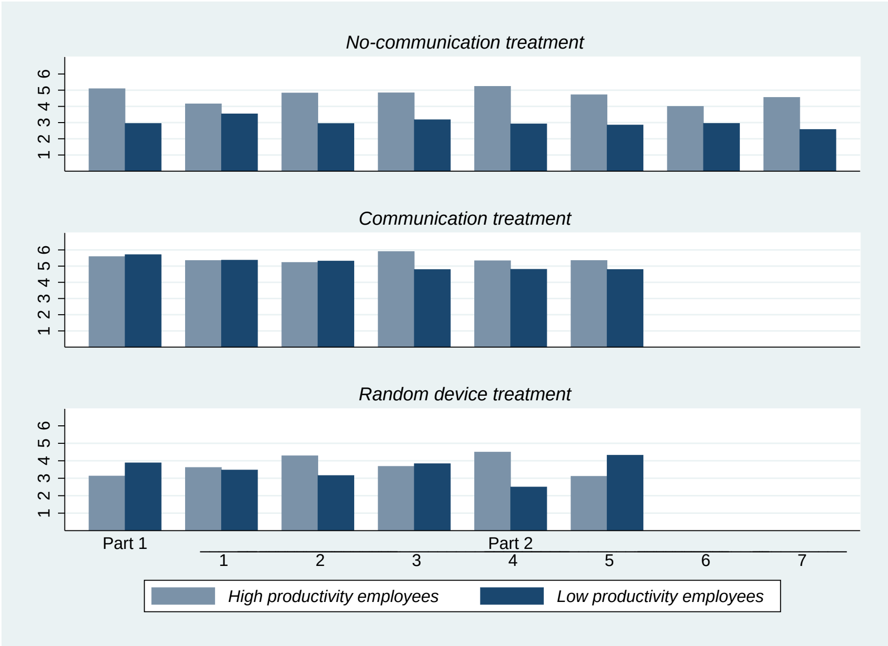


Figure 2.3: What Effort do Employers Expect from the Chosen Employees?

The most plausible explanation for the low productivity workers in the *no-communication* treatment to not to reciprocate more than the high productivity workers, given on being selected, seems to be the fact that they did not anticipate the expectation of the principals. This may be interpreted as a signal of the power of communication, which seems to be more effective on low productivity agents. While the fact that overall effort provided is higher in the *communication* than in the *no-communication* treatment is in line with literature concerning results on communication increasing closeness and reciprocal responses, it also seems as if the principals can effectively transmit that they *do* expect a reciprocal response due to the choice and the differences in productivity when choosing the low productivity agent. Hence, this points into the direction that the entitlement effect is emphasized when there is the possibility to send a message. On the contrary, in the *no-communication* treatment, it may have played a role that we assigned the level of productivity randomly - we did not want to have strategic interaction in the assignment of productivity - if such was perceived as strongly random by the employees, the entitlement effect would not exist. This intuition is partly confirmed by looking at the answers of the post-experimental, unincentivized questionnaires.[16]

[16]Further research is required to investigate different time horizons and reciprocal behavior. As seen in the *communication* treatment, the effect of low productivity agents providing a higher effort than the high productivity agents disappears once we look at what happens when the agents have to provide effort over several periods. This is in line with previous research (Kube, Marechal, and Puppe, 2010; Gneezy and List, 2006), yet, it also indicates that even if it *seems* to be an profitable choice for the principal in the first place to choose the second best, it is important to keep in mind long term effects. As we find that the reciprocal response decreases, the short term benefit of a principal interested in increasing the profit in the short term by making use of reciprocity of a low productivity agent will result in a loss in the long term.

Result 5 *The shares among the high and low productivity employees that mention reciprocity do not differ in either treatment.*

In the *no-communication* treatment, of the low productivity workers that are chosen, only 29% (6/21) mention reciprocity as the motivation for why they have chosen the specific effort level, while a similar share of 27% (10/37) of the selected high productivity agents mentions such as a motivation (Two sample test of proportions, $z = 0.12$, $p = 0.90$, Binomial probability test, $p = 0.80$). On the contrary, in the *communication* treatment, 47% (8/17) of the low productivity agents mention reciprocity as a motivation for their choice, and 39% (16/41) of the high productivity agents (Two sample test of proportions, 0.57, $p = 0.90$, Binomial probability test, $p = 0.35$). The difference in proportions over treatments among the low productivity agents mentioning reciprocity is not significant (Binominal probability test, two-sided, p =0.12)[17].

To investigate the impact of the surprise of being chosen, we take a look to the expectations of the employees regarding the choice of the employer.

Result 6 *Most employees expect the employers to choose the high productivity agent.*

[17]Further research is required to control for this, by giving the employees a stronger entitlement feeling. As noted by Fahr and Irlenbusch (2000), several experimental studies have shown that in ultimatum games the behavior of the first mover changes when participants earn different property rights. Hence, by including a tournament to let agents compete for who has the higher productivity, we should find that the high productivity agents feel more entitled to be chosen; similarly, the low productivity agents should feel less entitled, and hence they should be more reciprocal when being chosen. Another explanation for the lack of higher effort provision by part of the low productivity agents may be the presence of peer effects (Falk and Ichino, 2006; Mas and Moretti, 2009; Abeler, Altmann, Kube, and Wibral, 2009; Montinari, 2011). Horizontal fairness concerns may influence in a way that the low productivity agents, when selected, consider this as being so unfair towards the high productivity agents, that they do not want to reciprocate any more. However, if such would be the reason behind the behavior of our agents, then why did the principals not anticipate such? Furthermore, we do not find any evidence of such reasoning in the post-experimental questionnaires. Also overconfidence may play a role (Billett and Qian, 2008; Charness, Rustichini, and van de Ven, 2011). If this effect is stronger than the entitlement effect, then indeed we do not find a difference among the two levels of effort exerted by part of the high / low productivity agents - but again, while the employers do not seem to fear such effect, we also do not find any evidence of such mentioned in the post-experimental questionnaires. Similarly, it is important to take into account the cross effects from selecting the second best agent; not only the peer effects that arise in the selected agent, but also the impact such behavior has on the agent that is not selected. While several papers have taken into account the effect of different contracts given to agents that differ in productivity (Güth, Königstein, Kovács, and o, 2001; Abeler, Altmann, Kube, and Wibral, 2009; Gächter and Thöni, 2010; Charness and Kuhn, 2007; Gaechter, Nosenzo, and Sefton, 2010), thereby making reference to the productivity of the disadvantaged agent and informational concerns, up to now we are not interested in such effect - we do not consider the effort of the second worker. However, obviously this might be an important matter in future research, especially when thinking about a setting with a longer time horizon; once the agent anticipate that it might be more beneficial to be the second best than the best, this might have disastrous effects on motivation.

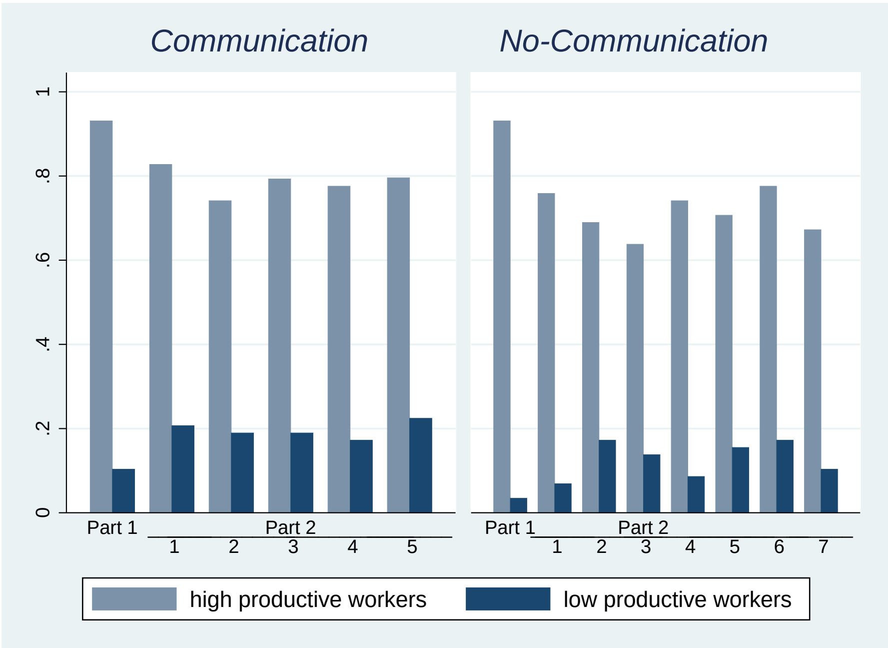


Figure 2.4: Fraction of Employees that Expect to be Chosen by the Principal, by Productivity

	Part 1	**Part 2**						
		Round 1	2	3	4	5	6	7
Communication treatment								
by A^H	0.93	0.83	0.74	0.79	0.78	0.80		
by A^L	0.10	0.21	0.19	0.19	0.17	0.22		
2-sample test of: z	8.91	6.69	5.96	6.50	6.51	5.66		
proportions; p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01		
No-communication treatment								
by A^H	0.93	0.75	0.69	0.64	0.74	0.71	0.78	0.67
by A^L	0.03	0.07	0.17	0.14	0.09	0.16	0.17	0.10
2-sample test of: z	9.66	7.54	5.62	5.53	7.16	6.00	6.51	6.29
proportions; p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Difference in the *communication* and *no-communication* treatment								
Diff. A^H: MW-test: z	0.00	0.92	0.62	1.85	0.43	1.06		
p-value	1.00	0.36	0.54	0.06	0.66	0.29		
Diff. A^L: MW-test: z	1.47	2.15	0.24	0.75	1.38	0.92		
p-value	0.14	0.03	0.80	0.45	0.17	0.36		

Table 2.6: Fraction of Employees that Expect to Be Chosen by the Employer

Results are reported in Figure 2.4 and Table 2.6. When asking the agents about their (unincentivized) expectations of being chosen, we find that in the *no-communication* treatment,the 93.10% of high productive agents declare that they expected to be chosen. The "opposite" is true for low productive agents: only the 3.44% of the low productive workers expected to be chosen. The same holds for the *communication* treatment: the 93.10% of high productive expected to be chosen, and the 10.34% of the low productive workers expected to be chosen. These results are in line with the incentivized answers to the question about the expected percentage of employers that choose the high productivity agent, see Table 2.7. In the *no-communication* treatment, on average the subjects expect that the employer chooses in 81% of the cases the high productivity agent. The distribution of the relative guesses of high productive and low productive agents does not differ (78% and 83% respectively, Fisher exact test, $p = 0.62$ and Two-sample Kolmogorov-Smirnov test, $p = 0.28$). In the *communication* treatment, on average the subjects expect that the employer chooses in 76% of the cases the high productivity agent; also here the distribution of the relative guesses of high productive and low productive agents does not differ significantly (78% and 75% respectively, $p = 0.98$ and $p = 1.00$). Yet, the difference in the percentages between treatments is driven by the expectations of the chosen low productivity employees (Mann Whitney test, $p < 0.01, z = 2.72$). Hence, it seems that the surprise effect plays a role: the chosen low productivity workers should be more surprised about being chosen than the chosen high productivity workers. Yet, this does not explain our results (it should go in the same direction as the entitlement effect, and there is no reason to assume that the surprise effect should be stronger in the *communication* treatment than in the no-communication treatment - all on the contrary, since here, while reading the message, not only the agents could get used to the idea of being chosen, in addition, they also got an explanation.

Hence, we do not find evidence for the "surprise effect" of being chosen driving our results, since we do not find evidence in the *no-communication* treatment for such mechanism being at work, and there is no reason why in this case such effect should not be present (being present in the *communication* treatment). On the contrary, while in the *communication* treatment the percentage of principals expected to choose low productivity workers among the low productivity workers is *higher* than such percentage in the *no-communication* treatment, here the low productivity workers do provide a higher effort than the high productivity workers. Therefore, while we cannot rule the surprise effect playing a role, if it does play a role as anticipated, then, excluding such, we should find the low productivity workers to provide a significantly lower effort than the high productivity workers in the *no-communication* treatment, and there are no reasons to expect such.

Another reason for why the effort provided among the high and low productivity agents does not differ in the *no-communication* treatment might be different reference points. Do our employees hold different reference points about the level of effort they provide? If the low productivity agents think they provide a significantly higher effort than the high productivity agents, then, while beliefs are still

	Communication treatment **Part 1**	*No-communication* treatment **Part 1**	Difference Mann Whitney test
by A^H	78%	78%	$z = 0.46, \quad p = 0.64$
by A^L	75%	83%	$z = 2.72, \quad p < 0.01$
Fischers exact test, p-value	0.98	0.62	
2 sample KS test, p-value	1.00	0.28	

Table 2.7: Fraction of Employers that is Expected to Have Chosen the High Productivity Employer

not in equilibrium, at least the same reasoning is at work. To investigate this issue, we elicited incentivized beliefs by part of the agents about the expected average level of effort chosen by part of the other agents in the session[18]. We do not find evidence of such mechanism being at work in any of the treatments:

Result 7 *In all three treatments, the low productivity employees expect their own types to provide an on average similar effort as the high productive types will provide. The same is true for the high productivity types.*

In the *no-communication* treatment, the low productive employees who are selected (N=5) believe that other low productive types selected will provide on average an effort of 4.00 while they think that the high productive types selected will provide on average an effort of 3.80 (Mann Whitney test, $z = 0.11$ $p = 0.92$). In the *communication* treatment, the low productivity workers who are selected (N=9) expect on average an effort of 5.80 from the other low productivity workers that are chosen, and on average an effort of 6.21 from the high ones ($z = 0.13$, $p = 0.89$). In the *random device* treatment, the low productivity workers who are selected expect on average an effort of 3.01 from the other low productivity workers that are chosen, and on average an effort of 3.52 from the high ones ($z = 0.55$, $p = 0.58$). Similar results hold for the high types that are chosen[19].

2.4 Conclusion

To summarize, our results confirm our hypothesis regarding the existence of a behavioral explanation based on a reciprocity concern induced by mediocrity, that

[18] In the treatment *no-communication*, we elicited the ex-post beliefs only for two sessions, for a total of 10 selected employees, of which 15 were high productive and 5 selected low productive employees. On the other treatment, we elicited ex-post beliefs in all the sessions.

[19] In the *no-communication* treatment, the high type employees who are selected (N=15) believe that other high productive types selected will provide an average effort of 3.43 while they think that the low productive types selected will provide an average effort of 4.17 (Mann Whitney test, $z = 0.79$ $p = 0.43$). In the *communication* treatment, the high productive employees who are selected (N=29) believe that other high productive employees selected will provide an average effort of 4.49, while they think that the low productive employees selected will provide an average effort of 5.01 ($z = 0.75$, $p = 0.45$). In the *random device* treatment, the high productive employees who are selected (N=17) believe that other high productive employees selected will provide an average effort of 2.92, while they think that the low productive employees selected will provide an average effort of 2.94 ($z = 0.04$, $p = 0.97$).

leads to low productivity agents being selected into positions, even when (ex-ante) more qualified candidates are available.

There are however many interesting open questions for future research.

First, we only analyze a game between one principal and two agents. Therefore the choice of the principal was simply whether to select the high productive agent or the low productive one. When more than two agents with different productivities belong to the set of candidates, the choice of the principal is more complex. In fact, the principal faces a trade-off between the benefit of an increase in the amount of ex-post effort induced by the higher reciprocity concern of less ex-ante productive agents and the direct cost of hiring a less productive agent. Hence, is it more profitable for the principal to select the second ranked individual, who already feels indebted towards the principal and is quite productive or should she choose a *much* lower ranked individual who is *much* less productive but should be extremely grateful towards the principal?

Second, how does the reciprocity concern due to a different entitlement interact with a reciprocity concern due to the classical gift-exchange effect ? If the principal may not only select the agent to perform a task, but also may decide about the size of his wage, which is the most profitable way of inducing a higher non-contractible effort? Selecting a less ex-ante productive agent or paying a higher wage to the ex-ante most productive one?

Third, while we find that the presence of entitlement based reciprocity may lead to inefficiencies ex-ante, when low productivity agents are employed instead of available high productivity ones, we also find that, under some circumstances, this inefficiency persists also ex-post (in the *no-communication* treatment). Yet, this is not the only reason why inefficiencies may arises. It is important to keep in mind that in our setting, agents could only reciprocate by means of an increase in effort. However often agents may reciprocate not only by increasing their effort, but also transferring (private) benefits to the principal. While the literature on reciprocity in general states that reciprocity leads to an increase in effort, real world (employment) relationships are not limited to the exchange of labor and work, but also include other types of interaction. Hence, it is quite intuitive to imagine that a worker, instead of reciprocating with an increase in effort, reciprocates in different ways - by, for example, paying the lunch for the boss; giving the superior a ride in the car, or even just transferring some money to his benefactor. If it were the case that in such setting reciprocity results in an increase in private transfers and not in an increase in effort, this would explain why, up to now, results in field experiment on the fair wage hypothesis were very mixed (for an overview, see Hennig-Schmidt, Rockenbach, and Sadrieh (2010)). Does reciprocity only foster mediocrity or does it also induce (more) corruption?

References

Abeler, J., S. Altmann, S. Kube, and M. Wibral (2009): "Gift Exchange and Workers' Fairness Concerns - When Equality Is Unfair," Discussion Papers 2009-11, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.

Akerlof, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 97(4), 543–69.

Anderson, S. (2011): "Caste as an Impediment to Trade," *American Economic Journal: Applied Economics*, 3(1), 239–63.

Bertrand, M., F. K. A. S., and D. Thesmar (2007): "Politicians, Firms and the Political Business Cycle: Evidence from France," Working paper, Chicago.

Billett, M. T., and Y. Qian (2008): "Are overconfident CEOs born or made? Evidence of self-attribution bias from frequent acquirers," *MANAGEMENT SCIENCE*, 54(6), 1037–1051.

Bolton, G. E., and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90(1), 166–193.

Bramoulle, Y., and S. Goyal (20011): "Favoritism," Discussion paper.

Brandts, J., W. Guth, and A. Stiehler (2006): "I want YOU! An experiment studying motivational effects when assigning distributive power," *Labour Economics*, 13(1), 1–17.

Charness, G. (2004): "Attribution and Reciprocity in an Experimental Labor Market," .

Charness, G., G. R. Frechette, and J. H. Kagel (2004): "How Robust is Laboratory Gift Exchange?," *Experimental Economics*, 7(2), 189–205.

Charness, G., and P. Kuhn (2007): "Does Pay Inequality Affect Worker Effort? Experimental Evidence," *Journal of Labor Economics*, 25, 693–723.

Charness, G., and M. Rabin (2002): "Understanding Social Preferences With Simple Tests," *The Quarterly Journal of Economics*, 117(3), 817–869.

Charness, G., A. Rustichini, and J. van de Ven (2011): "Overconfidence, self-esteem, and strategic deterrence," Discussion paper, Working Paper.

Crawford, V. (1998): "A Survey of Experiments on Communication via Cheap Talk," *Journal of Economic Theory*, 78(2), 286–298.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–51.

Deresiewicz, W. (2008): "The Disadvantages of an Elite Education," *Blog: The American Scholar*.

Dufwenberg, M., P. Heidhues, G. Kirchsteiger, F. Riedel, and J. Sobel (2008): "Other-Regarding Preferences in General Equilibrium," CEPR Discussion Papers 6815.

Dufwenberg, M., and G. Kirchsteiger (2004): "A theory of sequential reciprocity," *Games and Economic Behavior*, 47(2), 268–298.

Durante, R., G. Labartino, and R. Perotti (2011): "Academic Dynasties: Decentralization and Familism in the Italian Academia," Working Paper 17572, National Bureau of Economic Research.

Fahr, R., and B. Irlenbusch (2000): "Fairness as a constraint on trust in reciprocity: earned property rights in a reciprocal exchange experiment," *Economics Letters*, 66(3), 275–282.

Falk, A., and A. Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39–58.

Farrell, J., and M. Rabin (1996): "Cheap Talk," *Journal of Economic Perspectives*, 10(3), 103–118.

Fehr, E., and S. Gächter (1998): "How effective are trust-and reciprocity-based incentives," *Economics, vlues, and organization (BOOK)*, Avner Ben-ner and Louis Putterman.

Fehr, E., S. Gachter, and G. Kirchsteiger (1997): "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica*, 65(4), 833–860.

Fehr, E., E. Kirchler, A. Weichbold, and S. Gaechter (1998): "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets," *Journal of Labor Economics*, 16(2), 324–51.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108(2), 437–59.

Fehr, E., and K. M. Schmidt (1999): "A Theory Of Fairness, Competition, And Cooperation," *The Quarterly Journal of Economics*, 114(3), 817–868.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Gächter, S., and C. Thöni (2010): "Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis," *Journal of Economic Behavior and Organization*, 76(3), 531 – 543.

Gaechter, S., D. Nosenzo, and M. Sefton (2010): "The Impact Of Social Comparisons On Reciprocity," Discussion Papers 2010-10, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.

Gambetta, D., and G. Origgi (2009): "L-worlds: The curious preference for low quality and its norms," .

Gneezy, U. (2004): "Do High Wages Lead to High Profits? An Experimental Study of Reciprocity Using Real Effort," *Working Paper, Graduate SChool of Business, University of Chicago.*

Gneezy, U., and J. A. List (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74(5), 1365–1384.

Green, J. R., and N. L. Stokey (1980): "A Two-Person Game of Information Transmission," Discussion Papers 418, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

Greiner, B. (2003): "An Online Recruitment System for Economic Experiments," *Forschung und wissenschaftliches Rechnen*, 63, 79–93.

Güth, W., M. Königstein, J. Kovács, and E. o. Z.-M. o (2001): "Fairness Within Firms: The Case Of One Principal And Multiple Agents," *Schmalenbach Business Review (sbr)*, 53(2), 82–101.

Hannan, R. L., J. H. Kagel, and D. V. Moser (2002): "Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences, and Effort Requests on Behavior," *Journal of Labor Economics*, 20(4), 923–951.

Hennig-Schmidt, H., B. Rockenbach, and A. Sadrieh (2010): "In Search Of Workers' Real Effort Reciprocity-A Field and a Laboratory Experiment," *Journal of the European Economic Association*, 8(4), 817–837.

Koukoumelis, A., M. V. Levati, and J. Weisser (2009): "Leading by words: A voluntary contribution experiment with one-way communication," Jena economic research papers 2009,106, Jena.

Kräkel, M. (2009): "Competitive Careers as a Way to Mediocracy," Bonn econ discussion papers, University of Bonn, Germany.

Kramarz, F., and D. Thesmar (2007): "Social Networks in The Boardroom," CEI Working Paper Series 2006-20, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.

Kube, S., M. A. Marechal, and C. Puppe (2010): "The Currency of Reciprocity - Gift-Exchange in the Workplace," *SSRN eLibrary*.

Levine, D. K., F. Weinschelbaum, and F. Zurita (2010): "The Brother-In-Law Effect," *International Economic Review*, 51(2), 497–507.

Mas, A., and E. Moretti (2009): "Peers at Work," *American Economic Review*, 99(1), 112–45.

Montinari, N. (2011): "The Dark Side of Reciprocity," (2011-052).

Pande, R. (2003): "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India," *American Economic Review*, 93(4), 1132–1151.

Prendergast, C., and R. H. Topel (1996): "Favoritism in Organizations," *Journal of Political Economy*, 104(5), 958–78.

Rabin, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83(5), 1281–1302.

Sobel, J. (2005): "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43(2), 392–436.

Appendix “Mediocrity leading to Reciprocity”

2.A Instructions *No-Communication* Treatment

Welcome! You are about to participate in an experiment funded by the Max

Planck Institute of Economics. Please switch off your mobile and remain quiet. It is strictly forbidden to talk to the other participants. Whenever you have a question, please raise your hand and one of the experimenters will come to your aid. You will receive 2.50 Euros for showing up on time. Besides this, you can earn more. The show-up fee and any additional amounts of money you may earn will be paid to you in cash at the end of the experiment. Payments are carried out privately, i.e., the others will not see your earnings.

During the experiment we shall speak of ECUs (Experimental Currency Unit) rather than Euros. The conversion rate between them is 10 ECUs = 1 euro.

This means that for each ECU you earn you will receive 0.1 Euro. To simplify, in the following we are only speaking of male participants. This is to be understand gender neutral.

The experiment consists of two parts. The instructions for the first part follow on the next page. The instructions for the second part will be distributed after all participants have completed the first part. All instructions are identical for all participants and we read them aloud such that you can verify this.

2.A.1 Instruction of Part 1

Groups Formation

In this experiment you will be matched with two other participants to form a group of three persons. The three group members will interact with each other just once. We will refer to each group as a firm, and to the three group members as Employer, Employee A and Employee B:

- with 1/3 probability you will be the Employer;
- with 1/3 probability you will be Employee A;

- with 1/3 probability you will be Employee B.

This means that each participant has the same probability to be selected as Employer, Employee A and Employee B. You will be assigned to a group and to a role entirely at random. The computer will inform you of your role before the decision-making part of the experiment begins. No one of the participants will know at any point of the experiment the identity of the other people in his group. Therefore, all decisions are made anonymously.

Decisions within a Firm

The structure of the decision-making within each firm is as follows. There is a job to be offered. First, the employer

- chooses one of the two employees to hire.

Then the selected employee

- learns whether the employer selected him, and
- he chooses an effort level (effort): an integer number from 0 to 10 (included).

The employee who is not selected receives an unemployment benefit equal to 10 ECUs.

Distribution of earnings within a Firm

Earnings within the Firm are determined according to the following rules:

Earnings for the employer:

The Employer receives revenue from the effort chosen by the selected employee and incurs costs from the wage paid to him. The revenue produced by the effort chosen from the selected employee is the following:

If the selected worker is **employee A**, then the revenue produced equals 50 times the effort chosen by the employee, plus 0.05, times 0.10 . Therefore, the employer's earnings are:

Employer's earnings = [50*(Effort chosen by employee A+0.5)*0.10]

if effort >0

and

Employer's earnings = 0

if effort = 0

Table 2.A.1 reports the earnings (in ECUs) for the employer when employee A is selected for each effort level.

chosen effort	0	1	2	3	4	5	6	7	8	9	10
income of employer if selected employee is employee A	0	7.5	12.5	17.5	22.5	27.5	32.5	37.5	42.5	47.5	52.5

If the selected worker is **employee B**, then the revenue produced equals 50 times 0.10 times the effort he chooses.

Therefore, the employer's earnings are:

Employer's earnings = [50 * (Effort chosen by employee B) * 0.10]

if effort >0

and

Employer's earnings = 0

if effort = 0

Table 2.A.1 reports the Earnings (in ECUs) for the employer when employee B is selected for each effort level.

chosen effort	0	1	2	3	4	5	6	7	8	9	10
income of employer if selected employee is employee B	0	5	10	15	20	25	30	35	40	45	50

Please note that in both cases, the employer's earnings increase with higher effort levels. For example,

- when the Employer selects the employee A and
  - chooses effort = 0, the earnings for the employer are 0 ECUs;
  - if instead he chooses effort =5, the earnings for the employer are 27.5 ECUs.
  - if instead he chooses effort =9, the earnings for the employer are 47.5 ECUs.
- When the Employer selects the employee B and employee B:
  - chooses effort = 3, the earnings for the employer are 15 ECUs;
  - if instead he chooses effort =6, the earnings for the employer are 30 ECUs.
  - if instead he chooses effort =8, the earnings for the employer are 40 ECUs.

Please note that the numbers used in all the examples were selected arbitrarily. They are not intended to suggest how you might decide.

Earnings for the selected employee:

Once the employer chooses the employee, the selected employee chooses the level of effort to exert and consequently his earnings as shown in table 2.A.1:

effort	0	1	2	3	4	5	6	7	8	9	10
income of chosen empl.	0	50	49	48	46	44	42	40	38	35	32

For example,

- when the selected employee:
  - chooses effort=2, his earnings are 49 ECUs
  - chooses effort = 5, his earnings are 44 ECUs; if instead
  - he chooses effort =8, his earnings are 38 ECUs.

Earnings for the non selected employee:

The employee who is not selected receives an unemployment benefit equal to 10 ECUs.

What happens next?

- Before the experiment starts, in next screens, you will be asked to answer a few questions.
- When the experiment starts you will be informed about whether you are an employer or an employee in this experiment. In case you are an employee, it will be specified whether you are employee A or employee B.

Summary

We will now briefly summarize the content of the instructions you have just read. At the beginning of the experiment you will be randomly matched with two other participants to form a group of three people and you will be randomly assigned a role within this group which we will call "firm". You will be either the employer or employee A or employee B. The structure of the decision-making

within each firm is as follows.

- First, the employer chooses to hire one of the employees. The employer's earnings increase with higher effort levels of the selected employee.

- Next, the selected employee learns that he has been chosen. Then he chooses an effort level (Effort). The selected employee's earnings decrease with higher effort.
- The non selected employee receives an unemployment benefit = 10 ECUs

Please note that the decision task of this part of the experiment will be performed only once. Please, raise your hand if you have any questions.

Hypothetical examples for demonstration purposes

1. Assume that the employer chooses employee B.

The employee B chooses the effort = 9. This situation results in the following earnings:

Employer's earnings: The employer receives revenue from the effort of the employee A, i.e.: 50*9*0.10 = 45 ECUs. The earnings of the employer are 45 ECUs.

Employee A's earnings: The employee A receives an unemployment benefit of 10 ECUs. The earnings of employee A are 10 ECUs.

Employee B's earnings: The employee B receives a wage of 50 ECUs and chooses an effort = 9. The earnings of employee B are 35 ECUs.

2. Assume that the employer chooses Employee A.

The employee A chooses the effort =8. This situation results in the following earnings:

Employer's earnings: The employer receives revenue from the effort of the employee A, i.e.: 50*(8+0.5)*0.10 = 42.5 ECUs. The earnings of the employer are 42.5 ECUs.

Employee A's earnings: The Employee A receives a wage of 50 ECUs and chooses an effort = 8. The earnings of employee A are 38ECUs.

Employee B's earnings: The employee B receives an unemployment benefit of 10 ECUs. The earnings of employee B are 10 ECUs.

2.A.2 Instruction of Part 2

In this part you will face a situation similar to that encountered in the first part.

As before:

- you will be matched with two other participants;
- the three group members will be identified by as employer, employee A and employee B.
- your role is the same as in part 1 (i.e., you will be the employer, the employee A or the employee B, if you previously were, respectively, the employer, the employee A or the employee B);
- the employer in your group have to chose one between the two employee.
- The selected employee then chooses an effort level between 0 and 10.
- The employee who is not selected gets an unemployment benefit = 10 ECUs.
- The earnings within the firm are divided as in part 1.

But now

- This part 2 consists of 7 rounds. In every round you will be placed in a NEW group of three persons (i.e., the two participants you will be matched with are different ones in every round);
- You will never be informed of the identity of the participants you will be matched with, but it will never be the same in any group.
- How we determine your earning in part 2
  - Once part 2 is over, one participant will be randomly selected.
  - This participant will determine which of the 7 rounds of part 2 is paid, by making a random draw from the urn containing 7 balls (numbered 1 to 7).
  - The earnings corresponding to this round will be converted to Euros and paid out in cash.
  - The outcome of the urn-draw will apply to all the participants. Thus, only 1 of the 7 decisions you will make in this part will be paid out.
- Obviously each decision has an equal chance of being used in the end. So, think carefully when making your choice in each round!

2.A.3 Instructions *Communication* Treatment

The above instructions - from section 2.A onwards - were modified such as to include passages that explained how the communication worked. The following passages were added/changed (full instructions are available upon request by the authors), plus minor modifications.

Decisions within a Firm

The structure of the decision-making within each firm is as follows. There is a job to be offered.

First, the employer

- chooses one of the two employees to hire.
- Then, he can send one message to the chosen employee.

Then the selected employee

- learns whether the employer selected him,
- reads the message the employer has sent him, and
- he chooses an effort level (Effort): an integer number from 0 to 10 (included).

4. Communication

The employer may send one message to the employee he has selected. In this message, he can say anything he likes, including a suggestion what he think is the best approach to the experiment, what he intends with his selection, or what he would like the other to do.

The chosen employer can read the message before choosing the level of effort

However, there are two restrictions on the types of messages that he may send.

- First, he may not send a message that attempts to identify himself to other group members. Thus, he may not use his real name, nicknames, or self-descriptions of any kind "Tom Smith here," "I'm the guy in the red shirt sitting near the window," "It's me, Sandy, from French class," or even "As a woman [Latino, Asian- American, etc.], I think. . ."). To make sure that the rule of anonymity is adhered to, each message will be screened by a monitor who is a member of the experiment team before it is seen by the other member of your group.
- The second restriction is that there must be no threats or promises pertaining to anything that is to occur after the experiment ends.

To make sure that neither oft he two restrictions is violated, all messages are going to be read from a member oft he experimental staff before they are shown on the screen of the chosen employee.

If a message violates one of the restrictions, it is not going to be sent to the employee and the employer does not receive payment for the experiment.

The employer a maximum of 3 minutes available to write a message. A clock will show you how much time you have left in the communication period.

2.A.4 Instructions *Random Device* Treatment

The instructions - from section 2.A until 2.A.3 - were modified such as to include passages that explained how the random device worked. The following passages were added/changed (full instructions are available upon request by the authors).

Decisions within a Firm

The structure of the decision-making within each firm is as follows. There is a job to be offered. First, a random mechanism selects one of the two employees to work for the employer.

Then the selected employee

- learns whether the random mechanism selected him, and
- he chooses an effort level (Effort): an integer number from 0 to 10 (included).

5. Summary

We will now briefly summarize the content of the instructions you have just read. At the beginning of the experiment you will be randomly matched with two other participants to form a group of three people and you will be randomly assigned a role within this group which we will call "firm". You will be either the employer or employee A or employee B.

The structure of the decision-making within each firm is as follows.

- First, a random mechanism chooses one of the employees to work for the employer. The employer's earnings increase with higher effort levels of the selected employee.
- Next, the selected employee learns that he has been chosen. Then he chooses an effort level (Effort). The selected employee's earnings decrease with higher effort.
- The non selected employee receives an unemployment benefit = 10 ECUs.

2.A.5 Control Questions

The questions on the next screens will help us to understand if you have understand the instructions. Attention: the numbers are chosen randomly. They should not suggest you a way of playing the game.

Think about the role of the employer.

How much is the income of the employer, when he selects employee A, and employee A chooses an effort of 2? How much is the income of the employer, when he selects employee B, and employee B chooses an effort of 9?

Think about the role of the employee A.

How much is the income of the employee A, when he is selected by the employer and chooses an effort of 6? How much is the income of the employee A, when he is selected by the employer and chooses an effort of 0?

Think about the role of the employee B.

How much is the income of the employee B, when he is selected by the employer and chooses an effort of 8? How much is the income of the employee B, when he is selected by the employer and chooses an effort of 1?

2.A.6 Post-Experimental Questionnaire

QUESTIONS FOR EMPLOYER

Think to the FIRST part of the experiment. In this part you had to take just one decision.

1. Which employee have you chosen in the first part? (1)
2. Could you explain us why did you choose this employee and not the other one? (2)
3. Would you take the same decision again? If yes, why? If not, why not? (3)
4. Which effort would you have chosen in the first part if you were employee A and you were chosen? (4)
5. Which effort would you have chosen in the first part if you were employee B and you were chosen? (5)

Now think to the SECOND part of the experiment. In this part you had the same decision situation as in part 2 but you face this situation for 9 rounds.

1. In part 2, how many times (approximately) out of the 7 rounds did you choose employee A? (6)
2. How many times did you choose employee B? (7)
3. Could you motivate your choice? (8)
4. Would you take the same decisions again? If yes, why? If not, why not? (9)

QUESTIONS FOR CHOSEN EMPLOYEE

Think to the FIRST part of the experiment. In this part you had to take just one decision.

1. Do you remember the effort level you chose in the first part? If yes, how high was it? (11)

2. Would you take the same decision again? If yes, why? If not, why not? (12)

3. Could you explain us why did you choose this effort level? (13)

4. Did you expect to be chosen by the employer? (14)

5. Why do you think the employer has chosen you (and not the other employee)? Which employee would you have chosen if you had been in the role of the employer in part 1? (15)

Now think to the SECOND part of the experiment. In this part you and the other participants faced the same decision situation as in part 1 but this situation was repeated for 9 rounds.

1. Which effort did you choose in the later rounds when selected? (16)

2. Why? (17)

3. Would you take the same decision(s) again? If yes, why? If not, why not? (18)

4. Which employee would you have chosen if you had been in the role of the employer in part 2? Would you have chosen always the same type of agent (i.e., always agent A / always agent B)? (19)

QUESTIONS FOR EMPLOYEE THAT WAS NOT CHOSEN

Think to the FIRST part of the experiment.

1. Why do you think the employer has not chosen you in the first part? (21)

2. Did you expect to be chosen in the first round? (22)

3. Which effort level would you have chosen in the case you would have been selected? (23)

4. Which employee would you have chosen if you had been in the role of the employer? (24)

Think to the SECOND part of the experiment. In this part you and the other participants faced the same decision situation as in part 1 but this situation was repeated for 9 rounds.

1. In the cases you where chosen in one or several rounds - why do you think the employer has chosen you (and not the other employee)? (25)
2. Which effort level have you chosen? (26)
3. Would you take the same decision(s) again? If yes, why? If not, why not? (27)
4. Which employee would you have chosen if you had been in the role of the employer? Would you have chosen always the same type of agent (i.e., always agent A / always agent B)? (28)

QUESTIONS FOR EVERYBODY

1. Is there anything else you want to tell us? (10)
2. Please indicate your gender (Female/Male)

Chapter 3

Shifting the Blame when Delegating to a Powerless Intermediary

3.1 Introduction

Why do managers hire consultants to announce and implement layoffs, as well as outsource final production decisions?[1] Why might a company like Merck sell the patent to a cancer drug to another company (Ovation), likely anticipating that this will significantly inflate the price, instead of doing so directly?[2] While firms hire agents for reasons of efficiency, commitment or incentive provision (Aghion and Tirole, 1997; Bolton and Dewatripont, 2005; Schelling, 1960), responsibility-shirking and blame-shifting provide an additional rationale for delegating decision rights. Previous studies (Coffman, 2011; Fershtman and Gneezy, 2001; Bartling and Fischbacher, 2011; Hamman, Loewenstein, and Weber, 2010)[3] have found that players in experimental dictator games may avoid censure and costly punishment by delegating the allocation decision to an intermediary. However, because those studies provide the intermediary either with options with differing degrees of fairness, or with no choice whatsoever, it is not clear whether the willful choice of the intermediary of a selfish outcome over a less-selfish outcome is necessary for blame shifting, or whether the mere presence of a nominal intermediary is sufficient, regardless of her power to influence the fairness of the outcome.

We conducted an experiment in which a dictator may choose from one of three allocations of $20, including an equal split, among her four-person group.[4] She may also delegate her decision to an intermediary, but only in such a way that limits the intermediary's choice set to the two unfair allocations. Those differ merely in which of the two passive group members (the recipients) are hurt most. While in the first treatment (*Choice*) the intermediary has to choose among these two options, in another treatment (*Random*) he just initiates a random process that

[1]In the 2009 movie "Up In The Air", George Clooney portrays such a professional corporate downsizer. Hamman, Loewenstein, and Weber (2010) list other sources of illustrative examples, including "Letting a Stranger Do the Firing", Paul Brown 2007, The New York Times Nov. 10 and O'Rourke (1997), who considers a case study of Nike's working standards in outsourced plants as well as the accounting firm's labor and environmental audit.

[2]Coffman (2011) and Paharia, Kassam, Greene, and Bazerman (2009) provide the background on Merck. Merck sold the patent of a cancer drug to a small company, Ovation, that subsequently raised the price of this drug by more than a factor of ten. Since Merck had sold the rights for the drug, the public generally did not perceive it to be responsible for this increase in price. However, after paying a high enough purchase price, Ovation might have been compelled to increase the drug price to avoid a loss. Thus, by demanding a high enough price for the rights to the cancer drug, Merck might effectively limit Ovation's strategy space to 'unfair' price increases. Of course, this is conditional on Ovation choosing to buy the patent. An area ripe for future research is an agent's decision of whether or not to accept the intermediary role.

[3]Apart from the research discussed below, Fershtman and Gneezy (2001) and Hamman, Loewenstein, and Weber (2010) provide evidence of how dictators may use delegation to avoid negative judgment for harmful or selfish behavior. Fershtman and Gneezy examine the effects of strategic delegation in ultimatum games, showing that, when the proposer uses a delegee, his share increases. Hamman, Loewenstein, and Weber show that recipients receive significantly less when allocation decisions are made by agents that are hired by a principal to make a sharing decision on her behalf.

[4]In the following, we will refer to the dictator as "she", while the intermediators and the recipients will be referred to as "he".

selects one of the two. Finally, for a cost, the two receivers can reduce the earnings of some or all of the other players in the group, conditioning their deductions on the choices of the dictator and intermediary. Though the experiment instructions avoid all language pertaining to punishment, we follow Bartling and Fischbacher (2011) (hereinafter, BF) by interpreting the deductions chosen by a recipient as a form of punishment, reflecting how he attributes responsibility to the other players for the unfair outcome.

Our design most closely resembles the 'delegation and punishment' condition of BF, who find find that the dictator's punishment given an unfair allocation is greatly reduced when the dictator passes the choice to the intermediary. While in the BF design, the act of delegating forwards the whole action set to intermediary, in our design the intermediary can only choose between the unfair options. Thus, it is transparent that a dictator who delegates intends an unfair allocation to be obtained and that the intermediary has no say in whether or not the fair outcome was chosen.

We show that delegating shifts blame away from the dictator and *onto* the intermediary, even when the intermediary only faces unfair outcome choices. Given that the final allocation is unfair, the dictator is punished less if she delegates than if she had chosen directly. On the contrary, the intermediary's punishment increases after delegation. However, the latter is only true in the *Choice* treatment. In the *Random* treatment, punishment for the intermediary does not increase significantly. This shows that the intermediary escapes further blame if he is *completely* unable to influence the outcome. Yet, in any case we show that the results of BF are robust to altering and restricting the intermediary's choice set; delegating is profit maximizing for the dictator in both treatments. Even though the intermediary, if called upon, is powerless to implement the fair outcome, the dictator can effectively shift responsibility when delegating — as long as he has *some* choice. Hence, our findings complement and extend previous results on intermediation and blame-shifting. As in the *asymmetric* condition of BF, in which the dictator cannot choose the unfair outcome directly, delegation can be considered the least kind action available to the dictator (along with choosing an unfair allocation directly). This makes it all the more striking that the dictator can effectively shift blame to the intermediary by delegating and suggests that intentions alone cannot explain the punishment behavior. The fact that the intermediary is punished, even when his choice cannot possibly increase the likelihood of an unfair allocation, challenges the responsibility measure proposed by Bartling and Fischbacher (2011), which ties responsibility to the extent to which a choice increases the probability of an unfair allocation.

Instead, our findings, like those of Coffman (2011), suggest that the mere fact that the dictator does not *directly* determine the final allocation is sufficient for blame-shifting. Coffman (2011) allows a dictator to split money directly with a recipient or to pass any part of the surplus to an intermediary, who can share this amount with the recipient. He allows a fourth party to punish the dictator (without cost) and finds that the dictator's punishment is reduced when she implements an unfair outcome through the intermediary, even when she keeps the entire surplus, leaving the intermediary with nothing to share. Our findings reinforce Coffman's conclusion that sanctions for harmful behavior are reduced when the responsible party does not directly interact with the victim. We extend his results to a context in which the intermediary's interests are aligned with those of the dictator and show robustness to a different (costly) punishment technology. Furthermore, by allowing punishment of the intermediary, we identify the conditions under which blame actually shifts *onto* him. Thus, even when the intermediary can not in any way be viewed as complicit in the unfair behavior, she may be punished for it.

Not only did the participants in the intermediary role lack the ability to enforce a fair outcome, instead of choosing or contracting on their role as an intermediary, it was assigned to them. The fact that intermediaries are still punished for the unfair outcome in the *Choice* treatment raises the question of why an agent would agree to perform a blame-worthy task for a principal looking to avoid punishment. The stated beliefs of our participants suggest an explanation. While participants' beliefs are qualitatively in tune with the observed punishment patterns, the dictators and particularly the intermediaries underestimate the extent to which blame is shifted onto the intermediary.

Hence, given the reduction in punishment, a money-maximizing dictator with rational expectations has the incentive to delegate. Furthermore, by delegating a dictator might also avoid the uncomfortable decision of having to decide whom to harm most (see Dana, Cain, and Dawes, 2006; Dana, Weber, and Kuang, 2007; Grossman, 2010; Lazear, Malmendier, and Weber, 2009), while avoiding the cost of choosing the equal split. However, we find that roughly as many dictators directly choose an unfair allocation as use the intermediary. Surprisingly, the degree to which the dictator perceives intermediation as a way of avoiding punishment does not significantly increase the likelihood that the dictator will delegate. We conclude that delegating one's harmful action to someone else might "feel wrong" or violate a personal rule, imposing a psychological cost.

The rest of the paper is organized as follows. Section 3.2 explains our experimental design and the procedures. Section 3.3 presents the results, and 3.4 concludes with a discussion of how outcome- and intentions-based theories of social preferences have difficulty explaining our results.

3.2 Experimental Design

Each group played a once one-shot four-person dictator game over a \$20 surplus, with timing as follows (see also Figure 3.1). First, the dictator chooses between one of three allocations of \$20 between the four players, or she can delegate the decision to the intermediator. The dictator's four choices are presented in Table 3.1. Available allocations included an costly but fair split (\$5 each); an

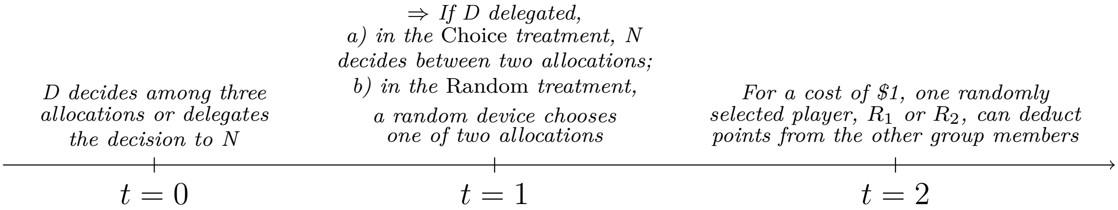


Figure 3.1: Timeline of the *Choice* and the *Random* Treatment

unfair option benefiting the dictator (D) and a second player who was the potential intermediary (N), allocating each \$9 each, while leaving \$2 to one of the remaining recipient players (R_1) and \$0 to the other (R_2); and a second unfair option that differed from the first only by switching the payoffs of R_1 and R_2. The dictator could choose one of these allocations directly or pass the decision to N. In the first treatment, (*Choice*), when the decision was delegated, N could choose either of the two unfair allocations, but not the fair allocation. In the second treatment, (*Random*), he could only click a single button, which caused the computer to randomly select one of the two unfair allocations. We call the split of the \$20 an *allocation*, whereas a terminal history of the decision process, including both the allocation and the player choosing it, is called an *outcome*.

Table 3.1: The Dictator's Four choices.[a]

Choice	Dollars allocated to			
	D	N	R_1	R_2
a	9	9	0	2
b	9	9	2	0
c	5	5	5	5
d	(Pass to N)			

[a] In the *Choice* treatment, the intermediary must choose either a or b if D passes. In the *Random* treatment, a random device selects a or b if D passes.

Having two recipients allows us, like BF, to double the number of punishment observations per group. Importantly, it also allows us to provide the intermediary with a choice between *two* unfair outcomes, instead of only one.

Following BF, we randomly selected R_1 or R_2 to assign punishment. The punisher could pay \$1 to deduct up to \$7 in any combination from any of the other three participants, with the restriction that the resulting payoffs must be non-negative. We used the strategy method to elicit the punishment choices for both R_1 and R_2; each specified a punishment (contingent on him being selected) for each of the five possible outcomes in a randomized order. In contrast, the intermediary only made a decision when called upon after the dictator delegated. We explained to the participants that we increased the show-up fee to \$6 from the usual \$5 so that a person who received \$0 from the dictator game would still have a minimum of \$5 if she chose a non-zero deduction. At the end of the experiment, the deductions specified for the realized outcome by the selected punisher were implemented.

After the experiment had finished, in four of the fourteen sessions in the *Choice* treatment and in all the six sessions in the *Random* treatment we elicited beliefs from the dictators and the intermediaries, with respect to the deduction behavior of the receivers. They had to indicate for each of the five scenarios whether they expect the receiver R_1 to deduct points, and, if yes, how much they expect R_1 to deduct from each of the respective subjects in their group. In the *Random* treatment, when the guessers' response was within +/- \$1 of the true punishment of the R_1 in his/her group, he/she would earn an additional \$5. In the *Choice* treatment, the benchmark was the average punishment level of the R_1s from the first ten sessions.

The experiment was computerized with the software Z-tree (Fischbacher, 2007). We conducted 20 sessions, lasting 30 - 45 minutes each and featuring a total of 236 participants, in between May 2010 and December 2011.[5] We used the online system ORSEE (Greiner, 2003) to randomly recruit participants from the University of Santa Barbara (UCSB) Experimental and Behavioral Economics Laboratory (EBEL) subject pool, largely comprised of UCSB students and staff. Upon arriving at the experiment, participants sat at computer terminals, were given a paper copy of the instructions, and followed along as the experimenter read them aloud. We randomly assigned participants into four person groups, in which the roles of the dictator, intermediator, and the two recipients were randomly assigned. We then gave participants a second packet of written instructions explaining their specific role more in detail. These instructions included exercises designed to verify participants' understanding of the instructions, which the experimenter verified before the decision-making began. Average payment was \$11.09, including the \$6 show-up fee. Instructions as well as screenshots are presented in the appendix.[6]

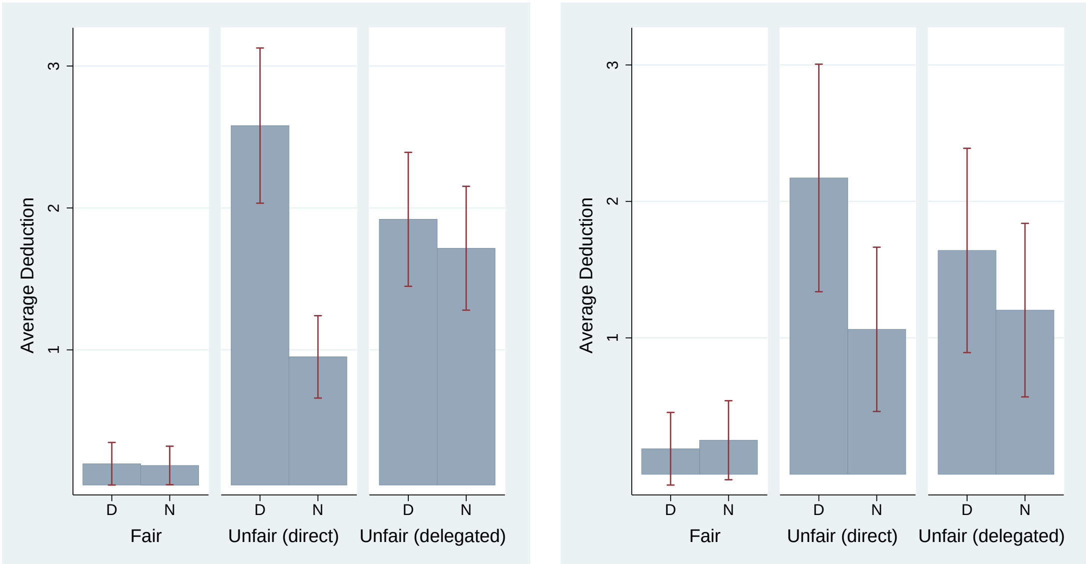
3.3 Results

Our primary interest is in how delegation affects the punishment following an unfair outcome, with the null hypothesis being that delegation does not affect punishment levels and the alternative being that it reduces D's punishment and increases that of N. Hence, we first examine the punishment behavior of the recipients. Figure 3.2 and Table 3.2 show expected punishment—averaged across both recipients— for each outcome by treatment and illustrate our main results. Recipients shift the blame for a delegated unfair outcome away from D, but only

[5]We dropped observations from two subjects who participated in more than one session in the choice treatment, and we dropped observations from four subjects in the random treatment where the program made a mistake.

[6]Full instructions and the software are available from the authors.

in the *Choice* treatment does it fall onto N. In both treatment, delegating is the profit maximizing choice for the dictator. In the following, we provide detailed evidence for these claims.



(a) *Choice* treatment (b) *Random* treatment

Figure 3.2: Average Punishment, by Outcome

Table 3.2: Average punishment and punishment frequency

	Choice ($N = 81$)			*Random* ($N = 32$)		
Outcome	% ded.	Mean deduction from D	Mean deduction from N	% ded.	Mean deduction from D	Mean deduction from N
Fair	9	0.20	0.19	9	0.19	0.25
Unfair direct	53	2.58	0.95	41	2.17	1.06
Unfair delegated	55	1.92	1.72	53	1.64	1.20

Recall that the two unfair allocations differ in which recipient is harmed the most: allocation a gave R_1 a payoff of \$0, as opposed to the \$2 he obtained from allocation b, with the reverse payoffs for R_2. Holding constant whether it was chosen by the intermediary or directly by the dictator, recipients do not appear to distinguish one of these allocations from the other in their punishment decision, in either treatment.

Result 1 *Punishment is not sensitive to which unfair outcome was chosen.*

In the *Choice* treatment, when the dictator chose an unfair allocation directly, 43 (53%) out of the 81 recipients who were *most harmed* (R_1 for $(9, 9, 0, 2)$ and R_2 for $(9, 9, 2, 0)$) chose to incur the $1 punishment cost and exactly the same number of recipients who were *least harmed* did. When the dictator delegated, 47 (58%) of the recipients who were most harmed chose to incur the punishment cost, while 42 (52%) of those who were harmed least chose to punish. A two sample test of proportions cannot reject the hypothesis that these punishment rates are the same ($z = 0.79$, $p = 0.43$).[7] Similar results hold for the *Random* treatment. When the dictator chose an unfair allocation directly, 17 (53%) out of the 32 recipients who were *most harmed* chose to incur the $1 punishment cost, and exactly the same number of the ones that were *least harmed* did. When the dictator delegated, 12 (38%) of the recipients who were most harmed chose to incur the punishment cost, while 14 (44%) of those who were harmed least chose to punish. Again we cannot reject the hypothesis that these punishment rates are the same ($z = 0.51$, $p = 0.63$).

Examining the mean and distribution of the deductions chosen by those who did punish leads to the same conclusion. In the *Choice* treatment a two-sample t-test with unequal variance cannot reject the hypothesis ($t = 0.48$, $p = 0.63$) that the same amount was deducted from the dictator on average by the least harmed ($4.95) and the most-harmed recipient ($4.77) when the dictator directly choose a selfish allocation. A two-sample Kolmogorov-Smirnov test cannot reject the hypothesis that the distributions of deductions are equal ($p = 0.90$). When D delegates and N chooses, $3.40 is deducted on average by the most-harmed recipient and $3.69 is deducted by the least-harmed. We cannot reject the hypothesis that these deductions are equal in means ($t = 0.52$, $p = 0.60$) or distributions (KS, $p = 1.00$). Similarly, in the *Random* treatment we cannot reject the hypothesis ($t = 0.20$, $p = 0.84$) that the same amount is deducted from the dictator on average by the least harmed ($4.00) and the most-harmed recipient ($4.18) when the dictator chooses unfair directly. Similarly we cannot reject the hypothesis that the distributions of deductions are equal (KS, $p = 0.92$). When D delegated and N had a choice, $2.5 was deducted on average by the most-harmed recipient and $4.14 was deducted by the least-harmed. Again we cannot reject the hypothesis that these deductions are equal in means ($t = 1.26$, $p = 0.25$) though they are significantly different in distributions (KS, $p = 0.06$).

Turning to the intermediary's punishment, when D chooses directly in the *Choice* treatment, the mean deduction from N by the most-harmed recipient, $1.81, is not significantly different ($t = 0.13$, $p = 0.90$) from that made by the least-harmed recipient, $1.77, nor are the distributions significantly different (KS $p = 1.00$). When D delegates, the mean deduction from N by the most-harmed recipient, $3.30, is not significantly different ($t = 0.84$, $p = 0.40$) from that made

[7] We can also exclude that there is a significant share of subjects that deduct only when they did receive at least a payoff of 2: 10 subjects do so. However, there are also 7 subjects that deduct only when they receive a payoff of 0: 7 subjects do so. The percentage is not statistically significant (two sample test of proportions, p = 0.44, z = 0.77).

by the least-harmed recipient, \$2.93, while the distributions again are not significantly different (KS $p = 0.98$). Also in the *Random* treatment, when D chooses directly the mean deduction from N by the most-harmed recipient, \$1.53, is not significantly different ($t = 1.24$, $p = 0.22$) from that made by the least-harmed recipient, \$2.47, nor are the distributions significantly different (KS $p = 0.64$). When D delegates, the mean deduction from N by the most-harmed recipient, \$2.08, is not significantly different ($t = 0.73$, $p = 0.47$) from that made by the least-harmed recipient, \$2.71, while also the distributions again are not significantly different (KS $p = 0.88$).

Hence, given that the recipients treat the two unfair allocations identically, for the remaining analysis we average the punishment data for the two unfair allocations within the two categories: unfair allocations chosen directly and unfair allocations that were the result of a delegated choice. We continue by observing that the recipients' deductions are consistent with punishing unfair or harmful behavior:

Result 2 *The recipients punish the unfair outcomes more than the fair outcome.*

Although Figure 3.2 alone convincingly shows that the fair allocation is punished much less than any of the unfair allocations selected by any means, we also provide hypothesis tests that support this claim.

The mean punishment in the *Choice* treatment is higher when either unfair option is chosen compared to when the fair option is chosen. A two-sample t-test with unequal variance cannot reject the hypothesis that the mean deduction for D is higher when the dictator chooses an unfair allocation than when he chooses fair ($t = 8.36$, $p = 1.00$) neither we can reject the hypothesis that the mean deduction is higher when she chooses to delegates than when choosing fair ($t = 6.92$, $p = 1.00$). A similar result holds for the intermediary: the mean deduction following the equal split is significantly smaller than the deduction followed a direct unfair choice ($t = 4.75$, $p < 0.01$). The same is true for the mean deduction following the equal split and the deduction followed a delegated unfair choice ($t = 6.67$, $p < 0.01$). Similarly, we cannot reject the hypothesis that the mean deduction for dictators in the *Random* treatment is higher when the dictator chooses directly an unfair allocation than when choosing fair ($t = 4.62$, $p = 1.00$) nor the hypothesis that the mean deduction is higher when she delegates than when choosing the equal split ($t = 3.73$, $p = 1.00$). For the intermediary, the mean deduction following the equal split is significantly smaller than the deduction followed a direct unfair choice ($t = 2.48$, $p < 0.01$), and this holds as well for the mean deduction following the equal split and the deduction followed a delegated unfair choice ($t = 2.78$, $p < 0.01$).

Furthermore, in the analysis that follows we do not comment on the punishment for the non-punishing recipient, which is negligible, even though punishment of the non-punishing recipient is non-zero. In the *Choice* treatment, the non-punishing recipient faces an average deduction of \$0.03 when the dictator directly chooses an unfair allocation, \$0.10 when she chooses fairly and \$0.02 when she delegates. In the *Random* treatment, the average deductions are \$0.09, \$0.11, and \$0.09, respectively.

Figure 3.2 also shows that although the dictator's allocation is the same for choosing an unfair allocation regardless of how it was chosen, her final payoff is not. Because, given an unfair outcome, delegation reduces the dictator's punishment, she earned the most when she delegated.

Result 3 *In the* Choice *treatment, delegating does not just effectively reduce the blame placed on the dictator, it effectively shifts it onto the intermediary. In the* Random *treatment, delegation does not shift punishment onto the intermediary.*

Given an unfair allocation in the *Choice* treatment, delegation reduces D's average punishment from \$2.58 to \$1.92, a drop that is significant according to both a two-sample, one tailed test of differences in means ($t = 1.82$, $p = 0.04$) and an OLS regression of the dictator's punishment following an unfair outcome, with a dummy for whether the dictator delegated (dummy coeff. $= 0.66, se = 0.17, p < 0.01$). In the *Random* treatment, the drop from \$2.17 to \$1.64 is of similar magnitude, though a smaller sample undermines the significance level in the t-test ($t = 0.97$, $p = 0.17$). However, the regression coefficient remains significant at a low level (dummy coeff. $= 0.53, se = 0.21, p = 0.02$), so we conclude that delegating maximizes expected profit if D has rational expectations.

Conversely, when D delegates, N's average punishment in the *Choice* treatment increases from \$0.95 to \$1.72, allowing us to reject the null hypothesis of equal punishment ($t = 2.91$, $p < 0.01$ and dummy coeff. $= 0.77, se = 0.16, p < 0.01$). In contrast, the increase in N's average punishment in the *Random* treatment from \$1.06 to \$1.20, only a fraction of the jump seen in the *Choice* treatment, is not significant ($t = 0.33$, $p = 0.37$ and coeff. $= 0.14, se = 0.22, p = 0.52$). Furthermore, N's average punishment following delegation is lower in the *Random* treatment than in the *Choice* treatment, with borderline significance ($t = 1.35$, $p = 0.09$). Thus, blame is not shifted when N must randomize.

Table 3.2 also displays punishment frequency by outcome for each treatment. A significantly higher fraction of recipients choose to punish the dictator for an unfair outcome than for the equal split. In the *Choice* condition, the punishment rate rises from 9% to an average of 54% (Two-sample test of proportions, $z = 6.84$, $p < 0.01$), while in the *Random* condition it rises from 9% to an average of 47% ($z = 3.65$, $p < 0.01$). In neither condition did the punishment rate for directly chosen unfair allocations differ significantly from the punishment rate for delegated allocations ($z = 0.24$, $p = 0.41$ and $z = 1.00$, $p = 0.16$, respectively). Thus, the difference in average punishment levels can be attributed mostly to the difference in conditional punishment levels.

Hence, we can conclude that

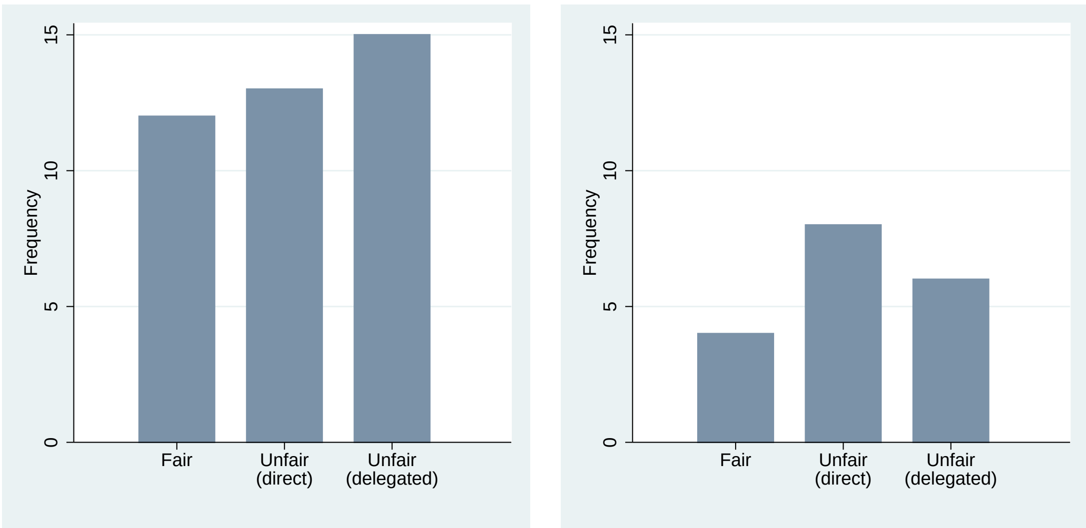
Result 4 *Delegation maximizes expected profits for the dictator.*

In the *Choice* treatment, expected profits for the principal is \$7.08 when delegating, while it is \$6.25 when choosing directly either a or b and \$4.80 when choosing a fair allocation. In the *Random* treatment, expected profits for the principal are \$7.36 when delegating, which is again more than \$6.83 for choosing directly unfair or \$4.81 for choosing fair.

Next we look at the dictator's behavior, which is summarized in Figure 3.3. First, we observe that not all dictators choose the profit-maximizing way to implement a selfish outcome.

Result 5 *A significant number of dictators forgo the intermediary and directly choose an unfair outcome.*

Fifteen (38%) out of 40 dictators in the *Choice* treatment delegate, while 12 (30%) choose fairly and 13 (33%) choose an unfair allocation directly. In the *Random* treatment, 6 out of 18 (33%) delegate, 4 (22%) choose fairly, and 8 (44%) choose an unfair allocation directly. In neither treatment can we reject the hypothesis (two sample test of proportions, $z = 0.47$, $p = 0.64$ and $z = 0.68$, $p = 0.49$, respectively) that dictators directly choose unfairly at the same rate at which they delegate.



(a) *Choice* treatment (b) *Random* treatment

Figure 3.3: The Dictators' Choices

While social preferences such as inequity aversion or simple altruism can explain why some dictators might choose the fair allocation, our results raise the question of why any dictator would choose an unfair allocation directly, instead of delegating the choice to the intermediary. To answer this question, we turn to the beliefs expressed by the dictators. First, we note the accuracy of the aggregate beliefs: mostly, the dictators correctly predict the qualitative features of the punishment schedule, on average.

Result 6 *The average beliefs of the dictators about the deduction behavior of the receivers are in line with the observed pattern of deductions.*

Table 3.3: Expected Punishment

		Choice	*Random*
		Expected Punishment	
Dictators			
For Dictator	if choosing if unfair direct	3.04	1.28
	if choosing unfair delegated	1.42	1.06
For Intermediator	if choosing unfair direct	1.42	2.17
	if choosing unfair delegated	2.25	1.44
Intermediators			
For Dictator	if choosing if unfair direct	3.04	3.70
	if choosing unfair delegated	1.33	1.58
For Intermediator	if choosing unfair direct	1.38	1.21
	if choosing unfair delegated	1.75	1.41

As can be seen in Table 3.3, in the *Choice* treatment, on average, dictators expect to be punished significantly *more* for choosing either of the two unfair allocations directly ($3.04) than for delegating ($1.42, $t = 2.83$, $p < 0.01$, $N = 12$). The intermediaries expect a similar pattern: we can reject the hypothesis ($t = 2.82$, $p < 0.01$) that the intermediaries expect the same average punishment for the dictator regardless of whether she chooses unfair directly ($3.04) or she delegates ($1.33). The beliefs of both the dictator and the intermediary about the intermediary's punishment conform to the observed pattern to a lesser degree: the differences are not statistically significant. The mean belief expressed by dictators about the average amount ($2.25) deducted from N when she delegates exceeds that following her direct choice of an unfair allocation ($1.42; $t = 1.53$, $p = 0.07$). Similarly, the intermediaries believe that the average punishment drops from $1.75 to $1.38 when the dictator chooses an unfair allocation directly instead of delegating ($t = 0.75$, $p = 0.46$).

Also in the *Random* treatment, dictators expect to be on average punished more for choosing either of the two unfair allocations directly ($1.28) than for delegating, even though this difference is not statistically significant (\$1.06, $t = 0.41$, $p = 0.34$, $N = 18$). On the contrary, we can reject the hypothesis ($t = 3.20$, $p < 0.01$) that the intermediaries expect the same average punishment for the dictator regardless of whether she chooses unfair directly (\$3.70) or she delegates (\$1.58). Also the mean belief expressed by dictators about the average amount (\$1.44) deducted from N when she delegates is lower than the ones following her direct choice of an unfair allocation (\$2.17); in line with the true punishment, the difference is again not statistically different ($t = 1.24$, $p = 0.11$). The intermediaries believe that the average punishment for themselves when the dictator chooses an unfair allocation directly (\$1.12) is lower than when she delegates (\$1.41), even though the difference is not statistically significant ($t = 0.65$, $p = 0.52$).

To what extent, though, do dictators' beliefs explain behavior at the individual level? Though the relatively small number of dictators (12 in the *Choice* treatment, 18 in the *Random* treatment) from whom we elicited beliefs does not permit much rigorous statistical analysis, we do not see any evidence that those who choose *a* or *b* directly perceive a smaller punishment reduction for delegating.

Result 7 *The difference in expected deduction between choosing unfairly directly and via the intermediary is not a good predictor of delegating.*

In the *Choice* treatment, eight out of twelve dictators for whom we have beliefs data chose an unfair allocation, with two choosing directly and six delegating. The two who choose directly unfair on average perceive that their punishment would be \$4.50 *lower* had they delegated, while those who delegated perceived the savings from doing so to be only \$1.25. In the *Random* treatment, the dictators who choose directly unfair (8) on average perceive that their punishment would be \$0.10 lower had they delegated, while those who delegated (6) perceived savings from doing so to be \$0.05 compared to choosing unfair directly. Thus, it is hard to argue from this data that those who failed to delegate did not perceive as much of a monetary advantage from doing so as those who did.

3.4 Conclusion

We conducted an experiment in which a dictator could choose either an equal split of a \$20 endowment among members of her four-person group, or one of two unfair allocations that increased the payoff for her and a second group-member, at the expense of the two remaining group members, or she could allow the second player to choose between the two unfair allocations in the *Choice* treatment, while the second player had to press a button to randomly implement one of the two unfair allocations in the *Random treatment* when the dictator delegated. One of the passive recipients was allowed to pay a small cost to deduct up to \$7 from any of the other participants, contingent on the outcome. We find that the recipients use the deductions to punish unfair behavior, but that dictators can effectively shift the blame for unfair behavior onto the intermediary, even though delegating necessarily leaves the intermediary with no choice but to select an unfair allocation.

Why can the dictator avoid punishment by delegating? Outcome-based theories of social-preferences (Charness and Rabin, 2002; Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) cannot explain the shift in punishment from the dictator to the intermediary when the intermediary, as opposed to the dictator, chooses an unfair allocation. Nor can intentions-based models of sequential reciprocity, such as Dufwenberg and Kirchsteiger (2004), which regard the dictator's decision to delegate as no less fair than choosing an unfair allocation directly and which do not regard any of intermediary's choices as unkind, given that she is limited only to unfair options.

Though, unlike Coffman (2011), our experiment was not designed with the explicit purpose of eliminating alternative explanations, such as limited reasoning or the mere presence of a third party, our result is entirely consistent with Coffman's finding that the punishment decreases if the dictator avoids interacting directly with the harmed recipient. However, much remains to be understood about the mechanism behind this effect. Our results show that it persists even when the intentions of the dictator and the absence of responsibility of the intermediary are very transparent, which suggests that delegating 'dirty work' can effectively reduce blame, creating a powerful motive for delegation in a wide variety of settings.

On the other hand, like BF, in the *Choice* treatment, we observed blame actually being shifted *onto* the intermediary, as opposed to merely off of the dictator. This highlights an important limitation on the effectiveness of delegation as a way to avoiding the costs of improper behavior. In some cases, in order to make a contract individually rational for an agent with rational expectations, a principal would have to compensate the agent for much of the punishment she herself would have incurred anyway. Further research with voluntary contracting between the principal and agent, as in Hamman, Loewenstein, and Weber (2010), is needed to understand the conditions under which principals can successfully avoid the costs of misbehavior in the marketplace.

Bartling and Fischbacher propose a formal measure of responsibility to explain punishment patterns. According to their measure, a player takes on responsibility for a bad outcome if and only if her action increases the probability that this outcome will result. While their responsibility measure outperforms outcome- or intentions-based social-preference models in predicting punishment behavior in their experiment, it does not pass the stress test to which we subject it in our experiment. In our setting, a delegating dictator increases the probability that an unfair outcome results, while the intermediary does not, for when the choice comes to her an unfair outcome is already guaranteed. Thus, the BF responsibility measure cannot explain why blame is shifted from the dictator onto the intermediary when the dictator delegates in the *Choice* treatment.

Using the intermediary is profit maximizing and may avoid possible psychological costs associated with having to choose whom to harm most. So why do roughly a third of the dictators choose an unfair outcome directly? Unlike in the experiment of Bartling and Fischbacher, who find a similar result, there is no risk that the intermediary will choose the fair outcome, leaving the dictator with a low payoff.

While it is possible that these dictators care about the wellbeing of the intermediaries and are willing to reduce their own payoff to avoid lowering that of the intermediary, this seems highly unlikely. These dictators can hardly be described as averse to inequality, having forgone the punishment-minimizing equal split. Furthermore, even if these dictators care about the intermediary to the exclusion of the recipients, perhaps including only the intermediary in her reference group because they both benefit from the unfair allocation, the fact that helping the intermediary increases inequality and yields a material payoff for the dictator that is lower than that of the intermediary renders this explanation even less plausible.[8]

Some dictators may have directly chosen an unfair allocation under the belief that the punishment savings from delegating would be trivial, or that the recipient would punish them *more* for delegating because doing so includes and innocent third party and highlights the dictator's "dubious motive" (Paharia, Kassam, Greene, and Bazerman, 2009).[9] However, we find no link between propensity to delegate and the perceived difference in punishment between direct versus delegated unfair allocations. A more plausible explanation is that delegating one's harmful action to someone else might "feel wrong" or violate a personal rule, imposing a psychological cost.

Individuals are known to exploit or even create situations in which observers cannot clearly link actions with their consequences, in order to reduce the social consequences of selfish or harmful behavior. We have shown that, through the use of an intermediary, a dictator may escape social sanctions even when her behavior is *transparently* harmful. Furthermore, despite the transparency, she may effectively shift the blame onto the intermediary. It remains to be seen just how transparently limited the role of the intermediary may be yet still permit the dictator to shift the blame.

These findings have important implications, not just for interpersonal or public relations, but also in organizational settings. By delegating an unpopular decision, a manager may limit the impact on employee morale and effort provision, deflecting negative feelings towards a "fall guy" or an outside company. However, we

[8]Charness and Rabin (2002) find that people have little willingness to sacrifice their own payoff to help those who have more than them.

[9]Paharia, Kassam, Greene, and Bazerman (2009) find that when asked to rate unethical behavior in hypothetical scenarios, subjects judge acts carried out through an intermediary more leniently, though only when shown them one-by-one.

still know little about the functioning of ‘directness’ and about the mechanics of blame-shifting and the delegator-intermediary relationship. Clearly, the punishment technology and the extent to which the intermediary is even subject to sanctions are important. Our finding of blame-shifting, and that of BF, were obtained when punishment was costly and constrained, while Coffman uses a costless and unconstrained technology, which may account for the fact that he did not observe high levels of intermediary punishment.

To what extent can one continue to shift the blame onto the intermediary over the long run? Surely an agent that is punished regularly in the place of the party for whom she serves as the intermediary would require compensation. A savvy principal might select an agent that cannot be easily or effectively punished, such as when a manager hires an outside consultant to implement an unpopular decision. Those passing judgment might eventually lay blame where it is deserved, which may make it difficult to rely on a blame-shifting strategy too heavily and persistently.

References

Aghion, P., and J. Tirole (1997): "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105(1), 1–29.

Bartling, B., and U. Fischbacher (2011): "Shifting the Blame: On Delegation and Responsibility," forthcoming in: Review of Economic Studies 32, Thurgauer Wirtschaftsinstitut, Universitaet Konstanz.

Bolton, G. E., and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90(1), 166–193.

Bolton, P., and M. Dewatripont (2005): *Contract Theory*, vol. 1 of *MIT Press Books*. The MIT Press.

Charness, G., and M. Rabin (2002): "Understanding Social Preferences With Simple Tests," *The Quarterly Journal of Economics*, 117(3), 817–869.

Coffman, L. C. (2011): "Intermediation Reduces Punishment (and Reward)," *American Economic Journal: Microeconomics*, 3(4), 77–106.

Dana, J., D. M. Cain, and R. M. Dawes (2006): "What you don't know won't hurt me: Costly (but quiet) exit in dictator games," *Organizational Behavior and Human Decision Processes*, 100(2), 193–201.

Dana, J., R. Weber, and J. Kuang (2007): "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 33(1), 67–80.

Dufwenberg, M., and G. Kirchsteiger (2004): "A theory of sequential reciprocity," *Games and Economic Behavior*, 47(2), 268–298.

Fehr, E., and K. M. Schmidt (1999): "A Theory Of Fairness, Competition, And Cooperation," *The Quarterly Journal of Economics*, 114(3), 817–868.

Fershtman, C., and U. Gneezy (2001): "Strategic Delegation: An Experiment," *The RAND Journal of Economics*, 32(2), 352–368.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Greiner, B. (2003): "An Online Recruitment System for Economic Experiments," *Forschung und wissenschaftliches Rechnen*, 63, 79–93.

Grossman, Z. (2010): "Strategic Ignorance and the Robustness of Social Preferences," University of California at Santa Barbara, Economics Working Paper Series 1469100, Department of Economics, UC Santa Barbara.

Hamman, J., G. F. Loewenstein, and R. A. Weber (2010): "Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship," *American Economic Review*, (100 (September 2010): 1826-1846).

Lazear, E., U. Malmendier, and R. Weber (2009): "Sorting and Social Preferences," NBER Working Papers 12041, National Bureau of Economic Research, Inc.

O'Rourke, D. (1997): "Smoke from a Hired Gun: A Critique of Nike's Labor and Environmental Auditing in Vietnam as Performed by Ernst & Young," *report published by the Transnational Resource and Action Center: San Francisco.*

Paharia, N., K. S. Kassam, J. D. Greene, and M. H. Bazerman (2009): "Dirty work, clean hands: The moral psychology of indirect agency," *Organizational Behavior and Human Decision Processes*, 109(2), 134–141.

Schelling, T. (1960): *The Strategy of Conflict.* Harvard University Press, Cambridge, MA.

Appendix “Shifting the Blame when Delegating to a Powerless Intermediary”

3.A Instructions for the *Choice* Treatment

The following instructions were given both in written and oral form, while in appendix 3.A.1 - 3.A.3 the instructions are presented which were read by each participant. The examples and exercises participants had to solve are presented in appendix 3.A.3. These reflect the Choice *treatment, with the modifications for the* Random *treatment summarized at the bottom.*

(Oral) Welcome and thank you for participating in this decision-making experiment. You will be paid privately in cash at the end of the experiment, which will last around half an hour. Research foundations have provided the funds. You will make a few decisions that will affect your payoff and possibly the payoffs of other participants. Other participants will simultaneously be making choices that may affect your payoff. Please pay careful attention to the instructions as a considerable amount of money is at stake. You are guaranteed a minimum payment of $5, and may earn as much as $15.

Your participation in this session and any information about your earnings will be kept strictly confidential. Your payment-receipt and consent form are the only places in which your name or perm number are recorded. You will never be asked to reveal your identity to anyone during the course of the experiment. In order to keep decisions private, please do not reveal your choices to any other participant. You will complete one task. Your earnings will be calculated based on your decision and the decision of other subjects, and at the end of the experiment you will be paid that amount plus a $6 show-up fee. If you have any questions during the experiment, please raise your hand and wait for assistance. Please note that for each screen, once you click OK you cannot go back to the previous screen. Please make sure you have read and understand everything completely before you move on.

In this experiment, you will be anonymously grouped together with three other people, so that your decision may affect the payoffs of these three, just as the decisions of the other people in your group may affect your payoffs. You will not know the identity of the other people, and the other people will not know your identity. Each group will consist of four kinds of players, one participant A, one participant B, one participant C and one participant D. Decisions will be made sequentially, in alphabetical order. Participant A starts and can decide how to divide 20 dollars between the four participants. Participant A can choose between four options:

- Distribution 1: Participant A and participant B each receive 9 dollars, participant C receives 2 dollars and participant D receives 0 dollars (this is: 9,9,2,0).
- Distribution 2: Participant A and participant B each receive 9 dollars, participant C receives 0 dollars and participant D receives 2 dollars (this is: 9,9,0,2).
- Distribution 3: Each participant receives 5 dollars (this is: 5,5,5,5).

Participant A can choose between these three distributions, or can pass the decision to participant B. Then, participant B can decide among the first two distributions:

- Distribution 1: Participant A and participant B each receive 9 dollars, participant C receives 2 dollars and participant D receives 0 dollars (this is: 9,9,2,0).
- Distribution 2: Participant A and participant B each receive 9 dollars, participant C receives 0 dollars and participant D receives 2 dollars (this is: 9,9,0,2).

If A delegates the decision to participant B, A cannot take any further decision. Once A, or, in case A delegates, B has made a choice about the allocation of the 20 dollars, participant C and participant D are informed about

- if A has delegated the decision or not
- what distribution was chosen.

Then, either participant C or participant D is randomly chosen. This participant has the possibility to deduct dollars from A, B and the other participant (either D or C), at the cost of one dollar. The randomly chosen player can deduct a maximum of 7 dollars, but can also deduct less. The player deducting points cannot deduct more point from a participant than that participant has earned through the chosen allocation.

Next you will be assigned a role. You will get further instructions on paper, which explain how you will make the decisions for your specific role. There are some examples, and seven short exercises designed to verify your understanding of the instructions. After you completed the exercises, please raise your hand. Once everybody has completed the exercises, we will go over them together.

3.A.1 Instructions for Player *D*

You are participant A. Either you or participant B will decide how to divide 20 points between the four participants in your group. Being participant A, you can choose between four options. If you do not pass the decision to participant B, then participant B will not take a decision. You take the decision. If you delegate the decision to participant B, then you cannot take any further decision. Participant B will take the final decision. In the following table we show you again an overview over all distributions between which you (or, in case you delegate, participant B) can choose.

		Your dollars	Dollars of B	Dollars of C	Dollars of D
You can choose	Distribution 1	9	9	0	2
	Distribution 2	9	9	2	0
	Distribution 3	5	5	5	5
	pass to B				
B can choose	Distribution 1	9	9	0	2
	Distribution 2	9	9	2	0

If you have chosen an allocation of the 20 dollars - or, in case you delegated, participant B has chosen an allocation of the 20 dollars, participant C and participant D are informed about

- whether you have delegated the decision or not, and
- what distribution was chosen.

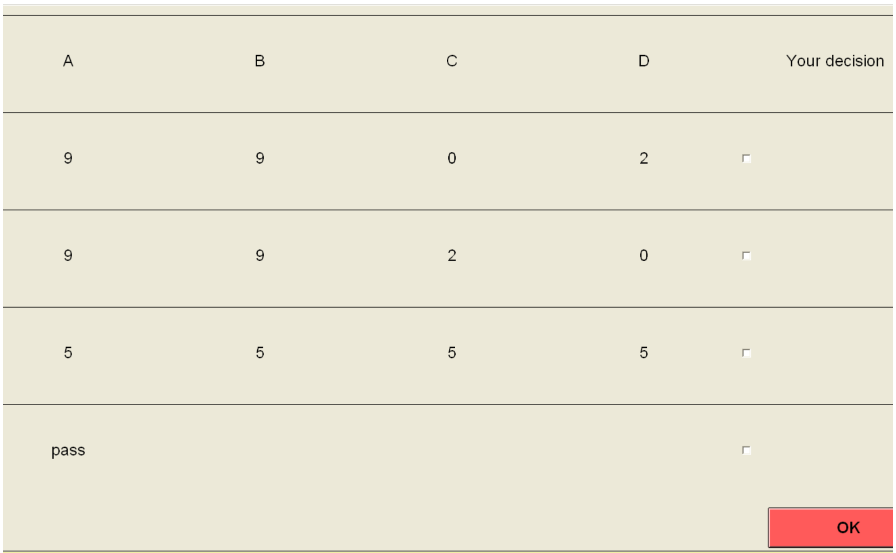
Then, either participant C or participant D is randomly chosen. This participant has the possibility to deduct dollars from you, participant B and the other participant (either D or C), at the cost of one dollar. The randomly chosen player can deduct a maximum of 7 dollars, but can also deduct less. The player can never deduct more dollars than the dollars you earned through the chosen allocation.

What will happen on the computer

Your insert your decision on a screen as the following:

If you want to choose distribution 1, then you click the top small square on the right side. If you want to choose distribution 2, you click the second square. If you want to choose distribution 3, you click the third square. If you want to delegate the decision to participant B, you click in the last square.

After choosing one distribution, you click on the OK-button on the bottom right. As long as you don't click on this button, you can rethink your choice, and select something else. After you (and / or participant B) and the randomly chosen player C or D have made the decision, the experiment is finished and you get your final payoff paid in cash. To summarize, you only take one decision, which you have to insert in the above screen. Think carefully about your decision. Do you have questions?



A	B	C	D	Your decision
9	9	0	2	☐
9	9	2	0	☐
5	5	5	5	☐
pass				☐

OK

3.A.2 Instructions for Player N

You are participant B. Either participant A or you will decide how to divide 20 dollars between the four participants in your group. Participant A can choose between three distributions, or can pass the decision to you, participant B. If A passes the decision, you can only decide among two distributions. If A does not pass the decision, then you will not take a decision. If A delegates the decision to you, you will take the final decision. In the following table we show you again an overview over all distributions between which A (or you, in case A delegates) can choose.

- INPUT SAME TABLE AS FOR PLAYER A -

If A has chosen an allocation of the 20 dollars - or, in case A delegated, you have chosen an allocation of the 20 dollars, participant C and participant D are informed about

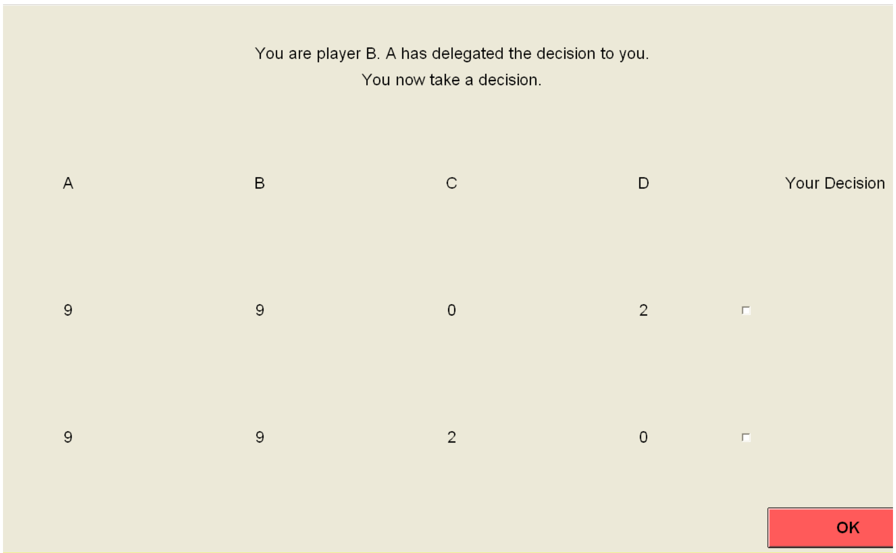
- whether A has delegated the decision or not, and
- what distribution was chosen.

Then, either participant C or participant D is randomly chosen. This participant has the possibility to deduct dollars from you, participant A and the other participant (either D or C), at the cost of one dollar. The randomly chosen player can deduct a maximum of 7 dollars, but can also deduct less. The player can never deduct more dollars than the dollars you earned through the chosen allocation.

What will happen on the computer

If participant A delegates the decision between distribution 1 and 2 to you, then you will see the following screen:

If you want to choose distribution 1, then you click the top small square on the right side. If you want to choose distribution 2, you click the second square.



After choosing one distribution, you click on the OK-button on the bottom right. As long as you don't click on this button, you can rethink your choice, and select something else. After you (and / or participant A) and the randomly chosen player C or D have made the decision, the experiment is finished and you get the final payoff paid in cash. To summarize, you only take one decision, which you have to insert in the above screen. Think carefully about your decision. Do you have questions?

3.A.3 Instructions for Player R_1 / R_2

You are participant C. Either participant A or participant B will decide how to divide 20 dollars between the four participants. Participant A can choose between three distributions, or can pass the decision to participant B, who then can only decide among two distributions. If A does not pass the decision to participant B, then participant B will not take a decision. A takes the decision. If A delegates the decision to participant B, then participant B will take the final decision. In the following table we show you again an overview over all distributions between which A (or, in case A delegates, participant B) can choose.

- INPUT SAME TABLE AS FOR PLAYER A -

If A has chosen an allocation of the 20 dollars - or, in case A delegated, participant B has chosen an allocation of the 20 dollars, you or participant D are chosen randomly. The participant who is chosen has the possibility to deduct dollars from A, participant B and the other participant (either D or C), at the cost of one dollar. The randomly chosen player can deduct a maximum of 7 dollars, but can also deduct less. The player can never deduct more dollars than the dollars a participant earned through the chosen allocation.

Your Decision

Before you get to known which decision participant A and / or participant B has chosen and before you get to known if you or participant D were chosen to deduct dollars, we ask you to make a decision for each of the following five cases:

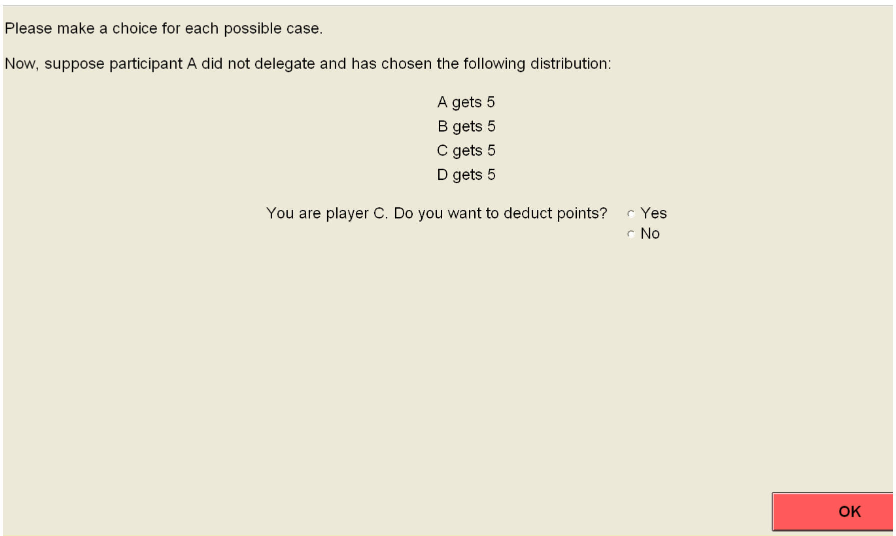
- Participant A does not delegate the decision and chooses allocation 1 (9,9,0,2)
- Participant A does not delegate the decision and chooses allocation 2 (9,9,2,0)
- Participant A does not delegate the decision and chooses allocation 3 (5,5,5,5)
- Participant A does delegate the decision and participant B chooses allocation 1 (9,9,0,2)
- Participant A does delegate the decision and participant B chooses allocation 2 (9,9,2,0)

In particular, this means that for each of the cases you have to say if you want to deduct dollars, and, if yes, how you want to distribute the deducted dollars on the other players. Participant A and / or participant B take the decision without knowing what you or participant D will do in each of the five cases.

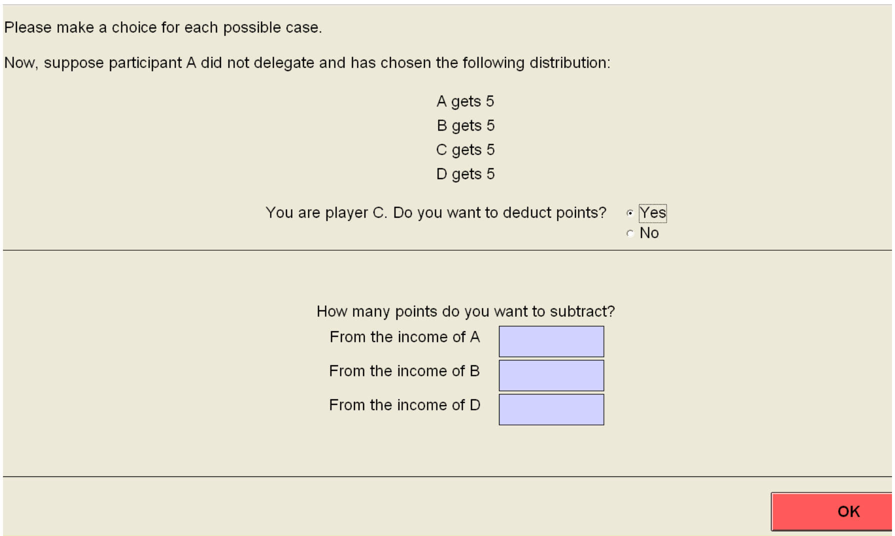
If you are chosen in the random selection process, then your decision is implemented for the case which results out of the decisions of participant A and / or participant B. Therefore, each of your five decisions can be determining for the final payments.

What will happen on the computer

Your decision in each of the cases you insert in five screens like the following:



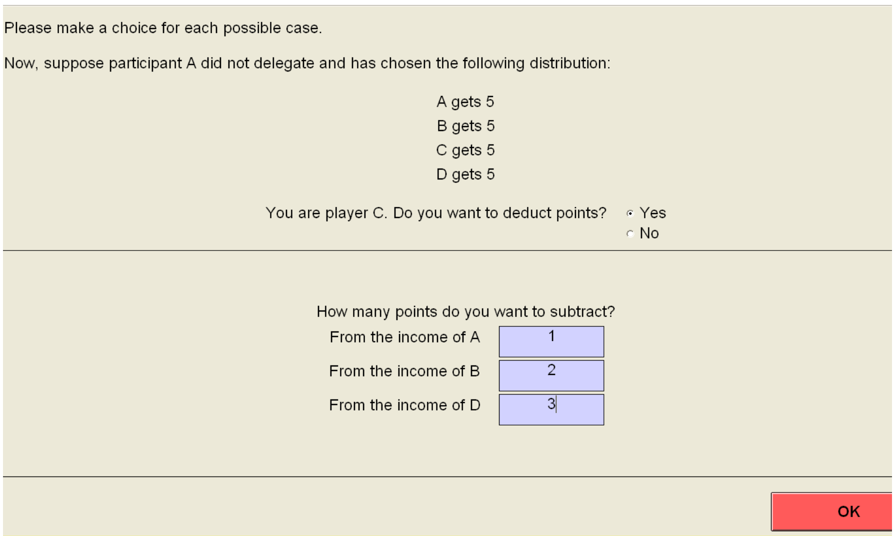
The above example-screen shows you the possibility “Participant A does not delegate the decision and chooses the allocation (5,5,5,5)”. The screens for the other four cases look similar - please pay attention for which case you take the



decision! If you click on “yes”, then the following screen appears (see next page):

After clicking on “yes”, you can insert the respective amount you want to deduct in the three boxes . If you choose to deduct a dollar of at least one other player, you will be deducted one dollar - and the player of whom you want to deduct dollars loses the amount you stated in the respective field. If you choose “no” when asked if you want to deduct dollars, then the three small boxes do not appear (or disappear again), and you cannot deduct dollars.

After choosing one distribution, you click on the OK-button on the bottom right. As long as you don’t click on this button, you can rethink your choice, and select something else.



In this example participant C wants to deduct dollars. Therefore, C clicked “yes”, and the three small boxes appeared (see the picture on the previous page). Participant C deducts 1 dollar of participant A, 2 dollars of participant B, and

3 dollars of participant D. (This is just an example and not a suggestion of how you should act.) In total you can deduct up to 7 dollars in each case. You can as well (as in the above example) deduct less than seven dollars. Thereby, you cannot deduct more dollars of a participant than what that participant received according to the respective allocation. In the above example, therefore, you cannot deduct of any player more than 5 dollars.

When you press the OK-button, you come to the next case. As long as this button is not pressed, you can still change all your entries. Do you have any questions?

Examples

The examples and the exercises are the ones that were presented to player D. For I, R_1 and R_2, they were modified respectively.

- Example 1: Distribution 2 is chosen (either by yourself or by participant B) and the randomly chosen participant is participant C. C choses to give up one dollar, to deduct 3 dollars from you and 4 dollars from participant B. Then, the following distribution results:

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 2	9-3 = 6	9-4 = 5	2 - 1 = 1	0

- Example 2: Distribution 3 is chosen (by yourself) and the randomly chosen participant is participant D. D choses to give up one dollar, to deduct two dollars from you, three dollars from participant B and one dollar of participant C. D choses to not to use the seventh possible dollar to deduct. Then, the following distribution results:

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 3	5-2 = 3	5-3 = 2	5-1 = 4	5-1 = 4

- Example 3: Distribution 2 is chosen (either by yourself or by participant B) and the randomly chosen participant is participant D. D choses to not to give up one dollar, to deduct dollars from other players. Therefore, the resulting distribution is as explained above, (9,9,2,0).

- Example 4: Distribution 1 is chosen (by yourself or by participant B) and the randomly chosen participant is participant C. C choses to give up one dollar, to deduct two dollars from you, two dollars from participant B and two dollars of participant D. C choses to not to use the seventh possible dollar to deduct. Then, the following distribution results:

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 1	9-2 = 7	9-2 = 7	0-1 = -1	2-2 = 0

Exercises Each Participant Had to Solve

Please answer the following questions. They only serve for helping you to get used to the experiment. The decisions and payments in the exercise here are chosen arbitrarily. Do not take them as suggestions for which allocation you should chose. Your answers here will not have any impact on the payments at the end of the experiment.

- Participant A has passed the decision to participant B. Whose decisions are relevant for the payments at the end of the experiment?
- Participant A did not pass the decision to participant B. Whose decisions are relevant for the payments at the end of the experiment?
- Distribution 2 is chosen. Participant C is randomly selected to deduct dollars, and wants to deduct the following bold printed amounts:

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 2	9	9	2	0
Deduction	**2**	**2**		**0**

Is this possible? If yes, please determine the resulting payment. If not, please make a mark where it is not possible.

- Distribution 2 is chosen. Participant D is randomly selected to deduct dollars, and wants to deduct the following bold printed amounts:

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 2	9	9	2	0
Deduction	**2**	**1**	**3**	

Is this possible? If yes, please determine the resulting payment. If not, please make a mark where it is not possible.

- Distribution 3 is chosen. Participant C is randomly selected to deduct dollars, and wants to deduct the following bold printed amounts:

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 3	5	5	5	5
Deduction	**2**	**3**		**3**

Is this possible? If yes, please determine the resulting payment. If not, please make a mark where it is not possible.

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 3	5	5	5	5
Deduction	**7**	**0**	**0**	

- Distribution 3 is chosen. Participant D is randomly selected to deduct dollars, and wants to deduct the following bold printed amounts:

  Is this possible? If yes, please determine the resulting payment. If not, please make a mark where it is not possible.

- Distribution 1 is chosen. Participant D is randomly selected to deduct dollars, and wants to deduct the following bold printed amounts: Is this pos-

	Your dollars	Dollars of B	Dollars of C	Dollars of D
Distribution 1	9	9	0	2
Deduction	**0**	**0**	**0**	

  sible? If yes, please determine the resulting payment. If not, please make a mark where it is not possible.

Please remember that these are only exercises, and that all numbers were chosen arbitrarily. You should not use the numbers to orient your decision on them. When you have solved the exercises, please raise your hand. You may think about your decision in the experiment.

Belief Elicitation

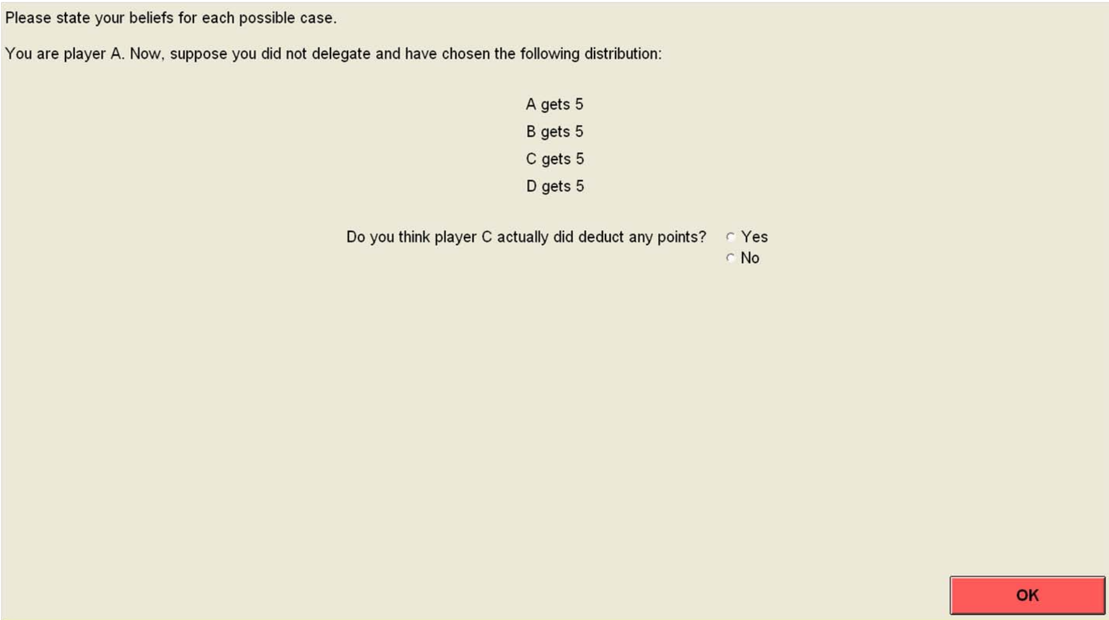
The screens used to elicit belief from the As were the following (the further screens were just adapted from those):

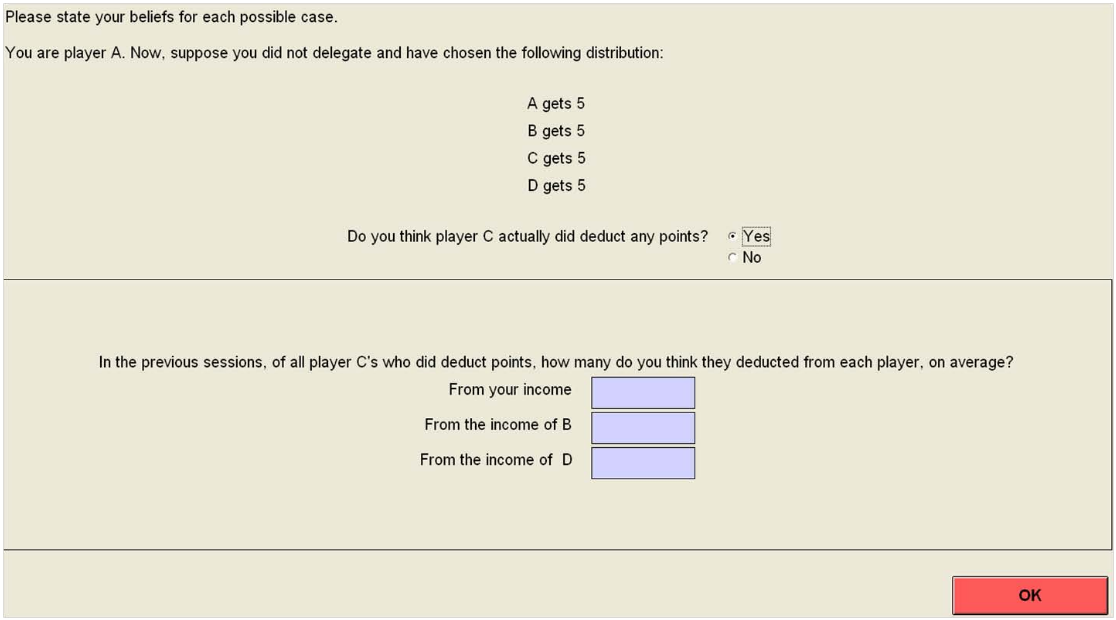
> Now that you have made your decision, we would like to give you a chance to earn some extra dollars.
>
> We will ask you what you belief the choice of participant C will be and has been in previous sessions. If your answer is within +/- $1 of the true average punishment of the C's in previous sessions, you will earn an additional $5.
>
> The average punishment is calculated using only the people that punished at least one person in the respective case. Hence, if you think player C did punish, we want you to state your beliefs about previous levels of punishment, conditional on punishing at all.
>
> If you think that in a case C did not punish, this input counts as if you state C deducted 0 points.
>
> If you have any questions, please raise your hand and wait for assistance.

3.B Instructions for the *Random* treatment

This treatment included the following modifications:

1. After displaying the available options for the dictator, we changed the text for all participants to the following:

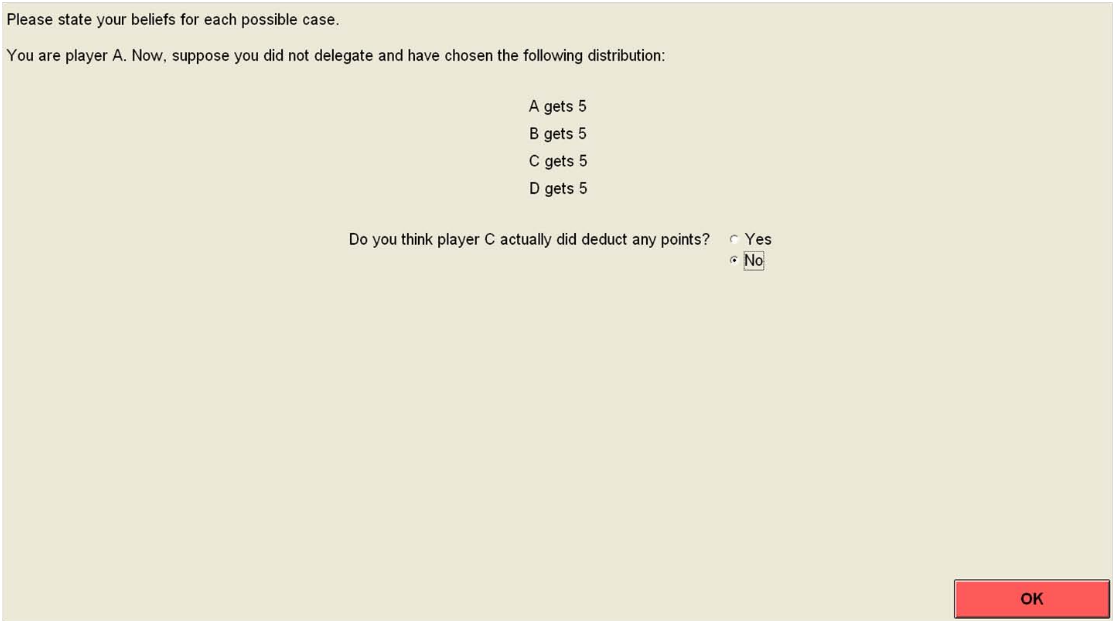




Participant A can choose between these three distributions, or can pass the decision to participant B. If A passes, participant B must click a button. When B clicks the button, the computer randomly selects either distribution 1 or 2. While each of these two distributions is equally likely, distribution 3 is never selected. If A passes the decision to participant B, A does not make any further decisions. Once A, or, in case A passes, the computer has selected how to distribute the 20 dollars, participant C and participant D are informed about

- if A has delegated the decision or not
- which distribution was chosen.

Then, either participant C or participant D is randomly selected to have the



opportunity to deduct dollars from A, B and the other participant (either D or C). Whichever person is chosen can deduct up to 7 dollars in any combination from the other participants. If that person does not make any deductions from any other participants, his or her earnings will be unaffected. However, it will cost her $1 if she does deduct anything from anyone. Note that even if this person obtained $0 from the chosen distribution, even if she pays the $1 to deduct money from others, she will still have positive earnings from the session because of the $6 show-up fee. The player deducting dollars cannot deduct more dollars from a participant than that participant has earned through the chosen allocation.

2. Similarly, the first part of the instructions of player A was changed to the following:

   You are participant A. Either you or participant B will decide how to divide 20 points between the four participants in your group. Being participant A, you can choose between four options. If you do not pass the decision to participant B, then participant B will not take a decision. You take the decision. If you delegate the decision to participant B, then you cannot take any further decision. By clicking on a button, Participant B will take the final decision: the computer will implement the final choice between distribution 1 and distribution 2. In the following table we show you again an overview over all distributions between which you (or, in case you delegate, participant B through his random choice) can choose.

3. The tables shown to the players were changed to the following:

4. The first part of the instructions of player B was changed to the following:

   You are participant B. Either participant A or your click will decide how to divide 20 dollars between the four participants in your group. Participant

		Your dollars	Dollars of B	Dollars of C	Dollars of D
You can choose	Distribution 1	9	9	0	2
	Distribution 2	9	9	2	0
	Distribution 3	5	5	5	5
	pass to B				
B clicks on a button, and the computer decides which of the two distributions to implement:	Distribution 1	9	9	0	2
	Distribution 2	9	9	2	0

A can choose between three distributions, or can pass the decision to you, participant B. If A passes the decision, you have to click on a button, and through this click it is randomly determined which of two distributions is implemented. If A does not pass the decision, then you will not take a decision. If A delegates the decision to you, your click will implement the final allocation. In the following table we show you again an overview over all distributions between which A (or you, in case A delegates) can choose.

5. The first part of the instructions of player C was changed to the following:

   You are participant C. Either participant A or your click will decide how to divide 20 dollars between the four participants in your group. Participant A can choose between three distributions, or can pass the decision to participant B. If A passes the decision, B has to click on a button, and through this click it is randomly determined which of two distributions is implemented. If A does not pass the decision, then B will not take a decision. If A delegates the decision to B, B's click will implement the final allocation. If A does not pass the decision to participant B, then participant B will not take a decision. A takes the decision.

Further small changes were made to adapt all instructions to the new setting. Full instructions are available from the authors.