THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Roberto ZEITOUNLIAN SARKISIAN

Le 9 juillet 2020

Essais dans les situations de contract avec préférences prosociales

---

Ecole doctorale : **Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
UMR 5314 -TSE-R

Thèse dirigée par
Ingela ALGER et François SALANIE

Jury

M. Guido Friebel, Rapporteur
M. Salvatore Di Falco, Rapporteur
Mme Ingela ALGER, Directrice de thèse
M. François Salanié, Co-directeur de thèse

Essays on Contracting under Other-Regarding Preferences

Roberto Sarkisian[*]

May 5, 2020

1 Introduction

Neoclassical theory assumes that economic agents are atomistic, in the sense that each individual is too small relative to the market to affect prices and each others' decisions. In such a setting, it is only natural to imagine that economic agents are purely selfish, only seeking to maximize own profit or utility. However, many interesting economic settings are characterized by the interaction among a small number of agents, in which they recognize that the combination of their strategic choices ultimately determine all participants' outcomes.

While many advances were made under the assumption that individuals' preferences were like in the neoclassical theory, experimental evidence pointed to deviations from the behaviors predicted by such game-theoretical models.[1] Rotemberg (2006) surveys both theoretical and empirical evidence for the presence of altruism in the workplace, be it displayed among employees towards one another, towards the employer, and also from the employer towards the subordinates. Capraro and Rand (2018) and Bilancini et al. (2020) find compelling

[*]Toulouse School of Economics. *E-mail:* roberto.sarkisian@tse-fr.eu

[1]The interested reader can refer to Kolm and Ythier (2006a,b) for an extensive review on the experimental and behavioral literatures of other-regarding preferences, in particular altruism, gift-giving and reciprocity.

experimental evidence that individuals display moral preferences, that ultimately induces them to do what they think is morally right.

On the theoretical side, many classes of preferences were proposed to explain individual behavior, such as reciprocity (Rabin, 1993; Falk and Fischbacher, 2006), inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), warm-glow (Andreoni, 1990), and altruism Becker (1974, 1976). More recently, Alger and Weibull (2013, 2016, 2017) has shown that a novel class of preferences, called *homo moralis*, is the one favored by long-term evolution.

Along with the efforts mentioned above, Andreoni and Miller (2002), Bellemare et al. (2008), Bellemare et al. (2011), Fisman et al. (2007), Bruhin et al. (2019), and Alger et al. (2019) among others try to empirically identify the underlying preference classes of individuals agents across different settings. Among their results, Bruhin et al. (2019) show that pure selfishness does not emerge as a temporally stable preference type, while altruism and behindness averseness do. Alger et al. (2019) expand on Bruhin et al. (2019)'s analysis to allow for strategic interactions and find that Kantian morality as part of the motivation behind individuals' choices.

In the following essays, I will study the effects of two classes of other-regarding preferences in contracting situations, namely altruism and *homo moralis*. My focus in this two classes is due to the fact that they are behaviorally indistinguishable in certain environments (Alger and Weibull, 2013; Bergström, 1995), which naturally leads to the questions of whether one preference class can be preferred over the other by a third party tasked with contracting such agents and if such prosociality can foster or hinder the development of certain markets.

The first three essays look closely at the problem faced by a principal who must design compensation schemes for *homo moralis* or altruistic agents, and how different types and degrees of other-regarding preferences will affect the optimal incentive scheme, the relative importance of variable income to fixed income in these contracts and, ultimately, which class of other-regarding preferences is the most preferred by the principal depending on the production technology and performance measures available to the principal. The fourth and last essay studies how other-regarding preferences may affect the emergence of markets. More precisely, I study how altruism can deter the emergence of formal insurance markets by sustaining cross-insurance transfers between individuals in detriment to the purchase of private insurance policies.

1.1 Contracting with Moral and Altruistic Agents

The first essay of this thesis focuses on the productive relationship between a principal and a team of agents. In this moral-hazard-in-teams setting, individuals can exert either a high level of effort or no effort, and thus stochastically affect the binary outcome of the principal. Because effort is costly, the principal must design a compensation scheme that induces the agents to exert the high level of effort, which is assumed to be preferred by the principal over no effort.

I assume that each team is composed by two symmetric agents, that can be characterized by two different types of prosocial preferences: altruism or *homo moralis*.[2] Moreover, the agents' preference type is common knowledge, and so is the common degree of prosociality. In this setting, the principal must then design the compensation schemes for each preference type, and then determine which type of prosocial preferences lead to higher expected profits.

This article builds upon the literature on moral hazard in teams (Holmström, 1982; Mookherjee, 1984; Itoh, 1991; Che and Yoo, 2001) and is related to the question of incentive provision for agents with prosocial preferences, such as the works of Itoh (2004), Rey-Biel (2008), Englmaier and Wambach (2010), and Livio (2015) among others. While the aforementioned studies analyze the optimal incentive scheme for a given preference class, my contribution is the comparison on the incentive structure proposed to the two preference classes, namely altruism and *homo moralis*.

It should come as no surprise that highly moral or altruistic agents are preferred by the

[2]Standard selfish preferences can be seen as a particular case of either type of prosocial preference, and the standard results on the literature are recovered by the appropriate choice of the degree of prosociality, namely zero.

principal to less prosocial individuals. Indeed, as prosociality increases, the cost of an agent shirking also increases due to the internalization of the reduced expected wages for their partner. Therefore, the principal can offer less high-powered contracts to the agents while incentivising both participation and high effort. However, for the same degree of prosociality, choosing a team of altruistic agents over a team of *homo moralis* agents or vice-versa also depends on the productive technology of the relationship between the two parties, a novelty in the literature.

The reason behind this important role of the production technology is how different prosocial types compute the gains of a deviation. On one hand, altruistic individuals are going to weight their own gain of lowering effort to the expected loss in wages of their partner while the latter holds his effort constant. On the other hand, a *homo moralis* agent would weight the gain of reduced effort to the expected loss in own wage if his partner were to shirk as well.

I thus show that *homo moralis* agents will only be preferred by the principal over altruistic agents if: (*i*) the production technology exhibits decreasing returns to efforts; (*ii*) the probability of a high realization of output conditional on both agents exerting effort is sufficiently high; and (*iii*) the outside option for the agents is zero or the degree of prosociality is sufficiently low.

1.2 Screening Moral and Altruistic Agents

While the model in the first essay shows how a principal can profit by choosing a team with certain type of prosocial preferences depending on the production technology and the degree of prosociality, the whole analysis is based on the strong assumption that the principal knows exactly whether he is facing a team of altruistic or *homo moralis* agents. The second essay relaxes this assumption, and asks whether the principal is capable of designing a menu of contracts that not only induces the agents to participate and exert the high level of effort, but also to induce them to reveal their preference type.

In this setting of screening followed by moral-hazard (Laffont and Martimort, 2002; Jullien et al., 2007; Ollier and Thomas, 2013; Maréchal and Thomas, 2018), I extend the analysis of Sarkisian (2017) to consider the relationship between heterogenous social preferences and the possibility of screening via a menu of contracts, much in the same spirit as von Siemens (2011). In contrast to von Siemens (2011), which screens different degrees of other-regarding preferences (namely inequity aversion), the goal in this chapter is to screen the two different classes of other-regarding preferences present in Sarkisian (2017).

Looking at space of contracts that satisfy the participation and incentive compatibility constraint for the moral-hazard-in-teams problem, one realizes that there exists equilibria in which the different preference types truthfully reveal their private information. However, all these equilibria are sustained by constructing a menu containing contracts at the intersection of the sets satisfying the aforementioned constrains for each type. This implies that, for any pair of contracts, two diametrically opposed equilibria exist: one in which each group accepts the contract designed for them, and another in which they take each others' offers.

I then ask whether the principal can design a more stringent menu, one satisfying the participation and incentive compatibility constraint for the moral-hazard-in-teams problem plus an incentive compatibility for the screening problem. The answer then is negative: separating equilibria will exist only if one of the moral-hazard-in-teams constraint is relaxed. If the participation constraint is the one abandoned, the principal offers a menu of contracts that only attracts one type of prosocial preferences, namely the cheapest to hire according to the results of the first essay. On the other hand, if the effort incentive compatibility constraint is relaxed, the principal can offer a menu of contracts inducing different levels of efforts for each preference class, and this additional dimension allows the principal to successfully screen *homo moralis* and altruistic agents.

Last, but not least, I show that pooling equilibria also exists for both high or no effort, by offering contracts satisfying the most stringent constraints of the moral-hazard-in-teams problem.

1.3 Optimal Contracting with Moral Agents

Up to this point, I have considered environments in which the principal can only observe a single performance measure of the agents' combined efforts, and is therefore limited to offering the later contracts conditioned on this single measure.

On the other hand, when each agent has an individual performance measure, the space of contracts than can be offered by the principal is substantially increased, since now she is able to contingent each agent's compensation scheme in all observable measures. In particular, the principal can now design tournaments, which have been shown to be optimal for the principal when the agents have selfish preferences and the efforts independently affect the individual performance measures (Che and Yoo, 2001).

In this third essay, I study what is the optimal incentive scheme proposed by a principal faced with two *homo moralis* agents and individual performance measures. I find that, contrary to the case with purely selfish preferences, tournaments can never be optimal when agents are risk averse.[3] Indeed, in most cases, a relative performance scheme, in which an agent is paid only if his own output is high, is the optimal scheme for incentivising *homo moralis* agents.

The worse performance of tournaments when agents are characterized by *homo moralis* preferences is directly related to how such individuals internalize a lower effort. Because an agent deviating to a lower effort believes his partner would follow him suit, the incentives provided by a tournament for an agent not to lag behind are reduced, and could only be compensated by an inefficiently high (and not profitable) prize offered by the principal.

This chapter contributes to the literature exploring optimal incentive schemes for other-regarding preferences, such as the analysis on inequity aversion (Itoh, 2004; Rey-Biel, 2008; Englmaier and Wambach, 2010), reciprocity (Livio, 2015), and altruism (Dur and Sol, 2010; Dur and Tichem, 2015). In particular, I derive the optimal incentive scheme for a novel class of other-regarding preferences, namely *homo moralis*, and show the suboptimality of

[3] Even with risk-neutral *homo moralis* agents, the principal is at best indifferent between tournaments and other incentive structures.

tournaments as an incentive scheme, in contrast to the existing literature.

1.4 Altruism and Insurance Markets

Which factors contribute to the emergence of formal institutions? Alternatively, which factors prevent it? It is not uncommon to observe interactions between market institutions and nonmarket trades in modern societies. One such example is agents relying not only on insurance policies to protect themselves against occasional losses, but also on family members and friends to help in adverse situations. Recent surveys suggest that in most cases reliance on nonmarket trades are more pronounced in either less developed societies (Cox et al., 2006; Cox and Fafchamps, 2008) or those where kinship among its members is higher (Costa-Font, 2010), while more developed societies correlate with stronger formal institutions, such as large banking and credit markets.

A large body of literature has been developed in the last decades focusing on remittances, credit cooperatives and risk-sharing either in rural areas or in developing countries. For the most part, transfers among agents in these settings rely on the repeated structure of the analysis, being largely sustained by the threats of punishment and exclusion from kin networks in the future.

I consider a static model where agents can engage in three actions: (i) *insurance policies purchase*, to cover their losses in an adverse situation; (ii) *self-protection*, i.e. exerting effort to avoid a loss from happening; and (iii) *cross-insurance*, by transferring resources to one another in order to share the risk each individual face.

The analysis is divided in three parts. The first one abstracts from the agents' effort choice, by taking the probabilities of suffering losses to be exogenous. I then show that a selfish agent can free-ride on an altruistic partner's transfers and not buy insurance. Qualitatively, a similar result hold in the second part of the analysis, where efforts are endogenous, the main difference being that the uninsured agent will generally exert higher effort than his counterpart due to the higher risk the former faces. The final part introduces a principal who can design the contracts being offered to the two agents. I show that although there are gains of trade to be had by all parties if both agents are obliged to buy insurance, self-protection and cross-insurance may crowd-out formal insurance when the decision to buy the firm's policies is a strategic choice of the agents.

Such results suggest that, in the case of remittances in particular, firms seeking to enter in a market where a high proportion of selfish agents live, and who rely on transfers coming from their altruistic partners elsewhere, may not have sufficient demand to operate. On the other hand, if both agents must buy insurance policies, I show that gains of trade can always be obtained among the three parties when both agents display the same degree of altruism.

References

Alger, I., B. Van Leeuwen, and J. W. Weibull (2019, November). Estimating Social Preferences and Kantian Morality in Strategic Interactions. TSE Working Papers 19-1056, Toulouse School of Economics (TSE).

Alger, I. and J. W. Weibull (2013, November). Homo moralis - preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2302.

Alger, I. and J. W. Weibull (2016). Evolution and kantian morality. *Games and Economic Behavior 98*(C), 56–67.

Alger, I. and J. W. Weibull (2017, Sep). Strategic behavior of moralists and altruists. *Games 8*(3), 38.

Andreoni, J. (1990, June). Impure altruism and donations to public goods: A theory of warm-glow giving? *Economic Journal 100*(401), 464–77.

Andreoni, J. and J. Miller (2002, March). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Becker, G. S. (1974, Nov.-Dec.). A theory of social interactions. *Journal of Political Economy 82*(6), 1063–93.

Becker, G. S. (1976, September). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature 14*(3), 817–26.

Bellemare, C., S. Krger, and A. van Soest (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica 76*(4), 815–839.

Bellemare, C., S. Krger, and A. van Soest (2011). Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool. *Journal of Economic Behavior & Organization 78*(3), 349–365.

Bergström, T. C. (1995, March). On the evolution of altruistic ethical rules for siblings. *American Economic Review 85*(1), 58–81.

Bilancini, E., L. Boncinelli, V. Capraro, T. Celadin, and R. D. Paolo (2020). 'Do the Right Thing' for Whom? An Experiment on Ingroup Favouritism, Group Assorting and Moral Suasion. *Judgment and Decision Making*, 11.

Bolton, G. E. and A. Ockenfels (2000, March). Erc: A theory of equity, reciprocity, and competition. *American Economic Review 90*(1), 166–193.

Bruhin, A., E. Fehr, and D. Schunk (2019). The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association 17*(4), 1025–1069.

Capraro, V. and D. G. Rand (2018, January). Do the Right Thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making 13*(1), 99–111.

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

Costa-Font, J. (2010, Winter). Family ties and the crowding out of long-term care insurance. *Oxford Review of Economic Policy 26*(4), 691–712.

Cox, D. and M. Fafchamps (2008, January). *Extended Family and Kinship Networks: Economic Insights and Evolutionary Directions*, Volume 4 of *Handbook of Development Economics*, Chapter 58, pp. 3711–3784. Elsevier.

Cox, D., E. Galasso, and E. Jiminez (2006, January). Private Transfers in a Cross Section of Developing Countries. Working Papers, Center for Retirement Research at Boston College wp2006-1, Center for Retirement Research.

Dur, R. and J. Sol (2010, July). Social interaction, co-worker altruism, and incentives. *Games and Economic Behavior 69*(2), 293–301.

Dur, R. and J. Tichem (2015, 09). Altruism and relational incentives in the workplace. *Journal of Economics & Management Strategy 24*(3), 485–500.

Englmaier, F. and A. Wambach (2010, July). Optimal incentive contracts under inequity aversion. *Games and Economic Behavior 69*(2), 312–328.

Falk, A. and U. Fischbacher (2006, February). A theory of reciprocity. *Games and Economic Behavior 54*(2), 293–315.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Fisman, R., S. Kariv, and D. Markovits (2007, December). Individual Preferences for Giving. *American Economic Review 97*(5), 1858–1876.

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

Itoh, H. (1991, May). Incentives to help in multi-agent situations. *Econometrica 59*(3), 611–36.

Itoh, H. (2004). Moral hazard and other-regarding preferences. *The Japanese Economic Review 55*(1), 18–45.

Jullien, B., B. Salanié, and F. Salanié (2007, January). Screening risk-averse agents under moral hazard: single-crossing and the CARA case. *Economic Theory 30*(1), 151–169.

Kolm, S. and J. M. Ythier (2006a). *Handbook of the Economics of Giving, Altruism and Reciprocity* (1 ed.), Volume 1. Elsevier.

Kolm, S. and J. M. Ythier (2006b). *Handbook of the Economics of Giving, Altruism and Reciprocity* (1 ed.), Volume 2. Elsevier.

Laffont, J.-J. and D. Martimort (Eds.) (2002). *The Theory of Incentives: The Principal-Agent Model* (1 ed.), Volume 1. Princeton University Press.

Livio, L. (2015). Friend or foes? optimal incentives for reciprocal agents. *Mimeo*, 1–47.

Maréchal, F. and L. Thomas (2018). The optimal contract under adverse selection in a moral-hazard model with a risk-averse agent. *Games 9*(1), 12.

Mookherjee, D. (1984). Optimal incentive schemes with many agents. *Review of Economic Studies 51*(3), 433–446.

Ollier, S. and L. Thomas (2013). Ex post participation constraint in a principalagent model with adverse selection and moral hazard. *Journal of Economic Theory 148*(6), 2383–2403.

Rabin, M. (1993, December). Incorporating fairness into game theory and economics. *American Economic Review 83*(5), 1281–1302.

Rey-Biel, P. (2008, 06). Inequity aversion and team incentives. *Scandinavian Journal of Economics 110*(2), 297–320.

Rotemberg, J. J. (2006). Altruism, reciprocity and cooperation in the workplace. In S. Kolm and J. M. Ythier (Eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Volume 1, Chapter 21, pp. 1371–1407. Elsevier.

Sarkisian, R. (2017). Team incentives under moral and altruistic preferences: Which team to choose? *Games 8*(3), 37.

von Siemens, F. (2011). Heterogeneous social preferences, screening, and employment contracts. *Oxford Economic Papers 63*(3), 499–522.

Team Incentives under Moral and Altruistic Preferences: Which Team to Choose?

Roberto Sarkisian*

May 5, 2020

Abstract

This paper studies incentives provision when agents are characterized either by *homo moralis* preferences (Alger and Weibull, 2013, 2016), i.e. their utility is represented by a convex combination of selfish preferences and Kantian morality, or by altruism. In a moral hazard in teams setting with two agents whose efforts affect output stochastically, I demonstrate that the power of extrinsic incentives decreases with the degrees of morality and altruism displayed by the agents, thus leading to increased profits for the principal. I also show that a team of moral agents will only be preferred if the production technology exhibits decreasing returns to efforts, the probability of a high realization if output conditional on both agents exerting effort is sufficiently high and either the outside option for the agents is zero or the degree of morality is sufficiently low.

Keywords: Moral hazard in teams, optimal contracts, *homo moralis* preferences, altruism.

*Toulouse School of Economics. *E-mail:* roberto.sarkisian@tse-fr.eu

1 Introduction

Teamwork permeates economic activities. In some cases, different skills are needed complementing each other to complete a project; in other cases, division of labor plays a crucial role in timely delivery of a product. Partnerships, group projects and team sports are but a few examples in which individuals team up to achieve a certain goal. With the exception of partnerships, it falls upon an employer to hire the team of employees to fulfill the task at hand. In particular, as expenditure in recruitment and assessment surges[1], with US companies spending on average around $US\$4000$ per hire, an increase of nearly 15% in the last four years, it is clear that recruitment and talent research divisions have turned their attention to more than the job applicants' professional abilities. As a matter of fact, common practice includes the analysis of criminal[2] and credit histories[3], and more recently, social networks as well[4].

While employees' technical skills are important, interest in their personal characteristics other than job-relevant skills may be related to the now widespread knowledge that economic agents are not purely selfish, often displaying other-regarding preferences (Fehr and Schmidt, 2006; Kolm and Ythier, 2006a,b; Kahneman et al., 1986, 1991). The literature in behavioral and experimental economics strongly suggests that social preferences affects outcomes in standard economic models (Dufwenberg et al., 2011; Bowles and Polania-Reyes, 2012). Bandiera et al. (2005), Bandiera et al. (2010), Barr and Serneels (2009) and Rotemberg (2006) analyze, in particular, the role other-regarding preferences play in interactions among employees in models of the workplace. The main findings in this literature show that the employees' concerns towards one another affect not only the provision of effort, but also the compensation schemes that are offered. Thus, it is only natural to wonder what an

[1]See O'Leonard et al. (2015).

[2]See the Society for Human Resource Management survey results at
https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/criminalbackgroundcheck.aspx

[3]See the Society for Human Resource Management survey results at
https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/creditbackgroundchecks.aspx

[4]See, for instance, Brown and Vaughn (2011).

ideal team would look like, for a given set of skills: would it be a team composed of selfish individuals, whose only concern is their own gains? Or maybe a group of altruistic agents, who would be content in increasing their workmates' wellbeing? Perhaps a crew of moral employees deriving satisfaction in choosing actions they think are the right ones? This is the question I address in this paper.

In view of the overwhelming experimental evidence of behaviors that are incompatible with purely selfish preferences (Kolm and Ythier, 2006a,b; Thaler, 1988; Tversky and Thaler, 1990), it is important to understand how prosocial preferences affect behavior in the workplace, and by extension, the design of contracts in the workplace. I propose a model to address this question.

Specifically, I focus on the optimal compensation schemes that should be used in a standard moral hazard setting to incentivise the employees to fulfill their tasks. In doing so, I am able to compare the profits obtained by the employer from a team composed of individuals with different kinds of prosocial preferences. Although I do not study the recruitment process *per se*, I am able to make predictions about which preferences the principal would prefer.

The framework utilized is the multiagent moral hazard model, as first proposed by Alchian and Demsetz (1972) and Holmström (1982), where a risk-neutral principal hires a team of two risk-averse agents. The agents can exert costly effort in order to stochastically affect the realization of output. By assumption, efforts are simultaneously and independently chosen by the agents, and cannot be observed by the principal. On the other hand, output is observable by third parties after being realized, and can thus be contracted upon.

In the behavioral economics literature several classes of prosocial preferences have been proposed. I analyze two of them. The first one is altruism (Becker, 1974), a class which has been extensively used in the literature on the voluntary contribution of public goods. This is natural since one can think of efforts made in the context of teamwork in a firm as contributions to a public good (the firm's profit). Second, in light of recent results by Alger and Weibull (2013, 2016), who show that a particular, novel, class of preferences stands out as being favored by evolution, I compare the optimal contract under altruism with the optimal contract under this class of *homo moralis* preferences, a convex combination of selfishness and morality. In sum, my model allows to address the following questions: if an employer could choose between a team of two moral agents and a team of two altruists, which team would he prefer, and why[5]?

I characterize the optimal contract for a team of equally altruistic agents and a team of equally moral agents, and compare them. First, I find that the trade-off between risk-sharing and incentive provision is present, as in the case with standard selfish preferences. However, as intuition would suggest, I find that high-powered incentives are less needed to induce effort as the agents become more concerned about *the right thing to do* or about each other's material payoff, and that the principal's expected profit obtained from the interaction with each team is increasing in the team's degree of morality or altruism. Second, if efforts are symmetric and could be contracted upon, the principal would be better off hiring a team of altruistic agents over the other ones, for any degrees of morality and altruism, because altruism towards one's partner reduces the payment necessary to induce participation, one effect that is not present with selfish or moral preferences. On the other hand, when efforts are not observable, which team is going to be preferred depends on the production technology: in particular, if the stochastic production technology displays increasing returns to efforts, the altruistic team is the cheapest to hire. This is a consequence of the different nature of each class of preferences. While altruistic agents derive benefits from increased material payoffs of their fellows, moral agents take satisfaction in doing the right thing. Intuitively, a higher effort under increasing returns drastically increases the expected material payoff of the agents, from which altruism is based upon. Meanwhile, the choice of the *right thing to do* depends only on the contract offered by the principal, and not on the production's underlying technology. Therefore, under increasing returns, altruistic agents possess higher

[5] Alger and Weibull (2013) shows that *homo moralis* and altruistic preferences are behaviorally alike in many situations, and a similar point can be found in Bergström (1995). However, this is not the case in this exposition, as will be seen later on.

intrinsic motivation to exert the high level of effort, thus demanding a less high-powered contract and saving costs for the principal.

This paper is closely related to the moral hazard literature, in particular to two of its strands: moral hazard in teams and moral hazard with prosocial preferences. Holmström (1982) and Mookherjee (1984) characterize the basic results on moral hazard in teams that are used to build the model below[6]. Itoh (2004), Englmaier and Wambach (2010), Rey-Biel (2008) and Livio (2015) study optimal incentive schemes under different prosocial preferences: the first three focus on inequity aversion, while the last models agents exhibiting reciprocity concerns towards each other. None of them, however, raises the question of which preferences yields the least cost to the principal.

The analysis below differs from the previous literature in two crucial points: first, it considers *homo moralis* preferences, which hasn't, to the best of my knowledge, been done before in a contracting setting, thus presenting a simple environment where the principal can profitably explore idiosyncracies generated by those and altruistic preferences. Second, and more importantly, it contrasts the optimal contracts under each class of preferences, and derives conditions under which the principal would prefer hiring one team over the other, therefore providing a rationale for firms to collect soft information on potential employees to compose teams that will minimize the total payments to be made.

I proceed as follows. Section 2 presents the environment, while Section 3 and 4 study the optimal contract assuming efforts are contractible and non-contractible, respectively. Section 5 concludes. For ease of exposition, all proofs are relegated to Appendix C.

2 The Model

I analyze the interaction between a principal and two agents, denoted by $i \in \{A, B\}$. The principal hires the two agents to work on a joint task, which generates revenue $x \in \{x^H, x^L\}$ to the principal, where $x^H > x^L$. Each agent can exert either a low or a high effort level

[6] Che and Yoo (2001) study optimal incentives for teams in a repeated setting.

$e_i \in \{0,1\}$. Efforts determine revenues stochastically, according to the following probability distribution

$$\begin{array}{cc} & \begin{array}{cc} e_B = 1 & e_B = 0 \end{array} \\ \begin{array}{c} e_A = 1 \\ e_A = 0 \end{array} & \left(\begin{array}{cc} p_2 & p_1 \\ p_1 & p_0 \end{array} \right) \end{array}$$

Throughout, I assume that revenue is never certain and that the probability of achieving a high outcome is increasing in the total effort exerted by the agents: $1 > p_2 > p_1 > p_0 > 0$.

If effort is costless, the assumption above indicates a preference of the principal for both agents to exert effort. However, effort is costly to each agent; for each $i = A, B$[7],

$$c(e_i) = \begin{cases} c > 0 & \text{if } e_i = 1, \\ 0 & \text{if } e_i = 0. \end{cases}$$

The principal offers the agents contracts $w_i(x)$, $i = A, B$, specifying payments that will follow each realization of revenues. The principal is assumed to be risk neutral, and his payoff is given by

$$V(x, w_A(x), w_B(x)) = x - w_A(x) - w_B(x).$$

Denote by $\pi(e_i, e_j, w_i(x))$ the expected *material* payoff accruing to agent i from the effort choices (e_i, e_j) and wage schedule $w_i(x)$, for $i, j \in \{A, B\}, i \neq j$. I restrict attention to wage schedules pairs $\mathbf{w}_i = (w_i^H, w_i^L)$ determining the payments following good and bad realizations of revenues. In what follows, the *material* payoff function takes the expected additively separable form

$$\pi(e_i, e_j, \mathbf{w}_i) = p_{e_i+e_j} \left[u(w_i^H) - c(e_i) \right] + (1 - p_{e_i+e_j}) \left[u(w_i^L) - c(e_i) \right], \tag{1}$$

where $u : \mathbb{R}_+ \to \mathbb{R}$ is the function that associates the agent's consumption utility to each amount of money. The dependence of i's expected material payoff on e_j comes from the effect of the other agent's effort on the probability distribution of revenues. The agents are

[7] In reality agents may differ in their respective cost of effort, but this is not pursued in this paper because it doesn't qualitatively change the results, at the same time it adds a more cumbersome notation.

risk averse towards wages: $u(w)$ is assumed to be twice-continuously differentiable, strictly increasing and strictly concave[8].

The principal faces either a team consisting of two agents characterized by *homo moralis* preferences with degree of morality $\kappa_i \in [0,1]$, represented by the utility function

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \kappa_i) = (1-\kappa_i)\pi(e_i, e_j, \mathbf{w}_i) + \kappa_i \pi(e_i, e_i, \mathbf{w}_i), \tag{2}$$

a team comprised of two altruistic agents, whose preferences are summarized by the utility function

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \alpha_i) = \pi(e_i, e_j, \mathbf{w}_i) + \alpha_i \pi(e_j, e_i, \mathbf{w}_j), \tag{3}$$

for $\alpha_i \in [0,1]$ and $i, j \in \{A, B\}$, $i \neq j$. Both specifications take the standard selfish preferences as a special case ($\kappa_i = \alpha_i = 0$), and this will allow comparisons between the results to be presented below and the benchmark moral hazard problem.

As pointed in Alger and Weibull (2013) and Bergström (1995), this specification of preferences for altruistic agents gives rise to the behavioral equivalence between *homo moralis* preferences and altruism for $\alpha_i = \kappa_i$ in many classes of games. With that in mind, I will make the following assumption for the rest of the exposition.

Assumption 1: $\alpha_A = \alpha_B = \kappa_A = \kappa_B = \theta$.

Thus, the agents' utility function are simplified to

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \theta) = (1-\theta)\pi(e_i, e_j, \mathbf{w}_i) + \theta \pi(e_i, e_i, \mathbf{w}_i), \tag{4}$$

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) = \pi(e_i, e_j, \mathbf{w}_i) + \theta \pi(e_j, e_i, \mathbf{w}_j). \tag{5}$$

The relationship among the three parties unfolds as follows. First, the principal offers each agent a contract $\mathbf{w_i}$, which can be either accepted or rejected by the agents. If at least one agent rejects the contract, the game ends and every party receives his own reservation

[8]The assumption that $u(\cdot)$ is strictly concave can be relaxed, and the same model below can be solved in a setting with risk-neutral agents and limited liability constraints, where the qualitative results are not changed from the analysis below.

utility. If both agents accept the principal's offers, they play a normal form game[9]: both of them must simultaneously and independently choose an effort level, from which revenues will be realized according to the probability distribution given by the production technology above. Payments are made according to the schedules proposed by the firm and the agents' payoffs in the normal form game are given by their expected utilities with regard to received wages, efforts and preferences. While each agent's effort choice is private information, revenues and wages are publicly observable. It is also assumed that the agents' preferences are common knowledge[10].

3 Studying the Benchmark: The Contractible Effort Case

As a starting point, I derive the optimal contract assuming efforts are observable and contractible by the principal, to serve as a benchmark for later results. In what follows, I assume that each agent possesses an outside option that gives him utility $\overline{u} \geq 0$ if he does not accept the principal's contract offer. Therefore, agent $i \in \{A, B\}$ is willing to participate in the proposed relationship iff

$$U(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) \geq \overline{u}. \tag{IR}$$

As discussed in the previous section, standard selfish preferences are a particular case of both *homo moralis* and altruistic preferences, and for ease of exposition, I begin this and the next section by analyzing the optimal contract for that instance. Thus, under contractible efforts, the standard Borch rule

$$\frac{p(e_i, e_j)}{u'(w_i^H)p(e_i, e_j)} = \frac{1 - p(e_i, e_j)}{u'(w_i^L)[1 - p(e_i, e_j)]}$$

[9]The normal form game here is comprised of the set of players $\{A, B\}$, the common set of pure strategies $S = \{0, 1\}$, and payoff function $U(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta)$.

[10]This is the reason why contracts are indexed by i.

gives

$$\frac{u'(w_i^H)}{u'(w_i^L)} = 1, \tag{BR_S}$$

which implies $w_i^H = w_i^L = w_i = u^{-1}(\overline{u} + c(e_i))$ for $e_i \in \{0,1\}$. The intuition here is the same as in the classical moral hazard problem with one principal and one agent: if effort is contractible, the principal optimally offers a constant wage schedule remunerating the agent according to his reservation utility and the cost of the principal's desired level of effort.

When the principal faces a team of altruistic agents, he solves

$$\begin{aligned}
\max_{\mathbf{w}_A,\mathbf{w}_B} \quad & p(e_A,e_B)\left(x_H - w_A^H - w_B^H\right) + (1-p(e_A,e_B))\left(x_L - w_A^L - w_B^L\right) \\
s.t. \quad & \left[p(e_A,e_B)u(w_A^H) + (1-p(e_A,e_B))u(w_A^L) - c(e_A)\right] \\
& \quad +\theta\left[p(e_A,e_B)u(w_B^H) + (1-p(e_A,e_B))u(w_B^L) - c(e_B)\right] \geq \overline{u} \quad (IR_A) \\
& \left[p(e_A,e_B)u(w_B^H) + (1-p(e_A,e_B))u(w_B^L) - c(e_B)\right] \\
& \quad +\theta\left[p(e_A,e_B)u(w_A^H) + (1-p(e_A,e_B))u(w_A^L) - c(e_A)\right] \geq \overline{u} \quad (IR_B)
\end{aligned}$$

An interior solution is characterized by the KKT first-order conditions

$$\begin{aligned}
&-p(e_i,e_j) + \lambda_i p(e_i,e_j)u'(w_i^H) + \lambda_j \theta p(e_i,e_j)u'(w_i^H) = 0 \\
&-[1-p(e_i,e_j)] + \lambda_i[1-p(e_i,e_j)]u'(w_i^L) + \lambda_j\theta[1-p(e_i,e_j)]u'(w_i^L) = 0,
\end{aligned}$$

so the Borch rule becomes[11]

$$\frac{u'(w_i^H)}{u'(w_i^L)} = 1, \tag{BR_{Alt}}$$

for any choices of effort $(e_A, e_B) \in \{0,1\}^2$ and $i, j \in \{A, B\}$, $i \neq j$. Therefore, if agents are altruistic, the optimal contract under verifiable efforts proposes a constant wage schedule, just as was the case in the benchmark selfish preferences, given by

$$w^{Alt} = u^{-1}\left(\frac{1}{1+\alpha}\overline{u} + c(e_i)\right), \tag{6}$$

which is well-defined for all $\theta \in [0,1]$. It is easy to see that for $\theta = 0$ this is exactly the same expression as for the optimal contract under verifiable efforts in the benchmark case, while that for any positive degree of altruism, it is lower than it would be for selfish agents.

[11] Since $1 \geq p_2 > p_1 > p_0 \geq 0$ and $u' > 0$ by assumption, the first-order conditions imply that $\lambda_i + \lambda_j\theta > 0$ for $i, j \in \{A, B\}$, $i \neq j$.

Finally, consider the team with *homo moralis* preferences, in which agent $i \in \{A, B\}$ is willing to participate in the proposed relationship iff

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \theta) = (1-\theta)\pi(e_i, e_j, \mathbf{w}_i) + \theta\pi(e_i, e_i, \mathbf{w}_i) \geq \overline{u},$$

i.e. iff

$$u(w_i^H)[(1-\theta)p(e_i, e_j) + \theta p(e_i, e_i)] + u(w_i^L)[(1-\theta)(1-p(e_i, e_j)) + \theta(1-p(e_i, e_i))] - c(e_i) \geq \overline{u}.$$

Some points are noteworthy. First, as mentioned before, in the case where $\theta = 0$ this participation constraint reduces to the usual (IR) constraint in the benchmark moral hazard problem, since the selfish preference is a particular case of this framework. Second, for $\theta = 1$, $U^{\kappa}(e_i, e_j, \mathbf{w}_i, 1) = \pi(e_i, e_i, \mathbf{w}_i)$. In this case, agent i's choice of effort does not depend on agent j's effort choice, and choosing e_i becomes an individual decision problem. Third, if $e_i = e_j = e \in \{0, 1\}$, the participation constraint collapses into

$$p(e, e)u(w_i^H) + (1 - p(e, e))u(w_i^L) - c(e) \geq \overline{u},$$

Note here that the agents' degrees of morality are irrelevant and the participation constraints are exactly the same as those that would be obtained in a symmetric equilibrium in the benchmark moral hazard problem: by imposing $e_i = e_j$, both expected material payoffs terms are identical, and since the utility function is constructed as a convex combination of these functions the expressions above are obtained.

By Assumption 1, every agent in each team is identical to his partner, since the only source of heterogeneity in the general formulation was given by the preferences. Therefore, I will restrict attention to symmetric choices of effort $e_A = e_B$ in the rest of the discussion[12].

Proposition 1: *Suppose Assumption* 1 *holds. Then, there exists* $c^* > 0$ *such that for all* $c \in (0, c^*)$ *the principal induces agents in the teams of moral or altruistic agents to exert high effort by means of a constant wage.*

[12] In Appendix B I show that relaxing both these assumptions leads to a Borch rule for moral agents that demands nonconstant wages when efforts are observable, in stark contracts to the literature with selfish and altruistic agents.

This result is not surprising: if efforts are contractible, the principal compensates the agents with a fixed transfer in case they exert the desired level of effort, or punish them if there is a deviation. Also, if the cost of exerting effort is small, then the amount the principal has to transfer back to the agents in order to have an increased chance of obtaining a high realization of revenues is also small, and thus profitable to implement. Moreover, since I restrict attention to symmetric equilibria, the degree of morality plays no role when I consider a team of *homo moralis* agents: the compensation schedule and effort choices are exactly the same as those obtained in the benchmark problem.

Now I can focus on the central question of the paper: given the optimal contracts that induce the desired level of effort, which team should the principal hire?

Proposition 2: *Suppose assumption* 1 *holds, efforts are verifiable and the principal wants both agents to make the high effort. For any* $\theta \in [0, 1]$*, the principal prefers hiring the team of altruistic agents team of selfish and moral agents.*

Intuitively, in situations where someone fails to do the right thing, a moral agent derives part of his utility from contemplating what would happen if everyone did the right thing. If both agents do in fact exert high effort, the contemplation in question does not add utility beyond the material utility that the agents thus obtain. By contrast, for altruistic agents, any choice but high effort decreases the material payoff of both agents, and consequently all the utility of each altruistic employee. Therefore, intrinsic motivation is larger for altruistic agents and a team comprised of such employees is less costly for the principal.

4 Moving to the Second Best: Non-contractible Efforts

Throughout the rest of the exposition, I focus on contracts that induce both agents to participate in the relationship and also exert the high level of effort ($e = 1$).

As a benchmark, focus first on standard selfish preferences. If efforts are non-contractible and the principal wishes to induce both agents to exert effort, he must solve, for $i, j \in \{A, B\}$,

$i \neq j$,

$$\begin{array}{lll} \max_{\mathbf{w}_A,\mathbf{w}_B} & p_2(x^H - w_A^H - w_B^H) + (1-p_2)(x^L - w_A^L - w_B^L) & \\ s.t. & p_2\left[u(w_i^H) - c\right] + (1-p_2)\left[u(w_i^L) - c\right] \geq p_1 u(w_i^H) + (1-p_1)u(w_i^L) & (IC_i) \\ & p_2\left[u(w_i^H) - c\right] + (1-p_2)\left[u(w_i^L) - c\right] \geq \overline{u}. & (IR_i) \end{array}$$

Manipulating the incentive compatibility constraint yields

$$u(w_i^H) - u(w_i^L) \geq \tfrac{c}{p_2 - p_1}. \quad (IC_i)$$

By assumption, $c > 0$ and $p_2 > p_1$. Thus, the incentive compatibility constraint implies a monotonicity constraint on the wages following a good and a bad realization of output, since $u(.)$ is assumed to be strictly increasing. Standard arguments show that both the incentive compatibility and the individual rationality constraints must bind at the optimum, so that the solution to the principal's problem is a contract $\mathbf{w}_S = (w_S^H, w_S^L)$ such that

$$\begin{aligned} u(w_S^L) &= \overline{u} - \frac{p_1 c}{p_2 - p_1}, \\ u(w_S^H) &= \overline{u} + \frac{(1-p_1)c}{p_2 - p_1}. \end{aligned}$$

Given the incentive compatibility constraint, it is clear that $\Delta w_S \equiv w_S^H - w_S^L > 0$.

Of course, if the principal wishes to induce the agents not to exert effort, a constant wage schedule $w^H = w^L = w = u^{-1}(\overline{u})$ would be optimal. Comparison of the principal's profits when agents exert effort and shirk show that the former is preferred by the employer for any $c \leq \overline{c}_S$, $0 < \overline{c}_S < c^*$.

Under altruistic preferences for the agents, the principal's problem is

$$\begin{array}{lll} \max_{w^H,w^L} & p_2(x^H - 2w^H) + (1-p_2)(x^L - 2w^L) & \\ s.t. & (1+\theta)[p_2 u(w^H) + (1-p_2)u(w^L) - c] \geq \overline{u} & (IR) \\ & (1+\theta)\left[p_2 u(w^H) + (1-p_2)u(w^L) - c\right] \geq & \\ & \quad \left[p_1 u(w^H) + (1-p_1)u(w^L)\right] + \theta\left[p_1 u(w^H) + (1-p_1)u(w^L) - c\right]. & (IC) \end{array}$$

Rewrite the incentive compatibility constraint as

$$u(w^H) - u(w^L) \geq \frac{c}{(1+\theta)(p_2 - p_1)}, \tag{7}$$

and notice that the right-hand side is strictly decreasing in the degree of altruism α. The intuition behind this is the tradeoff between explicit and intrinsic incentives. Indeed, as the agent cares less about his own material payoff relative to that of his teammate, the intrinsic incentive derived from an increase in the probability of a high realization of output (and a consequent raise in the expected material benefit of his partner) becomes larger than the explicit incentives given by a high powered contract in inducing the agent to exert the high level of effort.

The proposition below characterizes the optimal contract[13].

Proposition 3: *Suppose Assumption* 1 *holds. There exists* $\bar{c}_{Alt} > \bar{c}_S$ *such that, for all* $c < \bar{c}_{Alt}$, *it is optimal for the principal to induce both altruistic agents to exert effort,* $e_A = e_B = 1$, *by means of a contract* $\mathbf{w_{Alt}}^* = (w^H_{Alt}, w^L_{Alt})$ *such that*

$$\Delta w_{Alt} \equiv w^H_{Alt} - w^L_{Alt} \leq \Delta w_S$$

$\forall \theta \in [0, 1]$, *with strict inequality for any* $\theta > 0$.

Close inspection of the incentive compatibility constraint shows that any contract that would induce a selfish agent to exert the high effort would also induce an altruistic employee to do the same. Also, for any given contract an increase in θ would increase the utility of each agent. Hence, the principal can profit by reducing both wages[14]. This argument is formally stated below.

Corollary 1: *Suppose Assumption* 1 *holds. Then, the principal's expected profits are strictly increasing in* θ.

Last, I consider *homo moralis* preferences. The principal must choose wage vectors

[13]In Appendix A I show that the optimal contracts for both moral and altruistic agents may lead to multiplicity of equilibria in their effort choices, as in Holmström (1982).

[14]See Appendix C for the proof.

$\mathbf{w}_A, \mathbf{w}_B$ to solve

$$
\begin{aligned}
\max_{\mathbf{w}_A,\mathbf{w}_B} \quad & p_2\left(x_H - w_A^H - w_B^H\right) + (1-p_2)\left(x_L - w_A^L - w_B^L\right) \\
s.t. \quad & (1-\theta)\left[p_2(u(w_i^H)-c) + (1-p_2)(u(w_i^L)-c)\right] \\
& \quad +\theta\left[p_2(u(w_i^H)-c) + (1-p_2)(u(w_i^L)-c)\right] \geq \\
& \quad (1-\theta)\left[p_1 u(w_i^H) + (1-p_1)u(w_i^L)\right] \\
& \quad +\theta\left[p_0 u(w_i^H) + (1-p_0)u(w_i^L)\right] \qquad\qquad (IC_i) \\
& (1-\theta)\left[p_2(u(w_i^H)-c) + (1-p_2)(u(w_i^L)-c)\right] \\
& \quad +\theta\left[p_2(u(w_i^H)-c) + (1-p_2)(u(w_i^L)-c)\right] \geq \overline{u} \quad (IR_i)
\end{aligned}
$$

for $i = A, B$. Note that the individual rationality constraint can be rewritten in the simpler form

$$p_2 u(w_i^H) + (1-p_2)u(w_i^L) - c \geq \overline{u},$$

since, in equilibrium, the principal's offer induces the symmetric effort choice $e_A = e_B = 1$.

The incentive compatibility constraint also simplifies to

$$
\begin{aligned}
& p_2 u(w_i^H) + (1-p_2)u(w_i^L) - c \geq \\
& \quad (1-\theta)\left[p_1 u(w_i^H) + (1-p_1)u(w_i^L)\right] + \theta\left[p_0 u(w_i^H) + (1-p_0)u(w_i^L)\right].
\end{aligned}
$$

The right-hand side of this inequality highlights an interesting fact: a positive degree of morality implies the agent internalizes the cost of choosing a low effort by evaluating what would happen if the other agent also were to make the same decision. This is very different in nature to how an altruistic agent evaluates any deviation: while the latter considers only the effects of his own deviation on his own material payoff and on his partner's, the former would consider the effect of the same deviation being made by his partner on his own payoff. Besides, by force of the assumptions presented above, further manipulation of (IC_i) yields

$$u(w_i^H) - u(w_i^L) \geq \frac{c}{(p_2-p_1) + \theta(p_1-p_0)}, \tag{8}$$

where $\frac{c}{(p_2-p_1)+\theta(p_1-p_0)} > 0$. Because of the strict concavity of $u(\cdot)$, the incentive compatibility constraint for moral agents also implies a monotonicity condition on the optimum compensation schedules offered by the principal, even when the agents display the highest degree of morality. This last remark implies that the intrinsic incentives of the most moral agent are not sufficiently large to overcome the need to provide him with explicit incentives to exert the high level of effort.

Proposition 4: *Suppose Assumption* 1 *holds. There exists* $\bar{c}_{HM} > \bar{c}_S$ *such that, for all* $c < \bar{c}_{HM}$*, it is optimal for the principal to induce both moral agents to exert effort,* $e_A = e_B = 1$*, by means of a contract* $\mathbf{w_{HM}}^* = (w_{HM}^H, w_{HM}^L)$ *such that*

$$\Delta w_{HM} \equiv w_{HM}^H - w_{HM}^L \leq \Delta w_S$$

$\forall \theta \in [0, 1]$, *and strict inequality for* $\theta > 0$.

The intuition behind the monotonicity constraint is the same as in the benchmark model: if wages following a low realization of revenues were larger than their counterpart after a good realization, then agents would prefer to exert low effort in order to receive this higher compensation and save in the cost of exerting effort.

The novelty in the results relates to how the compensation schedules varies with respect to the degree of morality θ. Keeping in mind that $u' > 0$, one can see that as θ increases the right hand side of (IC_i) becomes ever smaller, albeit positive. This implies that the gap in wages following good and bad realizations of revenues must decrease, since the incentive compatibility constraint binds, but monotonicity still holds. Intuitively, the principal can reduce the compensation over high realizations of revenues given to an agent who is very concerned about *doing the right thing.* But at the same time, he must increase wages after bad outcomes in order to satisfy the participation constraint.

Because of this diminishing wage gap, intuition would suggest the first-best result is obtained for a sufficiently high degree of morality. However, this is not the case. To see this, take $\theta = 1$, where agent i's preferences are purely Kantian and, thus, his utility is completely characterized by the expected material payoff $\pi(e_i, e_i)$. Although the participation constraint doesn't vary with the agent's degree of morality[15], the same is not true for the incentive

[15]For symmetric equilibrium choices of effort.

compatibility constraint. Now, when considering the pros and cons of a deviation in terms of effort choice, agent i internalizes what would happen if agent j were to do the same. Specifically, this entails a reduction in the probability of the good revenue being realized from p_2 to p_0, instead of the reduction to p_1 in the selfish term. This internalization is reflected in the (IC_i) constraint, which becomes

$$u(w_i^H) - u(w_i^L) = \frac{c}{p_2 - p_0} > 0.$$

The denominator on the right-hand side is exactly the difference in the probabilities discussed above. Taking $\theta = 1$ makes the incentive compatibility constraint for a team of moral agents as easy to satisfy as possible, but it still binds, thus pushing the optimal contract away from the first-best one (constant wage schedule).

Given this behavior of wage schedules with respect to the degree of morality, a natural question to be asked is whether the principal is better off with highly moral agents or not. The answer is unconditional, and presented in the following result.

Corollary 2: *The principal's expected profit is strictly increasing in θ.*

Corollary 2 contrasts with the contractible effort case, where the principal's profits were identical when hiring a team of selfish agents or a team of moral agents, for any degree of morality the last would display. Mathematically, the result is a consequence of the individual rationality constraints being identical in both cases, while the incentive compatibility constraint has a smaller right-hand side under moral agents than under selfish ones. Intuitively, the principal exploits the agents' morality, as he did with altruistic employees as well, to induce high effort by means of less high-powered incentives, while inducing participation with a slightly increased payment after a bad realization of output. Thus, one concludes that the expected savings in wages after a good realization made by the principal by choosing a high morality agent offsets the expected increase in payments after low revenues.

One remark is in order here. Because of the assumption that $1 > p_2 > p_1 > p_0 > 0$ and the monotonicity condition implied by the incentive compatibility constraints for each preferences, it is the case that $U(0, 1, \mathbf{w}; \theta) \geq U(0, 0, \mathbf{w}; \theta)$, where $\mathbf{w}$ is the optimal contract offered by the principal. Thus, using the (IC) constraints again, I have that $U(1, 1, \mathbf{w}; \theta) \geq U(0, 0, \mathbf{w}; \theta)$, so the agents have no incentives to jointly deviate to shirking. The same is also true for altruistic agents.

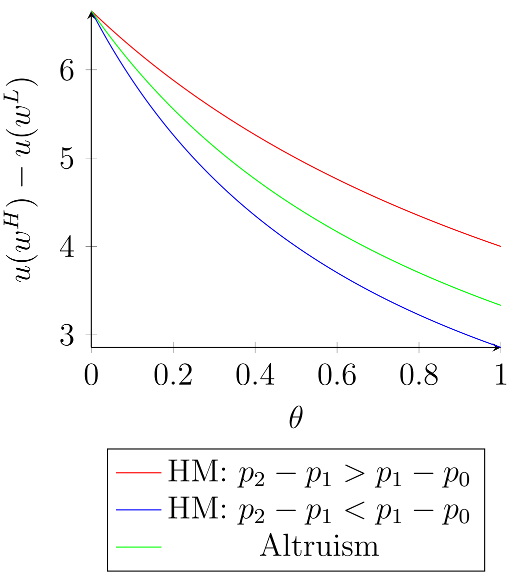
So far, I have showed that the principal can attain higher profits by exploiting the agents' morality or altruism, thus reducing high-powered explicit incentives in the optimal contract in such a way that participation and incentives to exert high effort are still satisfied. Therefore, from the employer's perspective, knowing which class of preferences demands the least amount of explicit incentives is crucial. Lemma 1 tells us that the answer to that question depends on the stochastic production technology.

Lemma 1: *Under Assumption* 1, $\left[u(w^H_{HM}) - u(w^L_{HM})\right] - \left[u(w^H_{Alt}) - u(w^L_{Alt})\right]$ *has the same sign as* $(p_2 - p_1) - (p_1 - p_0)$.

In other words, if the stochastic technology presents increasing returns on aggregate efforts, the optimal contract under *homo moralis* preferences is (weakly) more high-powered than its counterpart under altruism when $\kappa = \alpha$: any contract inducing moral agents to exert high effort would do the same to altruistic employees. The converse is true if the technology has decreasing returns on efforts. This can be seen in Figure 1. The middle (green) line represents the incentive compatibility constraint for altruistic agents, whose format is not affected by the production technology. The top (red) and the bottom (blue) lines are the graphic representations of the (IC) constraint for moral agents when $p_2 - p_1 > p_1 - p_0$ and $p_2 - p_1 < p_1 - p_0$, respectively[16].

[16] If $p_2 - p_1 = p_1 - p_0$, both lines coincide with the (IC) for altruistic agents.

Figure 1: Comparing the power of optimal contracts



Given the result above, one would expect that the principal's expected payoff will be uniformly higher if the agents are altruistic rather than moral when the production technology presents increasing returns to efforts, while the opposite would be true if decreasing returns are present. The flaw with such a logic is not considering the effects of the binding individual rationality constraints, which implied under contractible efforts that the team of altruistic agents was always the cheapest to hire. In particular, remember that for altruistic agents the outside option $\overline{u}$ is divided by $1+\theta$ in the participation constraint, a factor that is not present under selfish and moral preferences. This implies that w^L_{Alt} should also be smaller than w^L_{HM}[17]. However, the principal has clear preferences over the composition of the team, and the result below precisely states when one team is preferred over the other.

Theorem 1: *Assume the principal offers contracts* $\mathbf{w}_{HM}$ *and* $\mathbf{w}_{Alt}$ *to homo moralis and altruistic agents, respectively, inducing them to exert the high level of effort. Also, assume Assumption* 1 *holds. Then, if the stochastic production technology exhibits*

1. *increasing returns to efforts* ($p_2 - p_1 \geq p_1 - p_0$)*, the principal is better off hiring a team of altruistic agents over a team of moral agents;*

[17]This intuition in indeed right, and integrates the proof of Theorem 1 in the Appendix.

2. *decreasing returns to efforts ($p_2 - p_1 < p_1 - p_0$) and*

- *the outside option is zero ($\overline{u} = 0$), the principal prefers a team of moral agents; or*
- *the outside option is positive and the degree of morality is sufficiently low ($\overline{u} > 0$, $\kappa \to 0$), the principal prefers a team of moral agents only if $p_2 > \overline{p}_2 \in (0, 1)$.*

Under increasing returns to efforts, an altruistic team is cheaper for the principal because of two reasons. First, the wage that must be paid after a bad realization of output is smaller than its counterparts under selfish or moral preferences, and this is a consequence of the fact that the former's consideration with regards to the payoff of his partner slackens the participation constraint. On the other hand, such concern also slackens the incentive compatibility constraint in this case, because exerting efforts drastically increases the probability of being successful, thus providing implicit incentives for the altruistic worker to exert effort and requiring a less high-powered contract to be proposed by the employer.

Such a difference in the intrinsic incentives to exert effort disappear when the production technology has constant returns, so that the power of the contract remains the same for both teams. However, it is still the case that the principal exploits the fact that altruistic agents derive utility from each other's material payoff, and can thus pay them less.

The third case, with decreasing returns to efforts, is the most interesting, because the preference of the principal results from the net effect of two opposing forces. While it is still true that $w^L_{Alt} \leq w^L_{HM}$, Lemma 1 states that now the power of the contract required by moral agents is smaller than the one for altruistic agents. In the range where such a reduction is the most drastic, the principal will prefer the team of moral agents rather than altruistic ones. The first condition for this to happen is that the probability of a success when both agents are exerting effort is sufficiently high, as can be seen from the incentive compatibility constraints. The second condition is that either the outside option for the agents be zero, or that if it is positive, the degree of morality or altruism be close to zero. If both cases, the participation constraints for moral and altruistic agents become arbitrarily close (identical if $\overline{u} = 0$) so that the exploitability of altruistic preferences, described in the preceding paragraph, becomes small, and the principal profits by hiring the agents demanding the least powered contracts: the moral agents in this case.

Figures $2a$, $2b$ and $2c$ provide an example of Theorem 1 for $u(w) = \sqrt{w}$. Figure $2a$ represents the case where the production technology exhibits increasing returns to efforts. With the exception of $\theta = 0$, where both teams are identical to the selfish agents, the principal's profit is higher with a team of altruistic agents (V^{Alt}) than with a team of moral agents (V^{HM}).

Figure 2a: Comparing principal's profits for $p_2 - p_1 \geq p_1 - p_0$

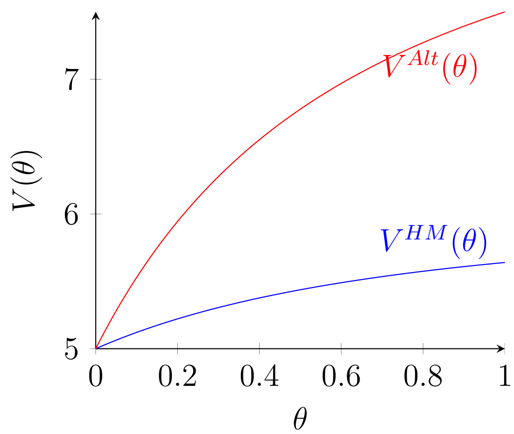
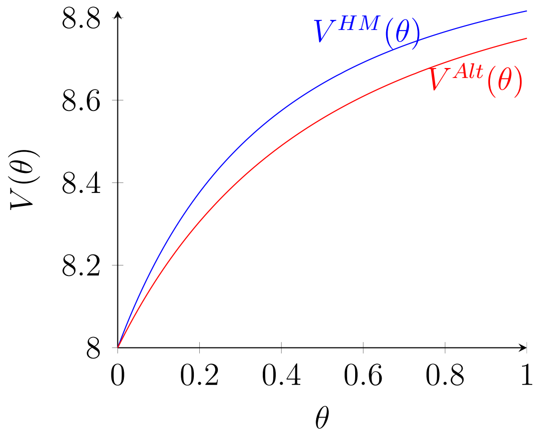


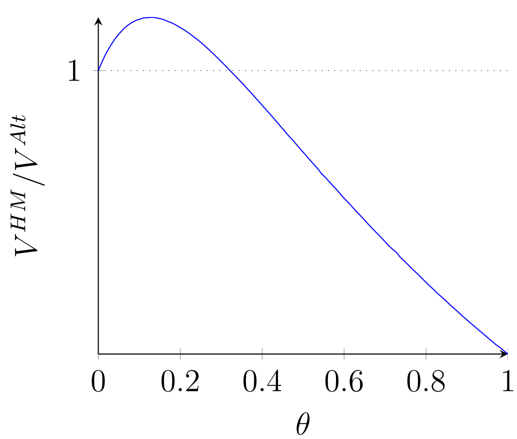
Figure $2b$ exemplifies the case with decreasing returns to efforts, zero outside option for the agents and a high probability of success if both agents exert effort (namely, I set $p_2 = 0.9$). As Theorem 1 states, under these conditions $V^{HM}(\theta) \geq V^{Alt}(\theta)$ for all equal degrees of morality and altruism.

Figure 2b: Comparing principal's profits for $p_2 - p_1 < p_1 - p_0$



Finally, Figure 2c plots the ratio V^{HM}/V^{Alt} for decreasing returns to efforts and $\overline{u} = 0.2$. Contrary to the previous case where $\overline{u} = 0$, the difference in the participation constraints for moral and altruistic agents make it unprofitable to the employer hiring the moral team if θ becomes larger, since the decrease in w^L_{Alt} would be, in expected terms, sufficient to compensate the savings related to the power of the contract. This is represented by the region in the figure in which $V^{HM}/V^{Alt} \leq 1$.

Figure 2c: Comparing the ratio of principal's profits for $p_2 - p_1 < p_1 - p_0$



Therefore, a rationale in terms of the principal's expected profits is given for trying to sort employees with respect to their preferences. If the production technology exhibits increasing returns with respect to efforts, the principal's choice is straightforward: always choose to employ altruistic agents. However, if the condition does not hold, employing moral individuals may lead to higher profits in comparison to both altruistic and purely selfish agents.

5 Concluding Remarks

This paper presents a comparison between the optimal contracts offered to teams of agents, who may be characterized by either *homo moralis* preferences or altruism towards each other. These contracts were explored in situations where the teams have only two agents with binary choices of efforts, affecting stochastically the revenues accrued by the principal.

Under contractible efforts, I show that altruistic agents are more exploitable by the principal, in the sense that the employers needs to pay a smaller wage to induce participation of the those agents when compared to the case where he would hire a team of selfish or moral employees. When efforts are no longer contractible, this exploitability also shows up for moral agents, and I show that the larger the degree of altruism or morality displayed by the members of each team, the higher the expected profits for the principal. Then, the natural question is which class of preferences would require smaller wages to exert effort and participate in the contractual relationship?

The main finding is that the principal obtains a higher expected profit hiring a team composed of moral agents under restrictive conditions: first, that the stochastic technology exhibits decreasing returns with respect to efforts; second, that the outside option of the agents yield zero utility or third, that the degree of morality is sufficiently low.

It is noteworthy that even in such a simple environment prosocial preferences affect the contractual design, by adding a third channel to the traditional trade-off between risk-sharing and incentive provision. In effect, the principal will be better off employing a team of either altruistic or moral agents instead of a team composed solely of selfish employees, since a higher degree of morality and altruism decreases the amount of explicit incentives provided by the optimal contracts to induce the agents to exert effort. However, this additional channel is not enough to completely extinguish the need for explicit incentives even then the agents are purely moral or altruistic. Because it is more costly to the principal to hire a team of selfish agents, the exploitability of prosocial preferences can thus explain costly acquisition of job applicants' soft information in the labor market. These departures in terms of the incentive compatibility constraints can be used to empirically test the underlying preferences of employees, by means of the powers of the contracts. This is particularly true when monitoring of the employees' activities is available to the employer, since selection of non-constant contracts would be evidence of moral agents.

A Multiplicity of Equilibria

The optimal contracts derived in the main text are such that the principal can induce both agents to exert the high level of effort $e = 1$, both for the *homo moralis* or altruistic teams. However, the strategy profile $(e_A, e_B) = (1, 1)$ may not be the unique Nash equilibrium of the simultaneous game played by the pair agents for a given contract. Indeed, let $U(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j; \theta)$ denote agent i's expected utility under this contract and degree of morality or altruism θ[18]. Then, the stage game played by agents 1 and 2 can represented by

	$e_B = 1$	$e_B = 0$
$e_A = 1$	$U(1, 1, \mathbf{w}_A, \mathbf{w}_B; \theta)$	$U(1, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$
$e_A = 0$	$U(0, 1, \mathbf{w}_A, \mathbf{w}_B; \theta)$	$U(0, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$

From the principal's problem, the incentive compatibility constraint implies that

$$U(1, 1, \mathbf{w}_A, \mathbf{w}_B; \theta) \geq U(0, 1, \mathbf{w}_A, \mathbf{w}_B; \theta),$$

i.e. given that the other agent is already exerting the high level of effort, it is not profitable for agent A to shirk when his compensation follows the optimal contract $\mathbf{w}^*$. Since this is true for both agents, it follows that $(e_A, e_B) = (1, 1)$ is a pure strategy Nash equilibrium for $\mathbf{w}^*$.

However, the comparison between $U(1, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$ and $U(0, 0, \mathbf{w}_A, \mathbf{w}_B; \theta)$ is not clear. In particular, if the latter is greater than the former, then $(e_A, e_B) = (0, 0)$ would constitute another pure strategy Nash equilibrium for $\mathbf{w}^*$.

[18]Given the symmetry of the problem, I focus attention on agent A and drop the subscripts. The same results would hold for agent B by simply reversing the effort choices e_A and e_B.

Under *homo moralis* preferences, note that

$$U(0,0,\mathbf{w}^*;\theta) \geq U(1,0,\mathbf{w}^*;\theta) \Leftrightarrow$$

$$p_0 u(w^H) + (1-p_0)u(w^L) \geq (1-\theta)[p_1 u(w^H) + (1-p_1)u(w^L) - c] + \theta[p_2 u(w^H) + (1-p_2)u(w^L) - c] \Leftrightarrow$$

$$c \geq u(w^H)[(1-\theta)p_1 + \theta p_2 - p_0] + u(w^L)[(1-\theta)(1-p_1) + \theta(1-p_2) - (1-p_0)] \Leftrightarrow$$

$$c \geq u(w^H)[(1-\theta)p_1 + \theta p_2 - p_0] - u(w^L)[(1-\theta)p_1 + \theta p_2 - p_0] \Leftrightarrow$$

$$c \geq [u(w^H) - u(w^L)][(1-\theta)p_1 + \theta p_2 - p_0] \Leftrightarrow$$

$$c \geq \frac{c}{(p_2 - p_1) + \kappa(p_1 - p_0)}[(1-\kappa)p_1 + \kappa p_2 - p_0] \Leftrightarrow$$

$$(p_2 - p_1) + \theta(p_1 - p_0) \geq (p_1 - p_0) + \theta(p_2 - p_1) \Leftrightarrow$$

$$(p_2 - p_1) - (p_1 - p_0) \geq \theta[(p_2 - p_1) - (p_1 - p_0)] \Leftrightarrow$$

$$(1-\theta)\left[(p_2 - p_1) - (p_1 - p_0)\right] \geq 0,$$

while for altruistic preferences

$$U(0,0,\mathbf{w_{Alt}}^*;\theta) \geq U(1,0,\mathbf{w_{Alt}}^*;\theta) \Leftrightarrow$$

$$(1+\theta)[p_0 u(w^H_{Alt}) + (1-p_0)u(w^L_{Alt})] \geq (1+\theta)[p_1 u(w^H_{Alt}) + (1-p_1)u(w^L_{Alt})] - c \Leftrightarrow$$

$$c \geq (1+\theta)(p_1 - p_0)[u(w^H_{Alt}) - u(w^L_{Alt})] = (1+\theta)(p_1 - p_0)\frac{c}{(1+\theta)(p_2 - p_1)} \Leftrightarrow$$

$$p_2 - p_1 \geq p_1 - p_0.$$

Observe that the inequality for moral agents is always satisfied if $\theta = 1$. On the other hand, if $\kappa \in [0,1)$, that inequality holds iff $p_2 - p_1 \geq p_1 - p_0$. Therefore, the following result holds.

Lemma 2: *Suppose Assumption* 1 *holds and that the principal offers the optimal contracts* $\mathbf{w}_{HM}$ *and* $\mathbf{w}_{Alt}$ *for the teams of moral and altruistic agents, respectively. Then,* $e_A = e_B = 1$ *is the unique symmetric Nash equilibrium of the simultaneous choice of effort game played by the agents iff* $p_2 - p_1 < p_1 - p_0$*, and* $\theta < 1$ *for homo moralis agents. Otherwise,* $e_A = e_B = 0$ *is also a symmetric Nash equilibrium.*

One remark about asymmetric equilibria must be made here. If $p_2 - p_1 = p_1 - p_0$, both an altruistic and a moral agent will be indifferent between shirking and exerting effort when their partners are shirking. Moreover, since the optimal contract satisfies the incentive compatibility constraint with equality for both types of pro-social preferences, the workers are also indifferent between shirking or not when their partner is exerting the high effort. Therefore, in this case, the asymmetric efforts $(e_A = 1, e_B = 0)$ and $(e_A = 0, e_B = 1)$ are also pure strategy Nash equilibria of the simultaneous choice of effort game.

B Obtaining the Borch Rule for Asymmetric Efforts Under *Homo Moralis* Preferences

Relax Assumption 1 and consider the contracting problem of a team of moral agents when efforts are observable. If the principal wishes to induce asymmetric choices of effort, constant wages are not optimal for moral agents, as they were to altruistic and selfish agents[19]. Note first that the Borch rule for teams of selfish and altruistic agents are derived for an arbitrary pair (e_A, e_B), and the ratio of marginal utilities with high or low wages are equal to 1 whether $e_A = e_B$ or not. For *homo moralis* preferences suppose, without loss of generality, that agent A exerts high effort while agent B exerts low effort. In this case, the principal solves

$$
\begin{aligned}
\max_{\mathbf{w}_A, \mathbf{w}_B} \quad & p_1\left(x_H - w_A^H - w_B^H\right) + (1-p_1)\left(x_L - w_A^L - w_B^L\right) \\
s.t. \quad & (1-\kappa_A)\left[p_1 u(w_A^H) + (1-p_1)u(w_A^L)\right] \\
& \qquad +\kappa_A\left[p_2 u(w_A^H) + (1-p_2)u(w_A^L)\right] - c \geq \overline{u} \qquad (IR_A) \\
& (1-\kappa_B)\left[p_1 u(w_B^H) + (1-p_1)u(w_B^L)\right] \\
& \qquad +\kappa_B\left[p_0 u(w_B^H) + (1-p_0)u(w_B^L)\right] \geq \overline{u} \qquad (IR_B)
\end{aligned}
$$

Close observation of the constraints reveals two differences between them. First, only (IR_A) contains the cost of effort, since agent A is the only one to exert high effort. Second, and more important, the probabilities of high and low realizations of revenues in the Kantian

[19] As shown by the Borch rules (BR_S), (BR_{Alt})

morality terms of the two constraints are different, but the same in the other term. This is true because each agent evaluates the consequence of his own effort should both agents choose this particular effort.

The wages must satisfy the Borch rule, given by

$$\frac{p_1}{u'(w_A^H)\left[(1-\kappa_A)p_1+\kappa_A p_2\right]} = \frac{1-p_1}{u'(w_A^L)\left[(1-\kappa_A)(1-p_1)+\kappa_A(1-p_2)\right]},$$
$$\frac{p_1}{u'(w_B^H)\left[(1-\kappa_B)p_1+\kappa_B p_0\right]} = \frac{1-p_1}{u'(w_B^L)\left[(1-\kappa_B)(1-p_1)+\kappa_B(1-p_0)\right]}.$$

Observe that the usual finding that $w_i^H = w_i^L = w_i^{FB}$ is only obtained if $\kappa_i = 0$, that is, only if both agents display the standard selfish preferences. If the degree of morality is not zero, the marginal utility ratios must be such that

$$\frac{u'(w_A^H)}{u'(w_A^L)} = \frac{(1-\kappa_A)p_1(1-p_1)-\kappa_A p_1 p_2+\kappa_A p_1}{(1-\kappa_A)p_1(1-p_1)-\kappa_A p_1 p_2+\kappa_A p_2} < 1,$$
$$\frac{u'(w_B^H)}{u'(w_B^L)} = \frac{(1-\kappa_B)p_1(1-p_1)-\kappa_B p_1 p_0+\kappa_B p_1}{(1-\kappa_B)p_1(1-p_1)-\kappa_B p_1 p_0+\kappa_B p_0} > 1,$$

which implies the optimal contract satisfies $w_A^H > w_A^L$ and $w_B^H < w_B^L$.

Therefore, should the principal want to induce the moral agents to undertake different efforts, two differences arise in comparison to the selfish and altruistic preferences cases. First, the general argument that the principal should pay a constant wage (that satisfies the participation constraint) in case the appropriate level of effort is exerted by the agent no longer holds. Indeed, for the agent exerting high effort, a monotonicity result similar to the one obtained in the second-best cases of the traditional moral hazard problems is observed. On the other hand, agent B, who is not supposed to exert effort, is paid according to a reverse monotonicity result: wage after a good realization of revenue must be lower than its counterpart after a bad realization. These results are summarized in the next proposition.

Proposition 5: *Suppose the principal restricts attention to asymmetric equilibria of the kind $e_i = 1 > e_j = 0$ for $i, j \in \{A, B\}$, $i \neq j$ when the agents exhibit homo moralis preferences. Then, there does not exist a constant contract ($w_i^H = w_i^L$ and $w_j^H = w_J^L$) that maximizes the principal's profits and satisfies the agents participation constraints.*

C Proofs

C.1 Proof of Proposition 1

Assume $e_i = e_j = e \in \{0, 1\}$ for $i \in \{A, B\}$. The discussion in the main text shows that the optimal contract under contractible efforts for teams of selfish or moral agents is

$$w^{FB} = u^{-1}\left(\overline{u} + c(e)\right),$$

while altruistic agents must be compensated according to

$$w^{FB}_{i,Alt} = u^{-1}\left(\frac{1}{1+\theta}\overline{u} + c(e)\right).$$

Denoting by V^{FB}_{11} and V^{FB}_{00} the principal's expected profits in the cases where both agents exert high and low effort, respectively, and the teams are comprised of either selfish or moral agents. Plugging in the optimal wages obtained above yields

$$V^{FB}_{11} = p_2 x^H + (1-p_2)x^L - 2u^{-1}(\overline{u} + c)$$
$$V^{FB}_{00} = p_0 x^H + (1-p_0)x^L - 2u^{-1}(\overline{u})$$

and thus $V^{FB}_{11} \geq V^{FB}_{00}$ if and only if

$$u^{-1}(\overline{u} + c) - u^{-1}(\overline{u}) \leq \frac{(x^H - x^L)(p_2 - p_0)}{2}.$$

By assumption, $p_2 > p_0$ and $x^H > x^L$, so the right-hand side is strictly positive, while $u' > 0$ implies the left-hand side is also positive since $c > 0$. By continuity of u, there exists $c' > 0$ such that $2\left(u^{-1}(\overline{u} + c') - u^{-1}(\overline{u})\right) = (x^H - x^L)(p_2 - p_0)$, and the inequality above holds for all $c \in (0, c']$.

Now, doing the same for altruistic agents, write

$$V^{FB}_{11,Alt} = p_2 x^H + (1-p_2)x^L - 2u^{-1}\left(\frac{1}{1+\theta}\overline{u} + c\right)$$
$$V^{FB}_{00,Alt} = p_0 x^H + (1-p_0)x^L - 2u^{-1}\left(\frac{1}{1+\theta}\overline{u}\right)$$

where an argument similar to the paragraph above implies that there exists $c'' > 0$ such that, $\forall c \in [0, c'')$, $V^{FB}_{11,Alt} > V^{FB}_{00,Alt}$.

Letting $c^* = \min\{c', c''\}$ concludes the proof.

C.2 Proof of Proposition 2

Since u is a strictly increasing strictly concave function by assumption, its inverse u^{-1} is, on its turn, a strictly increasing strictly convex function. Therefore, since

$$\overline{u} \geq \overline{u}\frac{1}{1+\theta}$$

for all $\theta \in [0,1]$, the amount of compensation dispensed by the principal is larger under *homo moralis* or selfish preferences.

C.3 Proof of Proposition 3

Under Assumption 1 and $e_A = e_B = 1$, the optimal symmetric contract $\mathbf{w_{Alt}}^* = (w^H_{Alt}, w^L_{Alt})$ offered by the principal must solve

$$\begin{array}{lll} \max_{w^H,w^L} & p_2(x^H - 2w^H) + (1-p_2)(x^L - 2w^L) & \\ s.t. & (1+\theta)[p_2 u(w^H) + (1-p_2)u(w^L) - c] \geq \overline{u} & (IR) \\ & u(w^H) - u(w^L) \geq \frac{c}{(1+\theta)(p_2-p_1)} & (IC) \end{array}$$

where the KKT conditions are necessary and sufficient for optimality, and given by

$$u'(w^H_{Alt})(1+\theta)[\lambda p_2 + \mu(p_2 - p_1)] = p_2 \tag{2.1}$$

$$u'(w^L_{Alt})(1+\theta)[\lambda(1-p_2) - \mu(p_2 - p_1)] = (1-p_2) \tag{2.2}$$

$$\lambda\left\{(1+\theta)[p_2 u(w^H_{Alt}) + (1-p_2)u(w^L_{Alt}) - c] - \overline{u}\right\} = 0 \tag{2.3}$$

$$\mu\left\{(1+\theta)(p_2-p_1)[u(w^H_{Alt}) - u(w^L_{Alt})] - c\right\} = 0 \tag{2.4}$$

$$(1+\theta)[p_2 u(w^H_{Alt}) + (1-p_2)u(w^L_{Alt}) - c] - \overline{u} \geq 0 \tag{2.5}$$

$$(1+\theta)(p_2-p_1)[u(w^H_{Alt}) - u(w^L_{Alt})] - c \geq 0 \tag{2.6}$$

$$\lambda \geq 0 \tag{2.7}$$

$$\mu \geq 0 \tag{2.8}$$

Note that $\lambda = 0$ cannot be a solution since it violates equation (2.2), because $u' > 0$ and $1 > p_2 > p_1 > p_0 > 0$ by assumption. Moreover, $\mu > 0$; otherwise, equations (2.1) and (2.2) would imply

$$u'(w^H_{Alt}) = \frac{1}{(1+\theta)\lambda} = u'(w^L_{Alt}),$$

which yields $w^L_{Alt} = w^H_{Alt}$ for all $\theta \in [0,1]$ since $u'' < 0$, thus violating the incentive compatibility constraint in (2.2). Therefore, any solution must have $\lambda, \mu > 0$ such that

$$\begin{cases} \lambda p_2 + \mu(p_2 - p_1) > 0 \\ \lambda(1-p_2) - \mu(p_2 - p_1) > 0 \end{cases}$$

Since the Lagrange multipliers are strictly positive, the optimal contract is fully characterized by the binding IC and IR, which rearranged result in

$$\begin{aligned} u(w^H_{Alt}) &= \frac{\overline{u}}{1+\theta} + c\frac{[(1-p_1) + \theta(p_2 - p_1)]}{(1+\theta)(p_2 - p_1)}, \\ u(w^L_{Alt}) &= \frac{\overline{u}}{1+\theta} - c\frac{[p_1 - \theta(p_2 - p_1)]}{(1+\theta)(p_2 - p_1)}. \end{aligned}$$

Differentiation of the incentive compatibility constraint with respect to θ leads to

$$\frac{\partial(w^H_{Alt} - w^L_{Alt})}{\partial\theta} < 0.$$

Given the optimal contract, one must again wonder whether the principal will induce both agents to exert high effort or not. If not, then the principal can offer the constant wage $w = u^{-1}\left(\frac{\overline{u}}{1+\theta}\right)$ as before, since this satisfies the participation constraint, but not the incentive compatibility constraint. Thus, the principal's expected payoff in this case is again given by

$$V_{00}(\theta) = p_0 x^H + (1-p_0)x^L - 2u^{-1}\left(\frac{\overline{u}}{1+\theta}\right),$$

while inducing high effort yields expected profits

$$V_{11}(\theta) = p_2\left[x^H - 2w^H_{Alt}\right] + (1-p_2)\left[x^L - 2w^L_{Alt}\right].$$

Consequently, it is only beneficial to the principal demanding high effort from both agents if $V_{11}(\theta) \geq V_{00}(\theta)$, that is,

$$\begin{aligned} (p_2 - p_0)(x^H - x^L) + 2u^{-1}\left(\frac{\overline{u}}{1+\theta}\right) \geq 2\Bigg[p_2 u^{-1}\left(\frac{\overline{u}}{1+\theta} + \frac{c[(1-p_1)+\theta(p_2-p_1)]}{(1+\theta)(p_2-p_1)}\right) \\ +(1-p_2)u^{-1}\left(\frac{\overline{u}}{1+\theta} - c\frac{[p_1 - \theta(p_2-p_1)]}{(1+\theta)(p_2-p_1)}\right)\Bigg]. \end{aligned}$$

Take $c = 0$. The the right-hand side reduces to $2[p_2 u^{-1}\left(\frac{\overline{u}}{1+\theta}\right) + (1-p_2)u^{-1}\left(\frac{\overline{u}}{1+\theta}\right)] = 2u^{-1}\left(\frac{\overline{u}}{1+\theta}\right)$, and the inequality is automatically satisfied, since $1 > p_2 > p_1 > p_0 > 0$ and $x^H > x^L$ by assumption.

Therefore, by continuity, $\exists \overline{c}_{Alt} > 0$ such that $\forall c \in (0, \overline{c}_{Alt}]$, $V_{11}(\theta) \geq V_{00}(\theta)$.

C.4 Proof of Corollary 1

Take $\theta_0, \theta_1 \in [0,1]$ such that $\theta_0 < \theta_1$, and let $\mathbf{w}^*_{Alt}(\theta) = \left(w^H_{Alt}(\theta), w^L_{Alt}(\theta)\right)$ denote the optimal wage offered by the principal when agents display the degree of morality $\theta \in [0,1]$. Moreover, for any $\theta \in [0,1]$, let $\mathcal{C}(\theta)$ denote the set of contracts satisfying both the IR and IC constraints for the degree of altruism θ, so that $\mathbf{w}^*_{Alt}(\theta) \in \mathcal{C}(\theta)$.

Then, using the constraints, one can check that

$$u(w^H_{Alt}(\theta_0)) - u(w^L_{Alt}(\theta_0)) = \frac{c}{(1+\theta_0)(p_2-p_1)} > \frac{c}{(1+\theta_1)(p_2-p_1)}$$

and

$$\overline{u} = (1+\theta_0)[p_2 u(w^H_{Alt}(\theta_0)) + (1-p_2)u(w^L_{Alt}(\theta_0)) - c] < (1+\theta_1)[p_2 u(w^H_{Alt}(\theta_0)) + (1-p_2)u(w^L_{Alt}(\theta_0)) - c],$$

so that $\mathbf{w}^*_{Alt}(\theta_0) \in \mathcal{C}(\theta_1)$. However, the KKT conditions imply that $\mathbf{w}^*_{Alt}(\theta_1)$ is the unique solution to the principal's problem for $\theta = \theta_1$. Then, it must be the case that $V^{Alt}_{11}(\mathbf{w}^*_{Alt}(\theta_1); \theta_1) > V^{Alt}_{11}(\mathbf{w}^*_{Alt}(\theta_0); \theta_1)$.

Now, observe that

$$\frac{dw^H_{Alt}(\theta)}{d\theta} = -\frac{1}{u'(w^H_{Alt}(\theta_0))}\left[\frac{(1-p_2)c}{(1+\theta)^2(p_2-p_1)} + \frac{\overline{u}}{(1+\theta)^2}\right] < 0.$$

Thus, keeping $w^L_{Alt}(\theta_1) = w^L_{Alt}(\theta_0)$ and taking $w^H_{Alt}(\theta_1) = w^H_{Alt}(\theta_0) - \varepsilon, \varepsilon \approx 0$, the principal satisfies both constraints while increasing his payoff by $2p_2\varepsilon > 0$. Therefore, the principal's expected profit is strictly increasing in θ.

C.5 Proof of Proposition 4

The principal's problem is given by

$$\begin{aligned}\mathbb{L} = \; & p_2\left[x_H - w_i^H - w_j^H\right] + (1-p_2)\left[x_L - w_i^L - w_j^L\right] \\ & + \textstyle\sum_{i=1}^{2} \lambda_i \left[p_2 u(w_i^H) + (1-p_2)u(w_i^L) - c - \overline{u}\right] \\ & + \textstyle\sum_{i=1}^{2} \mu_i \left[\left(u(w_i^H) - u(w_i^L)\right)\left((p_2 - p_1) + \kappa_i(p_1 - p_0)\right) - c\right]\end{aligned}$$

for $i = A, B$ and $j \neq i$. Then, the KKT conditions are given by the system of equations

$$-p_2 + \lambda_i p_2 u'(w_i^H) + \mu_i u'(w_i^H)\left((p_2 - p_1) + \kappa_i(p_1 - p_0)\right) = 0 \tag{3.1}$$

$$-(1-p_2) + \lambda_i(1-p_2)u'(w_i^L) - \mu_i u'(w_i^L)\left((p_2 - p_1) + \kappa_i(p_1 - p_0)\right) = 0 \tag{3.2}$$

$$p_2 u(w_i^H) + (1-p_2)u(w_i^L) - c \geq \overline{u} \tag{3.3}$$

$$\left(u(w_i^H) - u(w_i^L)\right)\left((p_2 - p_1) + \kappa_i(p_1 - p_0)\right) - c \geq 0 \tag{3.4}$$

$$\lambda_i \left[p_2 u(w_i^H) + (1-p_2)u(w_i^L) - c - \overline{u}\right] = 0 \tag{3.5}$$

$$\mu_i \left[\left(u(w_i^H) - u(w_i^L)\right)\left((p_2 - p_1) + \kappa_i(p_1 - p_0)\right) - c\right] = 0 \tag{3.6}$$

$$\lambda_i \geq 0 \tag{3.7}$$

$$\mu_i \geq 0 \tag{3.8}$$

Equations (3.1) and (3.2) clearly show that $\lambda_i = \mu_i = 0$ is not a possibility. Indeed, if that was the case, then $p_2 = 0$, which contradicts our initial assumption. Also, I cannot have $\mu_i > 0 = \lambda_i$, because this would imply equation (3.2) is not satisfied. So, I must either have $\lambda_i > 0 = \mu_i$ or $\lambda_i, \mu_i > 0$. Solving for the multipliers in equations (3.1) and (3.2) yields

$$\begin{aligned}\lambda_i &= \frac{(1-p_2)u'(w_i^H) + p_2 u'(u_i^L)}{u'(w_i^H)u'(w_i^L)} > 0 \\ \mu_i &= \frac{p_2(1-p_2)(u'(w_i^L) - u'(w_i^H))}{u'(w_i^H)u'(w_i^L)((p_2 - p_1) + \kappa_i(p_1 - p_0))} > 0\end{aligned}$$

so both the (IC_i) and (IR_i) constraints bind. Thus, using equations (3.3) and (3.4) one finds the optimal schedule must satisfy

$$\begin{aligned}u(w_i^L) &= \overline{u} - \frac{c[(1-\kappa_i)p_1 + \kappa_i p_0]}{(p_2 - p_1) + \kappa_i(p_1 - p_0)} \\ u(w_i^H) &= \overline{u} + \frac{c[(1-p_1) + \kappa_i(p_1 - p_0)]}{(p_2 - p_1) + \kappa_i(p_1 - p_0)}.\end{aligned}$$

Differentiating this expressions with respect to κ_i yields

$$\frac{dw_i^H}{d\kappa_i} = -\frac{(1-p_2)(p_1-p_0)}{(p_2-p_1)+\kappa_i(p_1-p_0)}\frac{c}{u'(w_i^H)} < 0$$
$$\frac{dw_i^L}{d\kappa_i} = \frac{p_2(p_1-p_0)}{(p_2-p_1)+\kappa_i(p_1-p_0)}\frac{c}{u'(w_i^L)} > 0.$$

Given the optimal contract, one must again wonder whether the principal will induce both agents to exert high effort or not. If not, then the principal can offer the constant wage $w = u^{-1}(\overline{u})$ as before, since this satisfies the participation constraint, but not the incentive compatibility constraint. Thus, the principal's expected payoff in this case is again given by

$$V_{00}(\kappa) = p_0 x^H + (1-p_0)x^L - 2u^{-1}(\overline{u}),$$

while inducing high effort yields expected profits

$$V_{11}(\kappa) = p_2\left[x^H - \sum_{i=A,B} u^{-1}\left(\overline{u} + \frac{c[(1-p_1)+\kappa_i(p_1-p_0)]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right]$$
$$+ (1-p_2)\left[x^L - \sum_{i=A,B} u^{-1}\left(\overline{u} - \frac{c[(1-\kappa_i)p_1+\kappa_i p_0]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right].$$

Consequently, it is only beneficial to the principal demanding high effort from both agents if $V_{11}(\kappa) \geq V_{00}(\kappa)$, that is,

$$(p_2-p_0)(x^H-x^L) + 2u^{-1}(\overline{u}) \geq \sum_{i=A,B}\left[p_2 u^{-1}\left(\overline{u} + \frac{c[(1-p_1)+\kappa_i(p_1-p_0)]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right.$$
$$\left. +(1-p_2)u^{-1}\left(\overline{u} - \frac{c[(1-\kappa_i)p_1+\kappa_i p_0]}{(p_2-p_1)+\kappa_i(p_1-p_0)}\right)\right].$$

Take $c=0$. The the right-hand side reduces to $\sum_{i=A,B}[p_2 u^{-1}(\overline{u}) + (1-p_2)u^{-1}(\overline{u})] = 2u^{-1}(\overline{u})$, and the inequality is automatically satisfied, since $1 > p_2 > p_1 > p_0 > 0$ and $x^H > x^L$ by assumption.

Therefore, by continuity, $\exists \overline{c}_{HM} > 0$ such that $\forall c \in (0, \overline{c}_{HM}]$, $V_{11}(\kappa) \geq V_{00}(\kappa)$.

C.6 Proof of Corollary 2

Denote again the principal's indirect expected profits by $V_{11}(\kappa_A, \kappa_B)$. Then,

$$\begin{aligned}
\frac{\partial V_{11}(\kappa_A, \kappa_B)}{\partial \kappa_i} &= -p_2 \frac{dw_i^H}{d\kappa_i} - (1-p_2)\frac{dw_i^L}{d\kappa_i} \\
&= \frac{c(p_1 - p_0)}{[(p_2 - p_1) + \kappa_i(p_1 - p_0)]^2}\left[\frac{p_2(1-p_2)}{u'(w_i^H)} - \frac{(1-p_2)p_2}{u'(w_i^L)}\right] \\
&= \underbrace{\frac{c(p_1 - p_0)(1-p_2)p_2}{[(p_2 - p_1) + \kappa_i(p_1 - p_0)]^2}}_{>0} \times \frac{u'(w_i^L) - u'(w_i^H)}{u'(w_i^L)u'(w_i^H)} > 0
\end{aligned}$$

since $w_i^H > w_i^L$ by the monotonicity implied by the (IC_i) and $u'' < 0$. Therefore, the principal's expected payoff is strictly increasing in each degree of morality κ_i.

In a similar fashion, let $U(\kappa_i)$ denote the agent's indirect utility under the optimal contract. Then,

$$\begin{aligned}
\frac{\partial U(\kappa_i)}{\partial \kappa_i} &= p_2 u'(w_i^H)\frac{dw_i^H}{d\kappa_i} + (1-p_2)u'(w_i^L)\frac{dw_i^L}{d\kappa_i} \\
&= \frac{c(p_1 - p_0)}{[(p_2 - p_1) + \kappa_i(p_1 - p_0)]^2}\left[-p_2(1-p_2) + (1-p_2)p_2\right] = 0.
\end{aligned}$$

C.7 Proof of Lemma 1

Let $\kappa = \alpha = \theta \in [0,1]$, $c > 0$ and $1 > p_2 > p_1 > p_0 > 0$ by assumption. Then

$$\begin{aligned}
&\frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)} \geq \frac{c}{(1+\theta)(p_2 - p_1)} \Leftrightarrow \\
&(p_2 - p_1) + \theta(p_2 - p_1) \geq (p_2 - p_1) + \theta(p_1 - p_0) \Leftrightarrow \\
&p_2 - p_1 \geq p_1 - p_0.
\end{aligned}$$

C.8 Proof of Theorem 1

Let $h^H = u(w^H)$ and $h^L = u(w^L)$, which are uniquely determined for any values of w since u is strictly increasing by assumption. The principal's problem can thus be rewritten as

$$\max\nolimits_{h^L, h^H} \quad p_2\left(x^H - 2u^{-1}(h^H)\right) + (1-p_2)\left(x^L - 2u^{-1}(h^L)\right)$$

subject to

$$(h^L, h^H) \in \mathcal{C}_{HM}(\theta) = \left\{(h_1, h_2) \in \mathcal{R}^2 : p_2h_2 + (1-p_2)h_1 - c \geq \overline{u}, h_2 - h_1 \geq \frac{c}{k_2 + \theta k_1}\right\}$$

if the principal is hiring a team of moral agents, and

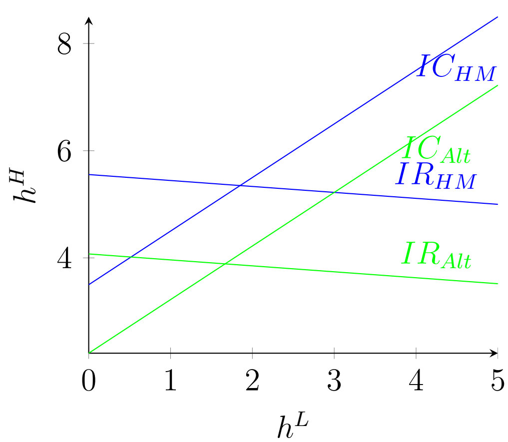
$$(h^L, h^H) \in \mathcal{C}_{Alt}(\theta) = \left\{(h_1, h_2) \in \mathcal{R}^2 : p_2h_2 + (1-p_2)h_1 - c \geq \frac{\overline{u}}{1+\theta}, h_2 - h_1 \geq \frac{c}{(1+\theta k_2}\right\},$$

where $k_1 = p_1 - p_0 > 0$ and $k_2 = p_2 - p_1 > 0$, if he considers a team of altruistic agents. The sets $\mathcal{C}_{HM}(\theta)$ and $\mathcal{C}_{Alt}(\theta)$ collect all the values of h^L and h^H satisfying the participation and incentive compatibility constraints for a given degree of morality or altruism $\theta \in [0,1]$.

First, notice that for any value of $\theta \in [0,1]$, a pair $(h_L, h_L) \in \mathcal{C}_{HM}(\theta)$ also satisfies the (IR) constraint in $\mathcal{C}_{Alt}(\theta)$: indeed, $p_2h_2 + (1-p_2)h_1 - c \geq \overline{u} \geq \frac{\overline{u}}{1+\theta}$ for $\overline{u} \geq 0$.

Suppose that $p_2 - p_1 \geq p_1 - p_0$, i.e $k_2 \geq k_1$. Then, by Lemma 1, $h^H_{HM} - h^L_{HM} \geq h^H_{Alt} - h^L_{Alt}$, which implies the optimal contract under *homo moralis* preferences also satisfies the incentive compatibility of altruistic agents, so that $(h^L_{HM}(\theta), h^H_{HM}(\theta)) \in \mathcal{C}_{Alt}(\theta)$. This implies that $\mathcal{C}_{HM}(\theta) \subset \mathcal{C}_{Alt}(\theta)$, one can conclude that $V^{Alt}_{11}(\theta) \geq V^{HM}_{11}(\theta)$. This can be graphically seen in Figure A.1.

Figure A.1: Optimal contracts for $p_2 - p_1 > p_1 - p_0$



Suppose now that $p_2 - p_1 < p_1 - p_0$, i.e. $k_2 < k_1$. In this case, the incentive compatibility constraint for moral agents is below the one for altruistic agents, as can be seen in Figure

$A.2$ and implied by Lemma 1, but because the reverse holds for the participation constraint, one cannot say that $\mathcal{C}_{Alt}(\theta) \subset \mathcal{C}_{HM}(\theta)$. Using the results in Propositions 2 and 3, one can check that

$$\begin{aligned} h^L_{HM} - h^L_{Alt} &= \overline{u} - \frac{\overline{u}}{1+\theta} + c - c + p_2 c \left(\frac{1}{(1+\theta)k_2} - \frac{1}{k_2 + \theta k_1} \right) \\ &= \frac{\theta \overline{u}}{1+\theta} + \frac{p_2 c \theta (k_1 - k_2)}{(1+\theta)k_2(k_2+\theta k_1)} \geq 0 \end{aligned}$$

for all $\theta \in [0,1]$, $\overline{u} \geq 0$, $c > 0$ and $0 < p_0 < p_1 < p_2 < 1$. Thus, the wage paid after a bad realization of output for a moral agent is larger than the corresponding wage paid to an altruistic agent if $k_1 > k_2$. Therefore, the principal can only be better off hiring a team of moral agents if the wage paid after a good realization of output to the latter is sufficiently smaller than the one paid for altruistic agents and the isoprofit curve is sufficiently flat. The former holds only if

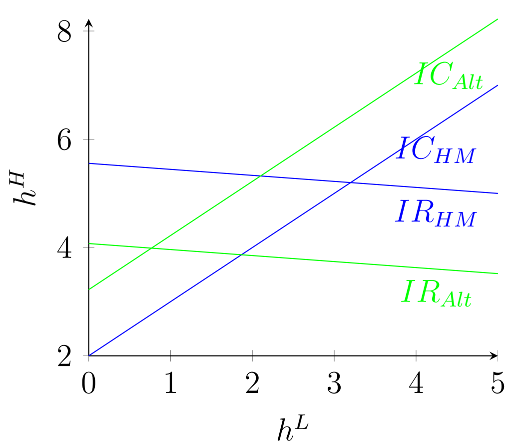
$$\begin{aligned} h^H_{HM} - h^H_{Alt} &= h^L_{HM} - h^L_{Alt} + c \left(\frac{1}{k_2 + \theta k_1} - \frac{1}{(1+\theta)k_2} \right) \\ &= \frac{\theta \overline{u}}{1+\theta} - \frac{(1-p_2) c \theta (k_1 - k_2)}{(1+\theta)k_2(k_2+\theta k_1)} < 0. \end{aligned}$$

For $\theta \in (0,1]$, $c > 0$ and $0 < p_0 < p_1 < p_2 < 1$, the inequality above holds iff

$$\overline{u}(k_2 + \theta k_1) < \frac{1-p_2}{p_2 - p_1}(k_1 - k_2),$$

that is, for $\overline{u} = 0$ or small values of θ for $\overline{u} > 0$.

Figure A.2: Optimal contracts for $p_2 - p_1 < p_1 - p_0$



Now, remember that the slope of the isoprofit curve for the principal is given by $\frac{dh^H}{dh^L} = -\frac{1-p_2}{p_2}\frac{u'(w^H)}{u'(w^L)} < 0$, which becomes flatter as p_2 approaches 1. Thus, if $k_1 > k_2$, the principal is better off with a team of moral agents if p_2 is close to 1 and either $\overline{u} = 0$ or $\overline{u} > 0$ and $\theta \to 0$.

C.9 Proof of Proposition 5

Existence of the contract follows from the KKT conditions written on the main text. The same goes for the inequalities on the wage schedules.

References

Alchian, A. A. and H. Demsetz (1972, December). Production , information costs, and economic organization. *American Economic Review 62*(5), 777–95.

Alger, I. and J. W. Weibull (2013, November). Homo moralis - preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2302.

Alger, I. and J. W. Weibull (2016). Evolution and kantian morality. *Games and Economic Behavior 98*(C), 56–67.

Bandiera, O., I. Barankay, and I. Rasul (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics 120*(3), 917–962.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *Review of Economic Studies 77*(2), 417–458.

Barr, A. and P. Serneels (2009, March). Reciprocity in the workplace. *Experimental Economics 12*(1), 99–112.

Becker, G. S. (1974, Nov.-Dec.). A theory of social interactions. *Journal of Political Economy 82*(6), 1063–93.

Bergström, T. C. (1995, March). On the evolution of altruistic ethical rules for siblings. *American Economic Review 85*(1), 58–81.

Bowles, S. and S. Polania-Reyes (2012, June). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature 50*(2), 368–425.

Brown, V. R. and E. D. Vaughn (2011). The writing on the (facebook) wall: The use of social networking sites in hiring decisions. *Journal of Business and Psychology 26*(2), 219.

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

Dufwenberg, M., P. Heidhues, G. Kirchsteiger, F. Riedel, and J. Sobel (2011). Other-regarding preferences in general equilibrium. *Review of Economic Studies 78*(2), 613–639.

Englmaier, F. and A. Wambach (2010, July). Optimal incentive contracts under inequity aversion. *Games and Economic Behavior 69*(2), 312–328.

Fehr, E. and K. M. Schmidt (2006). *The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories*, Volume 1 of *Handbook on the Economics of Giving, Reciprocity and Altruism*, Chapter 8, pp. 615–691. Elsevier.

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

Itoh, H. (2004). Moral hazard and other-regarding preferences. *The Japanese Economic Review 55*(1), 18–45.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986, October). Fairness and the assumptions of economics. *The Journal of Business 59*(4), S285–300.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991, Winter). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives 5*(1), 193–206.

Kolm, S. and J. M. Ythier (Eds.) (2006a). *Handbook of the Economics of Giving, Altruism and Reciprocity* (1 ed.), Volume 1. Elsevier.

Kolm, S. and J. M. Ythier (Eds.) (2006b). *Handbook of the Economics of Giving, Altruism and Reciprocity* (1 ed.), Volume 2. Elsevier.

Livio, L. (2015). Friend or foes? optimal incentives for reciprocal agents. *Mimeo*.

Mookherjee, D. (1984). Optimal incentive schemes with many agents. *Review of Economic Studies 51*(3), 433–446.

O'Leonard, K., R. Erickson, and J. Krider (2015, April). Talent acquisition factbook 2015: Benchmarks and trends in spending, staffing, and key recruiting metrics.

Rey-Biel, P. (2008, 06). Inequity aversion and team incentives. *Scandinavian Journal of Economics 110*(2), 297–320.

Rotemberg, J. J. (2006). *Altruism, Reciprocity and Cooperation in the Workplace*, Volume 1 of *Handbook on the Economics of Giving, Reciprocity and Altruism*, Chapter 21, pp. 1371–1407. Elsevier.

Thaler, R. H. (1988, Winter). Anomalies: The winner's curse. *Journal of Economic Perspectives 2*(1), 191–202.

Tversky, A. and R. H. Thaler (1990, Spring). Anomalies: Preference reversals. *Journal of Economic Perspectives 4*(2), 201–11.

Screening Teams of Moral and Altruistic Agents

Roberto Sarkisian*

May 5, 2020

Abstract

This paper studies the problem of screening teams of either moral or altruistic agents, in a setting where agents choose whether or not to exert effort in order to achieve a high output for the principal. I show that there exists no separating equilibrium menu of contracts that induces the agents to reveal their types unless the principal either (i) excludes one group from the productive relationship, or (ii) demands different efforts from different preference groups. I also characterize the contract inducing pooling equilibria in which all agents are incentivised to exert the high level of effort.

Keywords: Moral hazard in teams, screening, *homo moralis* preferences, altruism.

JEL Classification: D82, D86, D03.

1 Introduction

Sarkisian (2017) explores a moral hazard in teams problem where an employer has to choose between hiring a team of altruistic agents or a team of moral agents (as in Alger and Weibull (2013, 2016, 2017)). The key finding is that the principal sometimes prefers the team of

*This work is prepared under the supervision of professors Ingela Alger and François Salanié, as part of my Ph.D. thesis at Toulouse School of Economics. I thank Renee Bowen, Leandro de Magalhaes, Roger Myerson and Jörgen Weibull for very valuable feedback. Finally, I also thank seminar audiences at Toulouse School of Economics, as well as conference participants at the 2018 Africa Meeting of the Econometric Society.

moral agents over the team of altruistic ones depending on the production technology and the common degree of morality or altruism. The author then argues that firms may have incentives to collect information about their prospective employees preferences in order to benefit from offering less costly contracts.

This last point, however, is not developed there. In particular, Sarkisian (2017) assumes that the agents preferences are common knowledge, i.e. the principal knows not only which kind of prosocial preferences the prospective employees have, but also what is the common degree of morality or altruism displayed by the agents.

The objective of this paper is to relax that strong assumption: in what follows, it is assumed that the degree of altruism or morality is known to all parties, but the utility function specification is private knowledge of the agents. The principal then seeks to distinguish the two groups by offering menus of contracts that induce participation, effort provision and revelation of private information by the employees.

This class of adverse selection followed by moral hazard problems has been analyzed before. To cite but a few, Jullien et al. (2007) considers the problem of screening risk-averse agents under moral hazard under the strong assumption that the utility function satisfies single-crossing and CARA properties.[1] As a result, they find that the power of incentives is decreasing with respect to risk-aversion. Ollier and Thomas (2013) study a two-output model with risk-neutral agent protected by limited liability and ex-post participation constraints, and find that a fully pooling contract is optimal. Maréchal and Thomas (2018) build upon the previous model by assuming that the agent is risk-averse, and also finds that pooling contracts are difficult to avoid.

All the papers cited above differ from the environment studied here in one important way: they assume that preferences are common knowledge, but that either the degree of risk-aversion or a productivity parameter is private information of the single agent. Here, as stated before, the utility function rather than the common degree of altruism or morality

[1]Laffont and Martimort (2002) and Bolton and Dewatripont (2005) have dedicated sections to this class of problems, and provide more references to the literature.

is private information of the agents.[2] The main results, however, are in line with Ollier and Thomas (2013) and Maréchal and Thomas (2018): separation is difficult to achieve by the principal if she desires the agents to exert effort in equilibrium. Intuitively, this is a consequence of the utility functions not displaying a single-crossing-like property, an assumption that is imposed in Jullien et al. (2007).

Screening prosocial preferences have been the central issue in some studies, both theoretically and empirically. von Siemens (2011) studies an environment with a single principal screening a continuum of workers that have private information about their ability and preferences over social comparisons. In particular, von Siemens (2011) contrasts the optimal employment contracts for selfish and inequity averse agents, and finds that it is impossible to screen workers of similar ability with respect to their social preferences within the firm, a result that is line with the ones found here. The main differences between von Siemens (2011) and the model in this study is that the former considers only the adverse selection problem faced by the principal when hiring a single agent, while the latter assumes teamwork and moral hazard. Closer in essence to this paper are the works of Cabrales and Charness (2011) and Demougin et al. (2006), who consider screening followed by moral hazard when agents' prosocial preferences are characterized by inequity aversion. Their results also suggest that screening agents according to their social preferences is not feasible.

The paper goes as follows. The next session presents the environment and the concept of separating equilibrium to be considered. Section 3 discusses screening and existence of separating equilibria, while Section 4 characterizes contracts that support pooling equilibria. Section 5 concludes. For ease of exposition, all proofs are relegated to an appendix.

[2]If the symmetry assumption used in Sarkisian (2017) and here was to be relaxed, the problem would differ in a second way, namely the moral-hazard in teams structure rather than the conventional single agent formulation.

2 The Model

Consider a single risk-neutral principal (she/firm) who faces a continuum of potential employees with total mass normalized to one.[3] The firm seeks to hire a pair of agents to work on a common task that yields output $x \in \{x^H, x^L\}$ to the principal, with $x^H > x^L$. The probability p of the high outcome being achieved depends on the binary choices of effort made by the agents employed in the firm, $e_i = 0, 1$ for $i \in \{A, B\}$. In particular,

$$Pr(x = x^H | e_A, e_B) = p_{e_A + e_B}, \tag{1}$$

where I assume that $1 > p_2 \geq p_1 \geq p_0 > 0$. The cost of exerting effort is identical to every agent, $C(e) = ce$, for $c > 0$.

Output is contractible upon, and the principal posts wage schedules $w_i(x)$ in order to attract the teams of agents. If the firm successfully attracts a pair of employees, her realized profit is

$$V(x, w_A, w_B) = x - w_A(x) - w_B(x). \tag{2}$$

Denote by $\pi_i(e_i, e_j, w_i(x))$ the expected *material* payoff accruing to agent i from the effort choices (e_i, e_j) and wage schedule $w_i(x)$, for $i, j \in \{A, B\}, i \neq j$. I restrict attention to wage schedules pairs $\mathbf{w}_i = (w_i^H, w_i^L)$ determining the payments following good and bad realizations of revenues.[4] In what follows, the *material* payoff function takes the expected additively separable form

$$\pi(e_i, e_j, \mathbf{w}_i) = p_{e_i+e_j} \left[u(w_i^H) - c(e_i) \right] + (1 - p_{e_i+e_j}) \left[u(w_i^L) - c(e_i) \right], \tag{3}$$

where $u : \mathbb{R}_+ \to \mathbb{R}$ is the function that associates the agent's consumption utility to each wage realization w. The agents are risk averse towards wages: $u(w)$ is assumed to be twice-continuously differentiable, strictly increasing and strictly concave.

Each pair of agents belongs to one class of preference group: altruists or moral. More precisely, each team is composed by two agents drawn from the same preference group, as

[3]The model can be restated by considering n pairs of potential employees, without loss of generality.

[4]This is in line with Maréchal and Thomas (2018) and Ollier and Thomas (2013), where the schedules are composed of a fixed plus a variable part.

in Sarkisian (2017). The principal only knows the proportion of the population that each group corresponds to: $\lambda \in (0,1)$ for altruist, and $1-\lambda$ for moral. The agents' preferences in each group are represented by the utility functions

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \alpha_i) = \pi(e_i, e_j, \mathbf{w}_i) + \alpha_i \pi(e_j, e_i, \mathbf{w}_j) \tag{4}$$

for the altruists and

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \kappa_i) = (1-\kappa_i)\pi(e_i, e_j, \mathbf{w}_i) + \kappa_i \pi(e_i, e_i, \mathbf{w}_i) \tag{5}$$

for the moral agents, where $\alpha \in [0,1]$ and $\kappa \in [0,1]$ represent the agents' degrees of altruism and morality, respectively. In what follows, as discussed in Sarkisian (2017), I assume that $\alpha_A = \alpha_B = \kappa_A = \kappa_B = \theta$, and focus on the comparable functions[5]

$$U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) = \pi(e_i, e_j, \mathbf{w}_i) + \theta \pi(e_j, e_i, \mathbf{w}_j), \tag{6}$$

$$U^{HM}(e_i, e_j, \mathbf{w}_i, \theta) = (1-\theta)\pi(e_i, e_j, \mathbf{w}_i) + \theta \pi(e_i, e_i, \mathbf{w}_i). \tag{7}$$

The timing of the game is depicted in Figure 1.

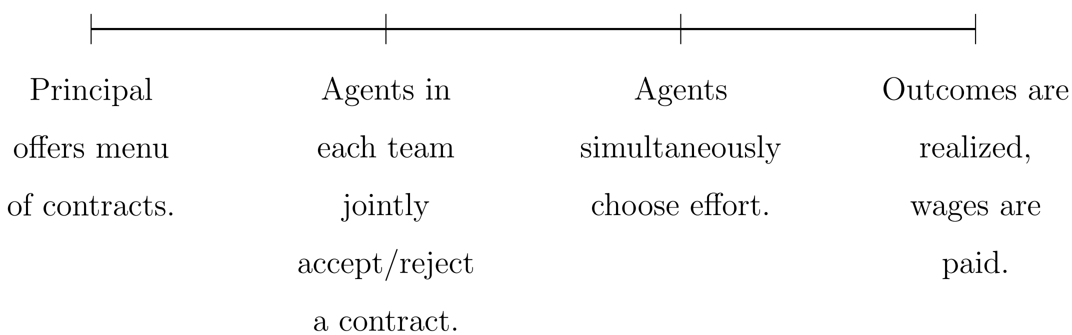


Figure 1: Timing of the game.

[5]As pointed in Alger and Weibull (2013) and also explored in Bergström (1995), this is the formulation that gives rise to the behavioral equivalence between *homo moralis* and altruistic preferences. Under an appropriate change of variables, the altruistic utility function could be rewritten as $U^{Alt}(e_i, e_j, \mathbf{w}_i, \mathbf{w}_j, \theta) = (1-\widetilde{\theta})\pi(e_i, e_j, \mathbf{w}_i) + \widetilde{\theta}\pi(e_j, e_i, \mathbf{w}_j)$, for $\widetilde{\theta} \in [0, 1/2]$.

3 Screening

Due to the assumption of a common degree of morality or altruism, I restrict attention to symmetric contracts offered to each team. These assumptions simplify the problem in the sense that both the incentive compatibility and individual rationality constraints are similar to the ones studied in the literature with a single agent, save for their dependence on the common degree of morality/altruism. For the pure moral hazard problem, these constraint are

$$(1+\theta)\left[p_2 u(w^H) + (1-p_2)u(w^L) - c\right] \geq \overline{u}_{Alt}, \tag{8}$$

$$u(w^H) - u(w^L) \geq \frac{c}{(1+\theta)(p_2 - p_1)} \tag{9}$$

for altruistic agents, and

$$p_2 u(w^H) + (1-p_2)u(w^L) - c \geq \overline{u}_{HM}, \tag{10}$$

$$u(w^H) - u(w^L) \geq \frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)} \tag{11}$$

for moral agents.

In contrast to Sarkisian (2017), I allow the different groups to have different reservation utilities. Two particular cases deserve a special mention. First, as in Sarkisian (2017), agents in each group may have exactly the same reservation utility $\overline{u}_{Alt} = \overline{u}_{HM} = \overline{u}$, which generates different utility levels for the participating agents whenever $\overline{u} > 0$ due to the utility function representing each prosocial preference. The second particular case is $\overline{u}_{Alt} = (1+\theta)\overline{u}_{HM}$, so that the participation constraints for both moral and altruistic agents are identical for any common degree of prosociality $\theta \in [0,1]$[6].

Let $\mathcal{C}_{Alt}$ denote the set of contracts that satisfy the participation and incentive compatibility constraints of altruistic agents, and similarly define the set $\mathcal{C}_{HM}$ for moral agents. The principal's screening problem is to choose $\mathbf{w}_{Alt} \in \mathcal{C}_{Alt}$ and $\mathbf{w}_{HM} \in \mathcal{C}_{HM}$ such that neither

[6] If $\theta = 0$, then both moral and altruistic agents behave as purely selfish individuals, and the screening problem becomes irrelevant.

group has an incentive to pick the contract designed for the other group[7]. The issue, however, is that the intersection between these two sets of feasible contracts is not empty, and thus one can always construct a separating equilibrium by selecting two contracts, $\mathbf{w}$ and $\mathbf{w}'$, in $\mathcal{C}_{Alt} \bigcap \mathcal{C}_{HM}$, and arguing that each group will self-select into one, and only one, of these contracts.

I will, therefore, focus on a stronger form of separation: I will require that a menu of contracts has at most one element in the intersection of the feasible sets. This will ensure that at least one group has no incentives to deviate and accept the contract designed for the other group.

Let $h = u(w)$, which is uniquely defined for each $w \in \mathbb{R}$ since u is strictly increasing by assumption. I can therefore rewrite the sets of feasible contracts using the linear constraints

$$(1+\theta)[p_2 h^H + (1-p_2)h^L - c] \geq \overline{u}_{Alt}, \tag{12}$$

$$h^H - h^L \geq \frac{c}{(1+\theta)(p_2 - p_1)} \tag{13}$$

for altruistic agents and

$$p_2 h^H + (1-p_2)h^L - c \geq \overline{u}_{HM}, \tag{14}$$

$$h^H - h^L \geq \frac{c}{(p_2 - p_1) + \theta(p_1 - p_0)} \tag{15}$$

for moral agents. I can easily draw the sets of feasible contracts for the cases in which the production technology displays decreasing or increasing returns to efforts, by appropriately choosing the reservation utilities $\overline{u}_{HM}$ and $\overline{u}_{Alt}$. In Figures 2 and 3, I assume that $\overline{u}_{HM} = \overline{u}_{Alt} > 0$. Notice, in Figure 2, that $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$, which implies that any feasible contract offered to moral agents is also accepted by altruistic ones at the same time that it also provides the latter with incentives to exert high effort. Meanwhile, in Figure 3, a contract in $\mathcal{C}_{HM}$ is also accepted by altruistic agents, but it may not necessarily induce them to exert the high effort.

[7] This is akin to the incentive compatibility constraint in the adverse selection problem. It can be seen as an additional set of constraints in the principal's maximization program.

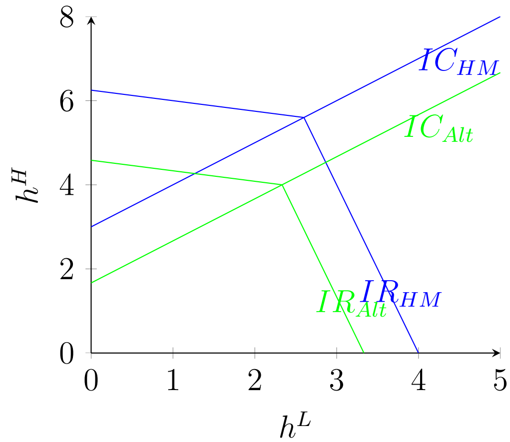


Figure 2: Feasible sets of contracts for $p_2 - p_1 > p_1 - p_0$.

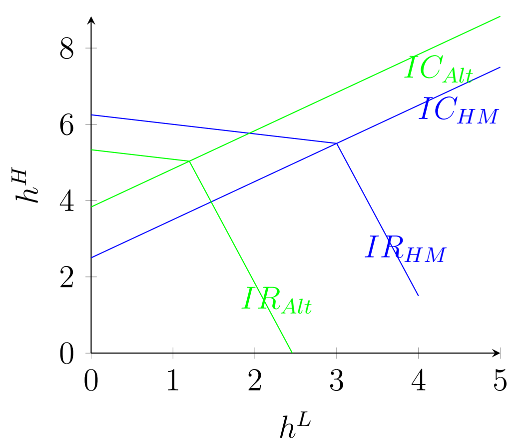


Figure 3: Feasible sets of contracts for $p_2 - p_1 < p_1 - p_0$.

Lemma 1: *Suppose that* $\overline{u}_{Alt} < (1+\theta)\overline{u}_{HM}$*. There exists no strictly separating menu of contracts that induces both agents to exert high effort. Similarly, if* $\overline{u}_{Alt} > (1+\theta)\overline{u}_{HM}$*, no strictly separating equilibrium exists that induces all agents to exert effort.*

Lemma 1 states that a separating equilibrium does not exist if the reservation utility of both groups are such that the participation constraints are never identical and contracts incentivize agents to exert the high effort. Then, for each case, one can find a profitable deviation for a group, i.e. either the moral agents are better off taking the contract designed for the altruistic teams, or altruists like the moral contracts better than their own.

Lemma 2: *Suppose that* $\overline{u}_{Alt} = (1+\theta)\overline{u}_{HM} \geq 0$*. There exists no strictly separating menu of contracts that induces both agents to exert high effort.*

The negative results in Lemmas 1 and 2 can be linked to the fact that the isoutility curves of moral and altruistic agents never cross in the region where both groups are incentivized to exert the high effort. Indeed, under the assumptions on each statement, the indifference curves of each kind of prosocial agent is either identical to one another, or they are parallel. It is this violation of a single-crossing-like property that prevents the principal from finding schedules that elicit the agents' preferences.

Proposition 1: *There exists no strictly separating menu of contracts that induces both types of agents to accept a contract and exert high effort for any* $\overline{u}_{Alt}, \overline{u}_{HM} \in \mathbb{R}_+$*.*

3.1 Separating Equilibria with Low Effort

A separating equilibrium also doesn't exist if the principal requires both types of agents to exert the low effort. Indeed, due to risk aversion by the agents, the principal can induce participation by offering the constant schedules $\overline{w}_{Alt}$ and $\overline{w}_{HM}$ for the altruistic and moral groups, respectively, satisfying the individual rationality constraints

$$u\left(\overline{w}_{Alt}\right) \geq \frac{\overline{u}_{Alt}}{1+\theta}, \tag{16}$$

$$u\left(\overline{w}_{HM}\right) \geq \overline{u}_{HM}. \tag{17}$$

By standard arguments, these constraints must bind in an equilibrium. However, if $\overline{u}_{HM} \neq \frac{\overline{u}_{Alt}}{1+\theta}$, one preference group always has incentives to deviate and accept the contract design to the second group. On the other hand, if $\overline{u}_{HM} = \frac{\overline{u}_{Alt}}{1+\theta}$, then $\overline{w}_{Alt} = \overline{w}_{HM}$ since u is strictly increasing, which implies that all workers accept exactly the same contract, and thus picking them apart is impossible for the principal. This argument is collected in the following result.

Proposition 2: *No separating equilibrium exists if the principal wishes to induce both preference groups to accept the contract and exert the low effort.*

3.2 Screening Preference Groups Through Exclusion

Propositions 1 and 2 have shown that the principal cannot screen moral agents from altruistic ones when she must induce both participation and high effort. However, the principal might be able to screen the different preference groups by offering a single (non-null) contract.

Turn once more to Figure 2, by assuming identical reservation utilities and increasing returns to effort. If the principal offers a menu with a single contract that satisfies both the participation and incentive compatibility constraint of the altruistic agent with equality (the intersection of the green lines), she will ensure that: (i) altruistic agents accept the offer and exert high effort; and (ii) moral agents choose not to participate in the relationship with the principal. The same can be achieved under decreasing returns to effort by offering a similar contract (Figure 3).

More generally, the principal can screen the preference groups by offering a singleton menu, where the contract offered necessarily satisfies with equality the participation constraint of the preference group with the lowest reservation utility.

Proposition 3: *Suppose that $\overline{u}_{HM} \neq \frac{\overline{u}_{Alt}}{1+\theta}$. The principal can screen different preference groups by offering a single contract that excludes the agents with the highest reservation utility.*

Proposition 3 holds either when the principal wishes to induce high or low effort. For the latter case, the argument behind Proposition 3 is even more compelling since the principal will offer a constant wage schedule to the risk-averse agents to exert zero effort, and therefore she can simply choose to employ the cheapest of the preference groups in terms of reservation utilities.

3.3 Screening with Different Efforts

So far, my analysis has focused on the case where both groups of agents are required by the principal to exert the same level of effort, either high or low. The negative results are basically a consequence of the indifference curves for the two groups being parallel to one another when efforts are the same: this implies that the contract offered to the group with the highest outside option also attracts the other team.

Although excluding one preference group from participating in the relationship with the principal is one way to screen agents, a second one exists: namely, requiring that only one group to exert exerts high effort.

If only one group is expected to exert effort, the incentive compatibility constraint with respect to effort can be neglected for that group. Moreover, a constant schedule should be offered to that same group due to the agents' risk-aversion. In what follows, I will denote by 1 the preference group that should exert effort, and by 2 the preference group who shouldn't exert effort. The feasible set of contracts for the principal will be given by all values of $\mathbf{w} = ((w_1^H, w_1^L), w_2)$ satisfying the incentive compatibility and individual rationality constraints for group 1, and the participation constraint for group 2.

One must, however, notice an important difference between the participation constraints for both groups. For group 1, which is bound by the incentive compatibility constraint, the

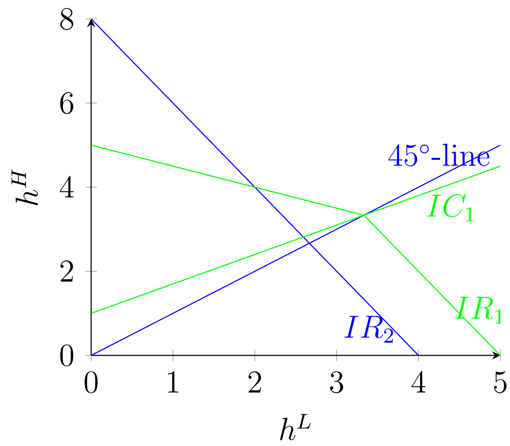


Figure 4: Feasible sets of contracts for different efforts

(IR) is given by

$$p_2 u(w_1^H) + (1 - p_2) u(w_1^L) \geq \overline{u}_1 \quad \text{for all } (w^H, w^L) \text{ that satisfies } (IC), \tag{18}$$

$$p_0 u(w_1^H) + (1 - p_2) u(w_1^L) \geq \overline{u}_1 \quad \text{otherwise.} \tag{19}$$

On the other hand, for group 2, participation must satisfy

$$u(w_2) = p_0 u(w_2^H) + (1 - p_0) u(w_2^L) \geq \overline{u}_2. \tag{20}$$

If the individual rationality curves never intersect, i.e. if either $\overline{u}_2 < \overline{u}_1$ or $\overline{u}_2 >> \overline{u}_1$, then a separating equilibrium doesn't exist, for the simple reason that the contract offered to the group with the highest outside option also attracts the agents of the other group, in much a similar manner to the case where the principal induces no group to high effort.

This is not true if the participation constraints intersect (which requires that $\overline{u}_2 \geq \overline{u}_1$). Using the linearization $h = u(w)$, the feasible set of contracts can be represented as in the figure below.

The contract offered to the agents in group 2 is given by the intersection of the 45°-line with the participation constraint for said group, since such a point has the principal proposing a constant schedule to the agents who are not expected to exert effort. On the other hand, agents in group 1 are offered the contract lying in the intersection between the two participation constraints, which they strictly prefer to the constant schedule of group 1 (while the latter is indifferent between the two contracts). The assumption that the participation constraints intersect also implies that the incentive compatibility constraint for group 1 is satisfied.

One remark is in order here: the principal leaves group 1 agents some rent for exerting the high effort, in the sense that the incentive compatibility constraint is not necessarily satisfied with equality (i.e. the pair (w_1^H, w_1^L) doesn't lie in IC_1). This can be interpreted as a *no distortion at the top* result: the principal's offer doesn't distort (downward) the effort demanded from the least costly group, but she must still pay a rent to that group.

Proposition 4: *Suppose that $\overline{u}_{HM}$ and $\overline{u}_{Alt}$ are such that $\overline{u}_{HM} \neq \frac{\overline{u}_{Alt}}{1+\theta}$ and the individual rationality constraints from both groups cross each other once. Then, a separating equilibrium exists if the principal induces only one preference group to exert the high effort.*

4 Pooling Equilibria

Let $\mathbf{w}_{HM}$ be the contract that satisfies both conditions in $\mathcal{C}_{HM}$ with equality, and similarly define $\mathbf{w}_{Alt}$. Also, denote by $\mathbf{w}_P$ the contract that satisfies both the participation constraint for moral agents and the incentive compatibility constraint for altruistic agents with equality. The following Proposition states the result formally.

Proposition 5: *Suppose that $\overline{u}_{Alt} = \overline{u}_{HM} = \overline{u} > 0$. $\mathbf{w}_{HM}$ constitutes a pooling equilibrium with both groups of agents exerting the high effort under increasing returns to efforts, while $\mathbf{w}_P$ constitutes such an equilibrium under decreasing returns to efforts.*

I do not claim in Proposition 5 that $\mathbf{w}_{HM}$ and $\mathbf{w}_P$ are the unique pooling equilibrium contracts under increasing and decreasing returns to efforts, respectively. Indeed, in the former case, any contract in $\mathcal{C}_{HM}$ indeed constitutes a pooling equilibrium. These two contracts, however, are completely characterized by a simple linear system of two equations. They also characterize one pooling equilibrium when $\overline{u}_{Alt} = (1+\theta)\overline{u}_{HM}$: in this case, the participation constraints for both groups are identical, and characterizing the feasible sets for the contracts depends only on comparisons of the incentive compatibility constraints. Moreover, they are the least costly for the principal to offer.

5 Discussion

The results presented above, in line with the literature on screening prosocial preferences, imply that the principal may be unable to construct a menu of contracts that is successful in screening teams of agents belonging to different preference groups. As a consequence, developing experiments to infer agents preferences in a static environment would present the same difficulties.

However, one possible strategy would be to offer the contracts sequentially. To fix ideas, suppose that the production technology exhibits increasing returns to efforts, and that $\overline{u}_{Alt} = \overline{u}_{HM}$. Under this circumstances, $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$ as was argued in the proof of Lemma 1. If agents are perfectly patient, than the principal could offer $\mathbf{w}_{Alt}$ in the first period, which would be accepted by all the altruistic agents but not by the moral ones, and only then offer $\mathbf{w}_{HM}$ to the remaining agents. Such sequential mechanism would make use of time to screen the agents, a channel that is not available in the static model described above.

There are two main issues with such an approach, at least from a theoretical viewpoint. First, if all the potential employees are aware that the employer would utilize the sequential offer mechanism above, altruistic agents would not accept $\mathbf{w}_{Alt}$ in the first period in order to contract under $\mathbf{w}_{HM}$ in the second period and therefore enjoy a higher utility. Clearly, such deviation by altruistic agents would again leave the principal unable to screen between the two preference groups.

Secondly, the sequential approach relies on the agents being infinitely patient and the two preference groups displaying the same reservation utility. The mechanism could still be employed in the situation where $\overline{u}_{Alt} < (1+\theta)\overline{u}_{HM}$ and $\delta_{Alt} < \delta_{HM}$, where $\delta_j \in (0,1)$ denotes group $j = \{Alt, HM\}$ discount factor. In this case, if the altruistic group discounts the future much more than its moral counterpart, the mechanism could indeed lead to full screening. Unfortunately, to the best of my knowledge, I do not know any research establishing conditions under which different prosocial preferences lead to heterogenous discount factors.

6 Conclusion

This paper extends the analysis in Sarkisian (2017) by relaxing the assumption that the agents' preferences are common knowledge in the contractual relationship. In particular, the interest lies in characterizing a separating equilibrium in which moral and altruistic individuals reveal their type and exert a high level of effort in the task proposed by the principal.

In effect, the results are negative, but in line with the literature of adverse selection followed by moral hazard: screening prosocial preferences is not possible. The empirical implication follows naturally: one cannot distinguish groups of agents characterized by the two classes of preferences described above when the degree of prosociality is the same for the two groups, at least when one considers a static environment. On the other hand, an alternative would be sequential mechanisms that offer contracts satisfying only one group's participation constraint in the first period, and only offering contracts satisfying the second group's participation constraint in the following period.

A Proofs

A.1 Proof of Lemma 1

The assumption that $\overline{u}_{Alt} < (1+\theta)\overline{u}_{HM}$ implies that the participation constraint for one group is different than the one for the other. In particular, setting $h = u^{-1}(w)$, and drawing the participation constraints on the plane (h^L, h^H) allows us to see that the participation constraint for the altruistic group is always below its counterpart for moral agents.

The proof then considers two cases in turn. Suppose first that the production technology is characterized by decreasing returns to efforts, i.e. $p_2 - p_1 < p_1 - p_0$. Then, any contract $\mathbf{w} \in \mathcal{C}_{HM}$ satisfies the participation constraint of altruistic agents with slackness, and thus it is profitable for this group of agents to deviate and take the contract designed for moral agents. Thus, no separating equilibrium exists in this case.

Under increasing returns to efforts, $p_2 - p_1 \geq p_1 - p_0$, on the other hand, one can readily check that $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$. Then, again, all altruistic agents would deviate and choose the contract designed for moral agents, since this contract would satisfy the former group's participation constraint with slackness.

For $\overline{u}_{Alt} > (1+\theta)\overline{u}_{HM}$, the proof if similar. If $p_2 - p_1 < p_1 - p_0$, then $\mathcal{C}_{Alt} \subset \mathcal{C}_{HM}$ and thus every altruistic agent has an incentive to deviate and take the contract designed for moral agents. Conversely, if $p_2 - p_1 > p_1 - p_0$, every contract in $\mathcal{C}_{Alt}$ satisfies the participation constraint of the moral agents with slackness, and therefore such agents have incentives to deviate and take the contract designed for the former group.

A.2 Proof of Lemma 2

Suppose first that $p_2 - p_1 \geq p_1 - p_0$, so that the incentive compatibility constraint of moral agents is always above the one for altruists. Then, $\mathcal{C}_{HM} \subseteq \mathcal{C}_{Alt}$ and the latter always prefer a contract designed for the former. If $p_2 - p_1 < p_1 - p_0$, the reverse holds: $\mathcal{C}_{Alt} \subseteq \mathcal{C}_{HM}$ and moral agents always prefer the contract designed for altruists rather than their own.

A.3 Alternative Proof of Proposition 1

Proposition 1 is a direct consequence of Lemma 1 and Lemma 2. Alternatively, the same result can be reached by the following reasoning. Suppose that $\mathbf{h}_{Alt} = (h^H_{Alt}, h^L_{Alt})$ and $\mathbf{h}_{HM} = (h^H_{HM}, h^L_{HM})$ are contracts that constitute a strictly separating equilibrium. Then, no team of agents have incentives to deviate and take the contract designed for the other group. In particular, this means that for altruistic agents the following inequality must hold

$$(1+\theta)[p_2 h^H_{Alt} + (1-p_2)h^L_{Alt} - c] > (1+\theta)[p_2 h^H_{HM} + (1-p_2)h^L_{HM} - c], \tag{21}$$

while

$$p_2 h^H_{HM} + (1-p_2)h^L_{HM} - c > p_2 h^H_{Alt} + (1-p_2)h^L_{Alt} - c \tag{22}$$

must hold for moral agents. Since $\theta \in [0,1]$, condition (21) reduces to

$$p_2 h^H_{Alt} + (1-p_2)h^L_{Alt} - c > p_2 h^H_{HM} + (1-p_2)h^L_{HM} - c, \tag{23}$$

which together with (22) imply that

$$p_2 h^H_{Alt} + (1-p_2)h^L_{Alt} > p_2 h^H_{HM} + (1-p_2)h^L_{HM} > p_2 h^H_{Alt} + (1-p_2)h^L_{Alt}, \tag{24}$$

a contradiction.

A.4 Proof of Proposition 3

For $\overline{u}_{HM} \neq \frac{\overline{u}_{Alt}}{1+\theta}$, the proof is analogous to that of Lemma 1. The impossibility of screening through exclusion when $\overline{u}_{HM} = \frac{\overline{u}_{Alt}}{1+\theta}$ comes from the argument of Proposition 2 if the principal does not wish to induce high effort, and generalizes straightforwardly to the case when she wishes to induce effort.

A.5 Proof of Proposition 5

Under increasing returns to efforts, $\mathcal{C}_{HM} \subset \mathcal{C}_{Alt}$, and thus $\mathbf{w}_{HM} \in \mathcal{C}_{Alt}$. In particular, as shown in Sarkisian (2017), this contract is the least costly one the principal can offer to moral agents in order to induce both participation and effort.

Under decreasing returns to efforts, $\frac{c}{(1+\theta)(p_2-p_1)} > \frac{c}{(p_2-p_1)+\theta(p_1-p_0)}$, and thus any contract satisfying the incentive compatibility constraint for altruistic agents automatically satisfies, with slackness, its counterpart for moral agents. Therefore, let us take the most restrictive set of constraints, namely the participation constraint for moral agents and the participation constraint for altruistic agents. Any contract satisfying both constraints, in particular $\mathbf{w}_P$, will necessarily belong to both $\mathcal{C}_{HM}$ and $\mathcal{C}_{Alt}$, so moral and altruistic agents alike accept such contract and exert the high level of effort.

References

Alger, I. and J. W. Weibull (2013, November). Homo moralis - preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2302.

Alger, I. and J. W. Weibull (2016). Evolution and kantian morality. *Games and Economic Behavior 98*(C), 56–67.

Alger, I. and J. W. Weibull (2017, Sep). Strategic behavior of moralists and altruists. *Games 8*(3), 38.

Bergström, T. C. (1995, March). On the evolution of altruistic ethical rules for siblings. *American Economic Review 85*(1), 58–81.

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

Cabrales, A. and G. Charness (2011). Optimal contracts with team production and hidden information: An experiment. *Journal of Economic Behavior and Organization 77*(2), 163 – 176.

Demougin, D., C. Fluet, and C. Helm (2006). Output and wages with inequality averse agents. *The Canadian Journal of Economics / Revue canadienne d'Economique 39*(2), 399–413.

Jullien, B., B. Salanié, and F. Salanié (2007, January). Screening risk-averse agents under moral hazard: single-crossing and the CARA case. *Economic Theory 30*(1), 151–169.

Laffont, J.-J. and D. Martimort (Eds.) (2002). *The Theory of Incentives: The Principal-Agent Model* (1 ed.), Volume 1. Princeton University Press.

Maréchal, F. and L. Thomas (2018). The optimal contract under adverse selection in a moral-hazard model with a risk-averse agent. *Games 9*(1), 12.

Ollier, S. and L. Thomas (2013). Ex post participation constraint in a principalagent model with adverse selection and moral hazard. *Journal of Economic Theory 148*(6), 2383–2403.

Sarkisian, R. (2017). Team incentives under moral and altruistic preferences: Which team to choose? *Games 8*(3), 37.

von Siemens, F. (2011). Heterogeneous social preferences, screening, and employment contracts. *Oxford Economic Papers 63*(3), 499–522.

Optimal Incentives Schemes under *Homo Moralis* Preferences*

Roberto Sarkisian

May 5, 2020

Abstract

I study optimal incentive schemes in a multi-agent moral hazard model, where each agent has other-regarding preferences and an individual measure of output, with both being observable by the principal. In particular, the two agents display *homo moralis* preferences as in Alger and Weibull (2013, 2016). I find that, contrary to the case with purely selfish preferences, tournaments can never be optimal when agents are risk averse and as the degree of morality increases, positive payments are made in a larger number of output realizations. Furthermore, I extend the analysis to a dynamic setting, in which a contract is initially offered to the agents, who then repeatedly choose which level of effort to provide in each period. As in Che and Yoo (2001), I show that the optimal incentive schemes in this case are similar to the ones obtained in the static setting, but for the role of intertemporal discounting.

Keywords: Moral hazard in teams, optimal contracts, *homo moralis* preferences.

JEL Classification: D82, D86, D03.

*This work is prepared under the supervision of professors Ingela Alger and François Salanié, as part of my Ph.D. thesis at Toulouse School of Economics.

1 Introduction

While most of the traditional economic literature on moral hazard has focused on agents' heterogeneous skills (Laffont and Tirole, 1986; Bolton and Dewatripont, 2005) and task allocation (Crawford and Knoer, 1981; Besley and Ghatak, 2005), it is crucial to also take into account social preferences in the context of incentive provision.[1] As pointed out by Nagin et al. (2002), a considerable fraction of the agents participating in their workplace experiment do not behave as selfishly as standard theory would predict. Fehr and Schmidt (2000) and Fehr et al. (2007) show that fairness concerns may drastically impact contractual designs in principal-agent environments. Dohmen et al. (2009) surveys experimental evidence of reciprocity both in stylized labor markets as well as in other decision settings. Bowles and Polania-Reyes (2012) survey finds evidence that explicit economic incentives can either reinforce or weaken prosocial behavior, and that the latter is more common, due to explicit incentives adversely affecting the individual's other-regarding preferences.

Here, I study the optimal incentives schemes a principal can offer to a team of two agents characterized by a novel class of other-regarding preferences, namely *homo moralis* preferences (Alger and Weibull, 2013, 2016). Using a multi-agent moral hazard environment, as first proposed in Alchian and Demsetz (1972) and Holmström (1982), I show that the optimal contracts offered to the teams of agents have to balance three different aspects: the agents' prosocial behavior, here characterized by their degree of morality, risk aversion and incentive provision.[2] I also consider the possibility of repeated interactions between the agents, as in Che and Yoo (2001), and show that the optimal incentive scheme in the dynamic setting largely maintains the structure of its static counterpart but for the effects of discounting in the wages paid by the principal.

[1] Besley and Ghatak (2005) explores the notion of a mission-oriented production of collective goods, emphasizing the role of matching between the mission preferences of principals and agents, since the former *economizes on the need for high-powered incentives.*

[2] The next section explores in more depth the concept and the utility function representing moral preferences.

More closely related to this paper are the theoretical contributions identifying the effects of other-regarding preferences in contract design and incentives provision. Many of those study inequity aversion, following the seminal work of Itoh (2004). While Demougin and Fluet (2003) considers inequity-averse agents in tournaments, Rey-Biel (2008) and Englmaier and Wambach (2010) look for the optimal incentive schemes under such preferences. While the former focus on binary effort choices by the agent (as in Itoh (2004)), the latter allows not only for continuous effort choice, but also consider incomplete contracts. In general, the results in this literature show that team incentives may outperform both individual and relative performance schemes when agents sufficiently dislike inequity.

In a similar vein to Itoh (2004) as well, Livio (2015) derives optimal incentive schemes for reciprocal agents, a class of preferences first modeled in normal form games by Rabin (1993). As a result, Livio (2015) finds that the optimal incentive scheme depends on the interplay between risk aversion and the degree of reciprocity. More precisely, a relative performance scheme, which induces negative reciprocity, is optimal when agents are not very risk averse, while a joint performance scheme inducing positive reciprocity is better when agents become more risk averse. A different form of reciprocity between agents is altruism.[3] Dur and Sol (2010) and Dur and Tichem (2015) study conditions under which explicit incentives can improve or damage altruism between co-workers.[4] In contrast to inequity aversion, and closer to the results in reciprocity, they find that both team performance and relative performance schemes can reinforce altruism in the workplace.

Differently than the literature above, I find that in most cases relative performance is the optimal scheme for incentivising moral agents. In one particular case, team performance is also optimal, but it is so because all other schemes are not available since limited liability constraints rule them out. Moreover, I also show that tournaments are never optimal, in

[3]Bénabou and Tirole (2006) study a model where agents have heterogeneous degrees of altruism (and greed). Their construction differs from Becker (1974) notion of altruism because on the latter it is the agents' concern about each other's wellbeing rather than their concern about own social reputation that induces prosocial behavior.

[4]See Kolm and Ythier (2006a,b) for more on altruism.

stark contrast to the studies of optimal incentive schemes with purely selfish individuals.

The choice of *homo moralis* preferences comes from the realization that, in all the literature listed above, other-regarding preferences are *assumed* based only on psychological and experimental results. Although in most cases assuming a certain type of preferences have an intuitive appeal, as in the intra-household models based on forms of altruism, a theoretical foundation for the choice of one or other preference representation was lacking. The missing link, then, is a specification of preferences that is robust in a general setting, or one that evolves endogenously over time in a population. Alger and Weibull (2013, 2016) provide such a link. They show that under incomplete information (agents' preferences are privately observed) and assortative matching, *homo moralis* preferences emerge as the evolutionarily stable ones, and that the degree of morality is given by the degree of assortativity of the matching process in which the individuals participate. Also, Alger and Weibull (2013, 2016) argue that the utility function representing *homo moralis* preferences is the only one that proves to be robust against invasion in monomorphic populations in the class of continuous utility functions. As described in their paper, these preferences can be understood as a convex combination of the well-known selfish *homo oeconomicus* preferences and Laffont (1975)'s concept of Kantian morality.

The paper continues in the following way. Section 2 introduces the model and the *homo moralis* utility function. Section 3 then analyses the problem faced by the principal in the static setting, while Section 4 extends the results to the dynamic environment. Section 5 concludes. For ease of exposition, all proofs are collected in the Appendix.

2 The Model

Consider a firm composed by one manager (principal) and two employees (agents), denoted by $i \in \{A, B\}$. Each agent produces an observable output $x_i \in \{x^H, x^L\}$, with $x^H > x^L$, which is stochastically determined by the agent's choice of either exerting effort or shirking, i.e. $e_i \in \{0, 1\}$. This production technology is characterized by the probability of achieving a high output conditional on the effort supplied:

$$Prob\left(x_i = x^H | e_i = 1\right) = p \in (0,1), \tag{1}$$

$$Prob\left(x_i = x^H | e_i = 0\right) = q \in (0,p). \tag{2}$$

This formulation assumes that the observable outputs x_A and x_B depend only on the corresponding agent's choice of effort and are independently drawn, and the production technology is symmetric. The cost of exerting effort is given by

$$C(e_i) = ce_i, \quad c > 0, i \in \{A, B\}.$$

The principal is assumed to be risk-neutral, and can use a remuneration scheme $\mathbf{w} = (\mathbf{w}_A, \mathbf{w}_B)$ to compensate her employees, which possibly depends on the output realizations x_A and x_B. Thus, the principal's expected payoff can be written as

$$V(x_A, x_B, \mathbf{w}) = \sum_i \mathbb{E}\left[x_i - w_i\right].$$

Each agent's material payoff is assumed to be additively separable in wages and effort, i.e.

$$\pi_i(w_i, e_i) = u_i(w_i) - C(e_i).$$

For ease of exposition, I assume that employees A and B value wages identically: $u_A(w) = u_B(w) = w^{1-\rho}$, for $\rho \in [0,1)$.[5] Therefore, their material payoffs can be rewritten as

$$\pi(w_i, e_i) = w_i^{1-\rho} - ce_i. \tag{3}$$

For any pair of effort choices (e_A, e_B), the space of possible output realizations is $\mathcal{S} = \{(x^H, x^H), (x^H, x^L), (x^L, x^H), (x^L, x^L)\}$, where each element $s \in \mathcal{S}$ is an ordered pair $s = (x_A, x_B)$. The principal can offer compensation schemes determining wages after each possible realization of output, namely

$$\mathbf{w_i} = \left(w_{iHH}, w_{iHL}, w_{iLH}, w_{iLL}\right),$$

[5]This specification allows to examine the behavior under risk neutrality ($\rho = 0$) as a limiting case of risk-averse agents ($\rho \in (0,1)$).

where w_{iHH} specifies, for instance, the wage received by agent i when both output realizations are high and w_{iHL} denotes the same agent's wage when his realized output is high while his partner's output realization is low. The agents' expected material payoff, conditional on efforts, is

$$\begin{aligned}\mathbb{E}\left[\pi(w_i, e_i)|e_i, e_j\right] = P(e_i)P(e_j)w_{iHH}^{1-\rho} + P(e_i)\left[1 - P(e_j)\right] w_{iHL}^{1-\rho} \\ + \left[1 - P(e_i)\right] P(e_j)w_{iLH}^{1-\rho} + \left[1 - P(e_i)\right]\left[1 - P(e_j)\right] w_{iLL}^{1-\rho} - ce_i,\end{aligned}$$

for $i, j \in \{A, B\}$, $j \neq i$.

Up to this moment the preferences of the employees haven't been fully described. In particular, I assume that the agents have *homo moralis* preferences[6], represented by the (expected) utility function

$$U_i(\mathbf{w}_i, e_i, e_{-i}; \kappa_i) = (1 - \kappa_i)\mathbb{E}[\pi(w_i, e_i)|e_i, e_{-i}] + \kappa_i\mathbb{E}[\pi(w_i, e_i)|e_i, e_i], \tag{4}$$

where $\kappa_i \in [0, 1]$ denotes agent i's *degree of morality*. Inspection of the above expression shows that this specification is the convex combination between the usual representation of selfish preferences (the first term) and agent i's material payoff if agent j were to choose the same action (second term). Also, the limiting cases are interesting: while taking $\kappa_i = 0$ reduces the utility function to the standard selfish preferences, $\kappa_i = 1$ captures a situation where agent i doesn't behave strategically: indeed, the problem in that case reduces to a single decision where $j \neq i$ choice of effort has not effect on agent i's utility.

Throughout the exposition, I assume that the difference $x^H - x^L > 0$ is large enough for the principal to always prefer to induce both agents not to shirk. Also, in order to focus on incentives provision, I assume that the workers are already employed by the firm, that contracts are bound by limited liability constraints and that preferences and costs are common information. Thus, the only private information is the agents' choices of effort. Timing is as follows: the principal sets her preferred incentive schemes (possibly contingent on both performance indicators (x_A, x_B)). The agents then simultaneously choose whether

[6] See (Alger and Weibull, 2013, 2016).

or not to exert effort. Finally, (x_A, x_B) is realized and payments are made according to the incentives schemes proposed by the employer.

Some remarks must be made. First, given any incentive scheme, agents A and B play a static game with complete information. Not only do they know the proposed incentive scheme, they also know their partner's degree of morality, and thus his preferences. Also, since this is a one-shot game, it is irrelevant whether the agents can observe each other's choice of effort after the outputs are realized or not, and thus discussions about commitment are outside the scope of this model. Second, assuming the agents are already employed by the firm somewhat relaxes the problem that will be solved by the principal, since participation constraints will not be considered.[7]

3 The Principal's Problem in the Static Framework

The principal's problem is

$$\begin{array}{lll} \max_{\mathbf{w}} & V(x_A, x_B, \mathbf{w}) & \\ s.t. & U_i(\mathbf{w}_i, 1, 1; \kappa_i) \geq U_i(\mathbf{w}_i, 0, 1; \kappa_i) & (IC_i) \\ & w_{iHH}, w_{iHL}, w_{iLH}, w_{iLL} \geq 0 & (LL_i) \end{array}$$

for $i \in \{A, B\}$. Given the risk neutrality and the linearity of the expectation operator, and assuming both agents will exert effort, the principal's expected profits can be rewritten as

$$V(x_A, x_B, \mathbf{w}) = \mathbb{E}[x_A + x_B] - \left[p^2 \sum_i w_{iHH} + p(1-p) \sum_i (w_{iHL} + w_{iLH}) + (1-p)^2 \sum_i w_{iLL} \right].$$

Since the principal maximizes over the incentives schemes, the problem above is equivalent to

$$\begin{array}{lll} \min_{\mathbf{w}} & p^2 \sum_i w_{iHH} + p(1-p) \sum_i (w_{iHL} + w_{iLH}) + (1-p)^2 \sum_i w_{iLL} & \\ s.t. & U_i(\mathbf{w}_i, 1, 1; \kappa_i) \geq U_i(\mathbf{w}_i, 0, 1; \kappa_i) & (IC_i) \\ & w_{iHH}, w_{iHL}, w_{iLH}, w_{iLL} \geq 0 & (LL_i) \end{array}$$

[7] I will consider, however, limited liability on wages. If the outside option on the participation constraint would be set to zero, then limited liability would imply the former.

Focus now on the incentive compatibility constraint. On the left-hand side both agents are exerting effort, so that $\mathbb{E}[\pi(\mathbf{w}_i, e_i^*)|e_i^*, e_j^*] = \mathbb{E}[\pi(\mathbf{w}_i, e_i^*)|e_i^*, e_i^*]$. Therefore, one obtains

$$\begin{aligned}
U_i(\mathbf{w}_i, 1, 1; \kappa_i) &= \mathbb{E}[\pi(\mathbf{w}_i, 1)|e_i^* = 1, e_j^* = 1] \\
&= p^2 w_{iHH}^{1-\rho} + p(1-p)w_{iHL}^{1-\rho} + (1-p)pw_{iLH}^{1-\rho} + (1-p)^2 w_{iLL}^{1-\rho} - c,
\end{aligned}$$

while the right-hand side writes

$$\begin{aligned}
U_i(\mathbf{w}_i, 0, 1; \kappa_i) = (1-\kappa_i)&\left[qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho}\right] \\
&+ \kappa_i\left[q^2 w_{iHH}^{1-\rho} + q(1-q)w_{iHL}^{1-\rho} + (1-q)qw_{iLH}^{1-\rho} + (1-q)^2 w_{iLL}^{1-\rho}\right].
\end{aligned}$$

Plugging in the above equations into the incentive compatibility constraint and rearranging the terms around the wages yields

$$\begin{aligned}
&w_{iHH}^{1-\rho}\left[p^2 - (1-\kappa_i)qp - \kappa_i q^2\right] \\
&\quad + w_{iHL}^{1-\rho}\left[p(1-p) - (1-\kappa_i)q(1-p) - \kappa_i q(1-q)\right] \\
&\quad + w_{iLH}^{1-\rho}\left[(1-p)p - (1-\kappa_i)(1-q)p - \kappa_i(1-q)q\right] \\
&\quad + w_{iLL}^{1-\rho}\left[(1-p)^2 - (1-\kappa_i)(1-q)(1-p) - \kappa_i(1-q)^2\right] \geq c.
\end{aligned}$$

This form of writing the (IC_i) is very convenient to observe how the degree of morality affects the incentives of agent i to exert effort. To start, take the term multiplying w_{iHH}, and suppose $\kappa_i = 0$. In this case, one obtains $p \cdot p - q \cdot p = (p-q) \cdot p$, which exactly describes the decrease in the probability of achieving the output realization (x^H, x^H) that would be observed under selfish preferences: agent i would take the action $e_{-i} = 1$ as given, and would only consider the effects caused by his own shirking. On the other hand, for $\kappa_i = 1$, the term would become $p \cdot p - q \cdot q = (p-q) \cdot (p+q) > (p-q) \cdot p$: everything else fixed, the principal would need a smaller wage w_{iHH} to incentivise agent i, since now agent i would evaluate his payoff as if both him and his partner were shirking. Similar reasoning can be applied to the remaining terms.[8]

[8]One interesting remark is in order at this point. Under standard *homo oeconomicus* preferences, both

For ease of exposition, the analysis will be divided in two: first, the risk-neutral case ($\rho = 0$) will be tackled. Then I proceed to characterize the optimal incentive schemes when the agents are risk averse ($\rho \in (0, 1)$).

3.1 Optimal Incentive Schemes for Risk-Neutral Agents ($\rho = 0$)

For now, focus is channeled towards risk-neutral agents ($\rho = 0$). Under this additional assumption, the principal's problem is a linear programming problem with five inequality constraints: the incentive compatibility and the four limited liability constraints. The first result states that the principal's problem accepts three widely known solution candidates, namely an *individual incentive scheme*, where the principal remunerates each agent i according to his observable measure of output x_i alone; a *team incentive scheme*, in which the basis for remuneration is the sum of the individual observable measures; and a *tournament scheme*, such that agent i receives a bonus if his output measurement has the highest value.

Lemma 1. *When agents are risk neutral with respect to wealth and have homo moralis preferences, the following two solution candidates implement* $e_i = 1$, $\forall \kappa_i \in [0, 1]$, $i \in \{A, B\}$*:*

1. *an individual incentive scheme, with*

$$w_{iHH} = w_{iHL} = \frac{c}{p - q} > w_{iLH} = w_{iLL} = 0;$$

agents are characterized by the same degree of morality $\kappa_i = 0$, and thus each multiplicative term is identical for employees A and B. However, if $\kappa_A \neq \kappa_B$, these terms may not be the same any longer, and the workers would behave as if they possess heterogeneous beliefs about the realizations of output. This would, therefore, give a rationale for different wages being proposed (and accepted in the case where participation constraints are included in the model) by agents facing the same disutility of effort and attitude towards risk. See de la Rosa (2011) for moral hazard problems with heterogenous beliefs. Observe, however, the two approaches are radically different at heart: while de la Rosa (2011) assumes agents have heterogeneous beliefs about the probability of success, thus implying that at least one of them have incorrect beliefs, in my model I assume both agents have correct beliefs about the probability of success, but differ only on their degree of morality.

2. *a team incentive scheme, such that*

$$w_{iHH} = \frac{c}{(p-q)(p+\kappa_i q)} > w_{iHL} = w_{iLH} = w_{ILL} = 0.$$

For $\kappa_i < \frac{1-p}{q}$, a tournament scheme also implements $e_i = 1$:

$$w_{iHL} = \frac{c}{(p-q)(1-p-\kappa_i q)} > w_{iHH} = w_{iLH} = w_{iLL} = 0.$$

Proof: *all proofs are in the Appendix.*

Inspection of the remuneration structures reveals two interesting insights. First, under the individual incentive schemes, the wage paid following a high realization of the observable measure of output does not depend on the agents' degrees of morality, in contrast with the remaining schemes. Intuitively, this is a consequence of the independence assumptions on the production technology and its stochastic measurement: together with an incentive scheme that relies solely on individual performance, this environment reduces to zero the effect of Kantian morality in the incentives provision; it is as if the employees are purely selfish.

Second, the tournament is only feasible if agent i does not exhibit a high degree of morality. The mechanism behind this is the asymmetric nature of this particular incentive scheme: an employee can only receive the bonus if he outperforms his colleague, thus conflicting the agent's urge *to do the right thing.* However, if $p + q \leq 1$, a tournament is feasible for all $\kappa_i \in [0, 1]$. In this case, since the probability of realizing a high output measure is sufficiently small, the incentives provided by the asymmetric scheme may overpower the agents' morality in order to induce both to exert effort.

In order to determine which scheme among the ones mentioned above is the most profitable for the principal, one must simply compare the expected payments made under each alternative structure.

Lemma 2. *When agents are risk neutral with respect to wealth and have homo moralis preferences, the principal is indifferent among the alternative schemes if $\kappa_i = 0$. If $\kappa_i \in (0, 1]$,* *the principal strictly prefers the team incentive scheme over the individual and tournament structures.*

The statement considers two distinct cases: one for $\kappa = 0$ and another for $\kappa > 0$. In the first case, the analysis boils down to standard *homo oeconomicus* preferences with risk-neutral agents. Thus, since the agents' are identical and risk-sharing is not an issue, all three structures provide exactly the same expected payments to the employees and, therefore, have the same expected cost for the principal. One concludes that the principal is indifferent among the alternative compensation schemes.

The interesting case, however, lies on $\kappa > 0$. When the employees display a concern with *doing the right thing*, the principal is strictly better off implementing a team incentive scheme. Such a scheme implies that the desired outcome is a high output realization for agents 1 and 2, which transforms exerting a high effort into being the right thing. Since both agents now display a positive degree of morality, the total expected cost of explicitly incentivising the agents is reduced.

Although Lemma 2 rules out individual performance and tournaments as the optimal incentive schemes (for $\kappa_i > 0$), it does not fully characterize the solution to the principal's problem. This is done in Proposition 1 below.

Proposition 1. *When agents are risk neutral with respect to wealth and have* homo moralis *preferences, the optimal incentive scheme for the principal is team performance.*

Proposition 1 strengthens Lemma 2: team incentives are the best scheme a principal can use to incentivise a team of moral and risk-neutral agents, among all schemes that satisfy the incentive compatibility and limited liability constraints.

The proof of Proposition 1 is constructed in four steps. First, I show that any optimal incentive scheme always has $w_{iLL} = 0$ for $i \in \{A, B\}$. Then, it is easy to show that the incentive compatibility constraint must be satisfied with equality. The third step uses Lemma 2, thus eliminating any incentive scheme such that $w_{iHL} > 0$. Then, the fourth and last step must only consider schemes with $w_{iHH}, w_{iLH} \geq 0$; finally, I show that the principal's expected transfers to the agents are minimized with a team incentive scheme for any $\kappa_i \in [0, 1]$.

Closer inspection of the optimal incentive scheme shows that the principal is better off with teams of highly moral agents. The mechanism behind this is that a larger degree of morality slackens the incentive compatibility constraint, thus demanding a smaller transfer from the employer to the employees. This is stated formally below.

Corollary 1. *Under the optimal incentive scheme with risk-neutral agents (team performance), the principal's expected profit is strictly increasing in the agents' degrees of morality.*

3.2 Optimal Incentive Schemes for Risk-Averse Agents ($\rho \in (0, 1)$)

Studying the risk-neutral case allows an understanding of the effects *homo moralis* preferences have on designing the optimal incentive scheme, without having to take into consideration the trade-off between incentive provision and risk sharing. In particular, the agents' urge *to do the right thing* makes team performance scheme the most profitable for the principal in that case. In this section, the risk neutrality assumption is relaxed, and the optimal incentive scheme will have to balance morality, incentive provision and risk aversion.

The assumption on a functional form for the utility function over wealth, namely $u(w) = w^{1-\rho}$ for $\rho \in [0, 1)$, comes in handy in this section since the results under risk neutrality can be treated as a particular case of this more general framework. Thus, at least for sufficiently high degrees of morality and low risk aversion, one expects team performance to be the optimal incentive scheme. The analysis below aims to specify the conditions for that claim to hold.

First, it is noteworthy that the usual incentive schemes (team, individual performance and tournaments) can be used by the principal to elicit effort. However, one other scheme must also be considered here: *relative performance.* In such a scheme, payments to agent i are made whenever his output realization is high, but it differs from an individual incentive scheme in allowing different wages following good or bad realizations of output from agent j. Under risk neutrality both schemes are identical because of the linearity of the utility function. However, under risk aversion, the concavity of u allows the principal to induce high effort by offering such a compensation scheme, since now any scheme must balance the trade-off between incentive provision and risk sharing.

Lemma 3. *When agents are risk averse with respect to wealth and have homo moralis preferences, the following incentive schemes implement $e_i = 1$ for $i \in \{A, B\}$:*

1. *an individual incentive scheme, for any $\kappa_i \in [0, 1]$, with*

$$w_{iHH} = w_{IHL} = \left(\frac{c}{p-q}\right)^{\frac{1}{1-\rho}} > w_{iLH} = w_{iLL} = 0;$$

2. *a team incentive scheme, for any $\kappa_i \in [0, 1]$, such that*

$$w_{iHH} = \left(\frac{c}{(p-q)(p+\kappa_i q)}\right)^{\frac{1}{1-\rho}} > w_{iHL} = w_{iLH} = w_{ILL} = 0;$$

3. *a tournament scheme, for $\kappa_i < \frac{1-p}{q}$, in which*

$$w_{iHL} = \left(\frac{c}{(p-q)(1-p-\kappa_i q)}\right)^{\frac{1}{1-\rho}} > w_{iHH} = w_{iLH} = w_{iLL} = 0;$$

4. *a relative performance scheme, for $\kappa_i < \frac{1-p}{q}$*

$$w_{iHH} = \left(\frac{c}{(p-q)(p+\kappa_i q) + A(\kappa_i, \rho)^{1-\rho}(p-q)(1-p-\kappa_i q)}\right)^{\frac{1}{1-\rho}} \geq$$
$$w_{iHL} = w_{iHH} \cdot A(\kappa_i, \rho) > w_{iLH} = w_{iLL} = 0$$

where $A(\kappa_i, \rho) = \left(\frac{p(1-p-\kappa_i q)}{(1-p)(p+\kappa_i q)}\right)^{\frac{1}{\rho}} \in [0, 1]$.

For the first three schemes, taking $\rho = 0$ yields exactly the same expressions shown in Lemma 1, which characterized such schemes for risk-neutral agents. Now, taking the limit as $\rho \to 0$ on the relative performance scheme yields the same expression as in the team performance: lacking the need for risk sharing, both schemes are identical.

Before characterizing the optimal incentive scheme for the principal, the following intermediate results deserves a few remarks.

Lemma 4. *For any $\rho \in (0,1)$ and $\kappa_i \in [0,1]$, $i = 1,2$, the principal prefers an individual incentive scheme over a tournament.*

The intuition for Lemma 4 is very simple: since a tournament imposes more risk on the agent than an individual incentive scheme, it must remunerate the agent for the increase in the riskiness of the contract. However, this compensation for risk is not profitable for the principal, for any degree of morality of the agent. Moreover, if the degree of morality is sufficiently high, such a scheme does not even satisfy the incentive compatibility constraint.

In contrast to the risk-neutral case, the optimality of a team performance scheme no longer holds for all values of κ_i, p and q. In particular, when compared to the individual performance scheme, the principal will only prefer the former if the agents' degrees of morality are very high, or if their coefficient of risk aversion is sufficiently low.

Lemma 5. *The principal strictly prefers team performance over individual performance schemes iff $\kappa_i > \overline{\kappa}(\rho) = \frac{p(1-p^\rho)}{qp^\rho}$.*

Again, observe this result extends the findings under risk neutrality: for $\rho = 0$, the right-hand side of the necessary and sufficient condition becomes 0, and thus any positive degree of morality will imply the optimality of team incentives over individual performance as was seen before. However, the right-hand side is strictly increasing[9] in ρ, which implies only a very high degree of morality can offset an increase in the degree of risk aversion in order for the principal to profit from the team incentive scheme. As $\rho \to 1$, the condition becomes $\kappa_i > \frac{1-p}{q}$, which can never be satisfied if $p + q \leq 1$. Counterintuitively, as the

[9]Indeed, $\frac{\partial\left(\frac{p(1-p^\rho)}{qp^\rho}\right)}{\partial\rho} = -\frac{p^{1-\rho}}{q}\ln p > 0$ since $p \in (0,1)$.

employee becomes more risk averse, the principal can benefit from a high degree of morality by offering the agent a contract associating positive payments to a larger number of possible output realizations. On the other hand, if the agent is not very risk averse but has a very high degree of morality, remunerating solely on the case where both agents are successful in obtaining the high output is optimal given the beliefs held by the moral agent.

Lemmas 4 and 5 rank the principal's preferences over team, individual and tournament schemes, but refrain from comparing them to relative performance schemes. Proposition 2 below strengthens the comparison, by determining the optimal incentive scheme for the principal depending on the probabilities of attaining the high output, the agent's risk aversion and degree of morality.

Proposition 2. *Suppose agents are risk averse with respect to wealth and have* homo moralis *preferences. Then, for any $\rho \in (0,1)$:*

1. *If $p+q>1$ and*
   - $\kappa \in \left[0, \frac{1-p}{q}\right]$*: a relative performance scheme, with $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$, is optimal;*
   - $\kappa \in \left[\frac{1-p}{q}, 1\right]$*: a team performance scheme, with $w_{iHH} \geq 0$ and $w_{iHL} = w_{iLH} = w_{iLL} = 0$, is optimal.*
2. *If $p+q \leq 1$ and*
   - $\kappa \in \left[0, \frac{p}{1-q}\right]$*: a relative performance scheme, with $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$, is optimal;*
   - $\kappa \in \left[\frac{p}{1-q}, 1\right]$*: a performance scheme with $w_{iHH}, w_{iHL}, w_{iLH} \geq 0$ and $w_{iLL} = 0$ is optimal.*

The interplay between risk aversion and morality leads to the optimality of relative performance schemes in most cases: it is profitable for the principal to offer compensation schemes that induce positive payments in as many output realizations as possible. One case, however, does the exact opposite by proposing an incentive scheme where the only positive payment comes only if both agents are successful in their tasks: if $p + q > 1$ and $\kappa_i \geq \frac{1-p}{q}$, the principal can profit by exploring the agent's high degree of morality and, thus, belief in the realization of high outcomes to concentrate transfer to that particular realization instead of promising positive transfers even when outputs are low.

Corollary 2. *Under the optimal incentive scheme with risk-averse agents, the principal's expected profit is non-decreasing in the agents' degrees of morality.*

As was the case under risk neutrality, the principal benefits from hiring agents with large degrees of morality, since they will need less explicit incentives embedded in the optimal compensation scheme in order to exert effort. However, the interplay of employees' morality and risk sharing demands compensation schemes that spread out payments more evenly across the possible realizations of output, in particular when the probability of realizing a high output is not very large (i.e. when $p + q \leq 1$).

4 Repeated Interactions

In what follows, I consider a repeated setting where the agents are expected to either exert effort ($e = 1$) or shirk ($e = 0$) in each period. As in Che and Yoo (2001), this arrangement is open-ended and can be terminated at the end of each period $t = 0, 1, \ldots$ with probability $1 - \delta \in (0, 1)$, where δ can also be thought of as the common discount factor for all three parties. A *history* at time t is a sequence of effort choices made by the employees until period $t-1$, and thus a *strategy profile* is a sequence of functions mapping from any possible history at each period into actions[10].

In this section, I will show that the optimal incentive schemes derived for the static model and stated in Proposition 2 are also capable of providing the incentives for both agents to

[10]More precisely, into a probability distribution over effort choices.

exert effort in the repeated setting. Firstly, note that a dynamic incentive compatibility constraint, for any incentive scheme $\mathbf{w}^*$, will be written as

$$U_i(\mathbf{w}^*,1,1;\kappa_i) \geq (1-\delta)U_i(\mathbf{w}^*,0,1;\kappa_i) + \delta\min\{U_i(\mathbf{w}^*,0,1;\kappa_i), U_i(\mathbf{w}^*,0,0;\kappa_i)\}. \quad (DIC_i)$$

Since $1 > p > q > 0$ by assumption, and together with the limited liability constraints, it is the case that $U_i(\mathbf{w}^*,0,1;\kappa_i) \geq U_i(\mathbf{w}^*,0,0;\kappa_i)$ under the three optimal incentive schemes in Proposition 2, so the relevant incentive constraint is

$$U_i(\mathbf{w}^*,1,1;\kappa_i) \geq (1-\delta)U_i(\mathbf{w}^*,0,1;\kappa_i) + \delta U_i(\mathbf{w}^*,0,0;\kappa_i), \quad (DIC_i')$$

which holds whenever the static incentive compatibility constraint is satisfied; indeed, for any of optimal static schemes and $\delta \in [0,1]$,

$$\begin{aligned} U_i(\mathbf{w}^*,1,1;\kappa_i) &\geq U_i(\mathbf{w}^*,0,1;\kappa_i) \\ &= (1-\delta)U_i(\mathbf{w}^*,0,1;\kappa_i) + \delta U_i(\mathbf{w}^*,0,1;\kappa_i) \\ &\geq (1-\delta)U_i(\mathbf{w}^*,0,1;\kappa_i) + \delta U_i(\mathbf{w}^*,0,0;\kappa_i). \end{aligned}$$

Moreover, one can easily check that $U_i(\mathbf{w},1,1;\kappa_i) \geq U_i(\mathbf{w},0,0;\kappa_i)$, so collusion in shirking is deterred by use of any of the three optimal incentive schemes in Proposition 2. However, the argument built until now does not imply that $e=0$ is a symmetric Nash equilibrium of the stage-game. If it is not, then the trigger-strategy here considered does not induce both agents to exert effort in the repeated game. Such issue does not arise if $U_i(\mathbf{w}^*,0,0;\kappa_i) \geq U_i(\mathbf{w}^*,1,0;\kappa_i)$, which can be written as

$$\begin{aligned} &q^2 w_{iHH}^{1-\rho} + q(1-q)w_{iHL}^{1-\rho} + (1-q)qw_{iLH}^{1-\rho} \geq \\ &\quad (1-\kappa_i)\left[pqw_{iHH}^{1-\rho} + p(1-q)w_{iHL}^{1-\rho} + (1-p)qw_{iLH}^{1-\rho}\right] \\ &\quad + \kappa_i\left[p^2 w_{iHH}^{1-\rho} + p(1-p)w_{iHL}^{1-\rho} + (1-p)pw_{iLH}^{1-\rho}\right] - c. \end{aligned}$$

Let $\overline{c}(\mathbf{w}^*,\kappa_i)$ denote the value of c that satisfies the condition above with equality for some optimal scheme $\mathbf{w}^*$ and degree of morality κ_i. I can now state the following result.

Proposition 3. *Consider an incentive scheme* $\mathbf{w}^*$ *characterized in Proposition 2. If* $c \geq \max\{\bar{c}(\mathbf{w}^*, \kappa_A), \bar{c}(\mathbf{w}^*, \kappa_B), 0\}$*, then the static optimal incentive scheme* $\mathbf{w}^*$ *induces both agents to cooperate in the repeated setting.*

An important point of Proposition 3 is that it holds for any value of the discount factor δ. That is, as long as the cost of exerting effort is sufficiently high to avoid $e = 1$ being a (weakly) dominant strategy for any of the employees, the optimal static incentive schemes of Proposition 2 also generate implicit incentives deterring shirking in the dynamic case irrespective of how patient the agents are. This is a consequence of the dynamic incentive compatibility constraint (DIC_i') being automatically satisfied by the schedules respecting its static version. Therefore, tournaments and individual performance schemes can also sustain effort in the dynamic game.

Corollary 3. *Tournaments* ($\mathbf{w}^{Tourn}$) *and individual performance scheme* ($\mathbf{w}^{Ind}$) *induce both agents to exert effort in the repeated setting if* $c \geq \max\{\bar{c}(\mathbf{w}, \kappa_A), \bar{c}(\mathbf{w}, \kappa_B), 0\}$.

Now, I want to focus on the more general principal's problem

$$\begin{aligned}
\min_{\mathbf{w}} \quad & p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) + (1-p)^2 w_{iLL} \\
s.t. \quad & U_i(\mathbf{w}^*, 1, 1; \kappa_i) \geq (1-\delta) U_i(\mathbf{w}^*, 0, 1; \kappa_i) \\
& \quad + \delta \min\{U_i(\mathbf{w}^*, 0, 1; \kappa_i), U_i(\mathbf{w}^*, 0, 0; \kappa_i)\} \qquad (DIC_i) \\
& w_{iHH}, w_{iHL}, w_{iLH} w_{iLL} \geq 0 \qquad (LL_i)
\end{aligned}$$

If $U_i(\mathbf{w}^*, 0, 1; \kappa_i) \leq U_i(\mathbf{w}^*, 0, 0; \kappa_i)$, the principal's problem is identical to the one in the static case, and the optimal incentive schemes described in Proposition 2 apply to the repeated setting. The more interesting case happens if $U_i(\mathbf{w}^*, 0, 1; \kappa_i) > U_i(\mathbf{w}^*, 0, 0; \kappa_i)$: for a large discount factor δ, the unique optimal incentive scheme will be either a team incentive scheme or a complete incentive scheme if $p + q < 1$ or $p + q > 1$, respectively. If, however, $p + q = 1$, then a relative performance scheme is uniquely optimal. The formal statement is given below.

Proposition 4. *Let $\underline{\kappa}(\delta)$ and $\overline{\kappa}(\delta)$ be such that*

$$\overline{\kappa}(0) = \frac{1-p}{q}, \quad \underline{\kappa}(0) = \frac{p}{1-q},$$

and

$$\frac{\partial \overline{\kappa}(\delta)}{\partial \delta} \begin{cases} > 0 & \text{if } p+q<1 \\ = 0 & \text{if } p+q=1 \\ < 0 & \text{if } p+q>1 \end{cases}, \frac{\partial \underline{\kappa}(\delta)}{\partial \delta} \begin{cases} < 0 & \text{if } p+q<1 \\ = 0 & \text{if } p+q=1 \\ > 0 & \text{if } p+q>1 \end{cases}.$$

Then, for any $\rho \in (0,1)$ and $\delta \in (0,1)$, the optimal incentive scheme for a risk-averse agent characterized by homo moralis *preferences is:*

1. *if $p+q>1$ and*
   - *$\kappa \in [0, \overline{\kappa}(\delta))$: a relative performance scheme, with $w_{iHH}, w_{iHL} > 0$ and $w_{iLH} = w_{iLL} = 0$;*
   - *$\kappa \in [\overline{\kappa}(\delta), 1]$: a team performance scheme, with $w_{iHH} > 0$ and $w_{iHL} = w_{iLH} = w_{iLL} = 0$.*
2. *if $p+q<1$ and*
   - *$\kappa \in [0, \underline{\kappa}(\delta)]$: a relative performance scheme, with $w_{iHH}, w_{iHL} > 0$ and $w_{iLH} = w_{iLL} = 0$;*
   - *$\kappa \in (\underline{\kappa}(\delta), 1]$: a performance scheme with $w_{iHH}, w_{iHL}, w_{iLH} > 0$ and $w_{iLL} = 0$.*
3. *if $p+q=1$, then $\underline{\kappa}(\delta) = \overline{\kappa}(\delta) = 1$ and a relative performance scheme, with $w_{iHH}, w_{iHL} > 0$ and $w_{iLH} = w_{iLL} = 0$, is optimal.*

An increase in the discount factor has two effects. The first one is the shifts in the thresholds $\underline{\kappa}(\delta)$ and $\overline{\kappa}(\delta)$. As δ approaches one, the values of the thresholds escape the interval $[0,1]$ that characterizes the degree of morality of the agents, and thus only one incentive scheme is optimal for each case.[11]

[11]In the proof of Proposition 4 in the Appendix, I show that the limits go to plus and minus infinity depending on whether $p+q>1$ or $p+q<1$.

The second effect is that an increase in the discount factor decreases the wage that must be paid to the agents, in particular if both output measures are high. This is true because the incentive schemes satisfying the dynamic incentive compatibility constraint carry implicit incentives for both agents to exert effort, by the existing threat of everlasting punishment in case of an unilateral deviation. Therefore, the principal benefits the more moral and patient his employees are, as intuition suggests.

5 Concluding Remarks

Studying optimal incentive schemes with other-regarding preferences highlights the fact that the traditional trade-off between risk sharing and incentive provision is not the only one to influence the characterization of optimal contracts, and thus, may provide a better understanding of why the contracts observed in reality are not as high-powered as the ones predicted in the theory. In this line, Englmaier and Wambach (2010), Itoh (2004) and Livio (2015), among others, explore the effects that altruism, inequity aversion and reciprocity have on compensation schemes.

Using the recent results on the evolution of preferences provided by Alger and Weibull (2013) and Alger and Weibull (2016), I study the problem of optimal incentive provision when agents display other-regarding preferences and different attitudes toward risk. In particular, I have shown that the optimal incentive scheme for moral agents may exhibit more risk than for selfish agents, in the sense that compensation is spread among more possible outcomes rather than aggregated around only one agent's output realization, as a consequence of the implicit incentives generated by morality. Also, in contrast to Itoh (2004) and Livio (2015), I show that tournaments are never optimal for positive degrees of morality.

Following Che and Yoo (2001), I extend the analysis to a dynamic environment and show that the optimal incentive schemes derived in the static case are also optimal when the agents are engaged in repeated interactions. The only difference between the former and the latter is intertemporal discounting, which affects the amount but not the underlying structure of the compensation schemes in the dynamic setting.

Appendix

Proofs

Proof of Lemma 1. As a starting point, I claim that any optimal contract must satisfy the incentive compatibility constraint with equality, and must be such that $w_{iLL} = 0$. To see this, I rewrite the (IC_i) constraint assuming $\rho = 0$ as follows:

$$w_{iHH}\left[p+\kappa_i q\right]+w_{iHL}\left[(1-p)-\kappa_i q\right]-w_{iLH}\left[p-\kappa_i(1-q)\right]-w_{iLL}\left[(1-p)+\kappa_i(1-q)\right] \geq \frac{c}{p-q}.$$

Note that for any $0 < q < p < 1$ and κ_i, w_{iLL} is multiplied by a strictly negative term, while w_{iHH} is multiplied by a strictly positive term. For w_{iHL}, the term multiplying it is strictly positive for $\kappa_i < \frac{1-p}{q}$ (and negative otherwise), and $w_i LH$ is multiplied by a strictly positive term whenever $\kappa_i > \frac{p}{1-q}$ (and negative otherwise).

Therefore, suppose, by contradiction, that $\mathbf{w}_i = (w_{iHH}, w_{iHL}, w_{iLH}, w_{iLL})$ is an optimal contract that satisfies the (IC_i) with some slack and also satisfies the limited liability (non-negativity) constraints. If $\mathbf{w}_i$ is such that $w_{iLH} > 0$, the principal can offer a new contract $\mathbf{w}_i' = (w_{iHH}, w_{iHL}, w_{iLH}, 0)$ which doesn't violate any of the constraints and make him better off. Indeed, note that

$$\begin{aligned}&w_{iHH}\left[p+\kappa_i q\right]+w_{iHL}\left[(1-p)-\kappa_i q\right]-w_{iLH}\left[p-\kappa_i(1-q)\right] > \\ &\quad w_{iHH}\left[p+\kappa_i q\right]+w_{iHL}\left[(1-p)-\kappa_i q\right]-w_{iLH}\left[p-\kappa_i(1-q)\right]-w_{iLL}\left[(1-p)+\kappa_i(1-q)\right],\end{aligned}$$

and

$$p^2 w_{iHH} + p(1-p)(w_{iHL}+w_{iLH}) < p^2 w_{iHH} + p(1-p)(w_{iHL}+w_{iLH}) + (1-p)^2 w_{iLL}.$$

Thus, any optimal contract must have $w_{iLL} = 0$. Moreover, a similar argument shows that $w_{iHL} = 0$ and $w_{iLH} = 0$ whenever their multiplying terms are strictly negative, i.e. whenever $\kappa_i \geq \frac{1-p}{q}$ and $\kappa_i \leq \frac{p}{1-q}$, respectively.

Now, see that the principal can reduce the expected transfers to the agents by offering a contract $\mathbf{w}_i'' = (w_{iHH} - \varepsilon_1, w_{iHL} - \varepsilon_2, w_{iLH} - \varepsilon_3, 0)$ where $\varepsilon_1 > 0$ and

$$\varepsilon_2 \begin{cases} > 0, & \text{if } \kappa_i < \frac{1-p}{q} \\ = 0, & \text{otherwise.} \end{cases}$$

and

$$\varepsilon_3 \begin{cases} > 0, & \text{if } \kappa_i > \frac{p}{1-q} \\ = 0, & \text{otherwise.} \end{cases}$$

For $\varepsilon_1, \varepsilon_2, \varepsilon_3 \approx 0$, the incentive compatibility constraint is still satisfied, while

$$p^2(w_{iHH} - \varepsilon_1) + p(1-p)\left[(w_{iHL} - \varepsilon_2) + (w_{iLH} - \varepsilon_3)\right] < p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}).$$

Thus, the principal reduces w_{iHH}, w_{iHL} and w_{iLH}, when the latter are not already zero, until the incentive compatibility constraint is satisfied with equality.

Given the argument above, attention can be restricted to incentives schemes such that $w_{iHH}, w_{iHL}, w_{iLH} \geq 0$ and $w_{iLL} = 0$. Thus, it is easy to see the three common incentive schemes, namely individual incentive scheme, team incentive scheme and tournament scheme, satisfy the conditions above. I analyze each in turn.

First, consider the individual incentive scheme, such that $w_{iHH} = w_{iHL} = w_{iH}, w_{iLH} = 0$. Substituting the first equality in (IC_i) yields

$$\begin{aligned} & w_{iH}\left[p^2 - (1-\kappa_i)qp - \kappa_i q^2 + p(1-p) - (1-\kappa_i)q(1-p) - \kappa_i q(1-q)\right] - c = 0 \\ & w_{iH} = \frac{c}{p-q} > 0, \end{aligned}$$

since $p > q$ by assumption.

A team incentive scheme would have $w_{iHH} \geq 0, w_{iHL} = w_{iLH} = 0$. By force of (IC_i) one obtains

$$w_{iHH} = \frac{c}{p^2 - (1-\kappa_i)qp - \kappa_i q^2} = \frac{c}{(p-q)(p+\kappa_i q)} > 0.$$

A tournament scheme consists of $w_{iHL} \geq 0, w_{iHH} = w_{iLH} = 0$. Again using the incentive compatibility constraint yields

$$w_{iHL} = \frac{c}{p(1-p) - (1-\kappa_i)q(1-p) - \kappa_i q(1-q)} = \frac{c}{(p-q)(1-p-\kappa_i q)}.$$

Observe that $w_{iHL} > 0$ here if and only if $1 - p - \kappa_i q > 0$, i.e. $\kappa_i < \frac{1-p}{q}$. Therefore, a tournament is a candidate solution if and only if agent i's degree of morality is not very high. ■

Proof of Lemma 2. The proof follows directly from the comparison of expected payments. For ease of exposition, they are written:

1. Individual incentive scheme: $\sum_i [p^2 + p(1-p)] w_{iH} = 2p\frac{c}{p-q}$;
2. Team incentive scheme: $p^2 \sum_i w_{iHH} = p^2 \sum_i \frac{c}{(p-q)(p+\kappa_i q)}$;
3. Tournament: $p(1-p) \sum_i w_{iHL} = p(1-p) \sum_i \frac{c}{(p-q)(1-p-\kappa_i q)}$, for $\kappa_i < \frac{1-p}{q}$.

First, compare the individual incentive scheme against the team incentive scheme. For $\kappa_i = 0$, both generate the same expected payment for the principal. However, for $\kappa_i > 0$, one has

$$\frac{pc}{(p-q)} \cdot \frac{p}{p+\kappa_i q} < \frac{pc}{(p-q)},$$

since $\frac{p}{p+\kappa_i q} < 1$. Therefore, for all $\kappa_i \in [0,1]$ and $i \in \{A, B\}$, the principal is weakly better off implementing a team incentives scheme.

Now, it is only left to compare a team incentives scheme with a tournament. To do so, suppose $\kappa_i < \frac{1-p}{q}$; otherwise the latter scheme does not satisfy the non-negativity constraints. Then, one can see that

$$\begin{aligned}
\frac{p^2 c}{(p_q)(p+\kappa_i q)} &\leq \frac{p(1-p)c}{(p-q)(1-p-\kappa_i q)} & \Leftrightarrow \\
\frac{p}{p+\kappa_i q} &\leq \frac{1-p}{1-p-\kappa_i q} & \Leftrightarrow \\
p - p^2 - \kappa_i pq &\leq p + \kappa_i q - p^2 - \kappa_i pq & \Leftrightarrow \\
\kappa_i q &\geq 0,
\end{aligned}$$

which is always satisfied, since $\kappa_i \in [0,1]$ and $q \in (0,1)$ by assumption. Therefore, a team incentives scheme is also weakly preferred by a principal over a tournament scheme. ■

Proof of Proposition 1. Building on the proof of Lemma 1, I restrict attention to schemes in which $w_{iLL} = 0$. As a first step, I show it is never optimal for the principal to offer a contract with $w_{iHL} > 0$ (given that $\kappa_i < \frac{1-p}{q}$). Indeed, suppose $\mathbf{w}_i = (w_{iHH}, w_{iHL}, w_{iLH}, 0)$ satisfy the (IC_i) with equality and the limited liability constraints; now, consider the alternative scheme $\mathbf{w}'_i = (w'_{iHH}, 0, w'_{iLH}, 0)$ such that

$$\begin{aligned} w'_{iLH} &= w_{iLH} \\ w'_{iHH} &= w_{iHH} + \frac{1-p-\kappa_i q}{p+\kappa_i q} w_{iHL}. \end{aligned}$$

Note this scheme also satisfy the incentive compatibility constraint with equality. Indeed,

$$\begin{aligned} &[p+\kappa_i q]w'_{iHH} - [p-\kappa_i(1-q)]w'_{iLH} \\ &= [p+\kappa_i q]\left(w_{iHH} + \frac{1-p-\kappa_i q}{p+\kappa_i q} w_{iHL}\right) - [p-\kappa_i(1-q)]w_{iLH} \\ &= [p+\kappa_i q]w_{iHH} + [1-p-\kappa_i q]w_{iHL} - [p-\kappa_i(1-q)]w_{iLH} \\ &= \frac{c}{p-q}. \end{aligned}$$

Now, observe the principal's expected transfers under $\mathbf{w}'_i$ are less or equal than under $\mathbf{w}_i$ if and only if

$$\begin{aligned} p^2 w'_{iHH} + p(1-p)w'_{iLH} &\leq p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) && \Leftrightarrow \\ p^2 w_{iHH} + p^2 \tfrac{1-p-\kappa_i q}{p+\kappa_i q} w_{iHL} + p(1-p)w_{iLH} &\leq p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) && \Leftrightarrow \\ p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) &\leq p(1-p) && \Leftrightarrow \\ (p+\kappa_i q)(1-p) &\geq p(1-p-\kappa_i q) && \Leftrightarrow \\ \kappa_i q &\geq 0, \end{aligned}$$

which is always satisfied, since $\kappa_i \in [0,1]$ and $q \in (0,1)$ by assumption (equality will only hold for $\kappa_i = 0$).

Therefore, any optimal contract must have $w_{iHL} = w_{iLL} = 0$, which rules out individual performance and tournament schemes. Note, however, a team incentive scheme may still be optimal. Therefore, the optimal incentives schemes must be such that

$$\begin{aligned} w_{iHH} &= \max\left\{0, \frac{1}{p+\kappa_i q}\left(\frac{c}{p-q} + [p-\kappa_i(1-q)]w_{iLH}\right)\right\} \\ w_{iLH} &\in \left[0, \frac{c}{(p-q)(\kappa_i(1-q)-p)}\right], \end{aligned}$$

for $\kappa_i > \frac{p}{1-q}$.

Given the contract described above, the principal's problem can be equivalently written as

$$\begin{aligned}
&\min_{w_{iLH}} p^2 \frac{1}{p+\kappa_i q}\left(\frac{c}{p-q} + [p-\kappa_i(1-q)]w_{iLH}\right) + p(1-p)w_{iLH} \\
&= \frac{p^2 c}{(p-q)(p+\kappa_i q)} + \left[p^2 \frac{p-\kappa_i(1-q)}{p+\kappa_i q} + p(1-p)\right] w_{iLH} \\
&= \frac{p^2 c}{(p-q)(p+\kappa_i q)} + \left[p - \frac{p^2\kappa_i}{p+\kappa_i q}\right] w_{iLH},
\end{aligned}$$

where I assume $w_{iLH} \geq 0$. Observe it is optimal for the principal to choose $w_{iLH} > 0$ iff

$$\begin{aligned}
p - \tfrac{p^2\kappa_i}{p+\kappa_i q} &< 0 \quad \Leftrightarrow \\
p + \kappa_i q &< p\kappa_i \quad \Leftrightarrow \\
\kappa_i &> \tfrac{p}{p-q}.
\end{aligned}$$

However, since $0 < q < p < 1$, the las inequality demands $\kappa_i > 1$, violating the assumption about the agents' degrees of morality. Thus, the principal optimally chooses the team incentives scheme, given by $\mathbf{w}_i^{opt} = (w_{iHH}^{opt}, 0, 0, 0)$, with

$$w_{iHH}^{opt} = \frac{c}{(p-q)(p+\kappa_i q)} > 0,$$

for all $\kappa_i \in [0,1]$, $i \in \{A, B\}$ and $0 < q < p < 1$. ■

Proof of Corollary 1. Simply take the derivative of w_{iHH} under a team incentive scheme with respect to κ_i, and note its sign is strictly negative. ■

Proof of Lemma 3. Consider the principal's problem described in the main text. The KKT conditions are necessary and sufficient to characterize the candidate solutions, and are given by

$$p^2 - \mu_i(1-\rho)w_{iHH}^{-\rho}[(p-q)(p+\kappa_i q)] - \lambda_{iHH} = 0 \tag{3.1}$$

$$p(1-p) - \mu_i(1-\rho)w_{iHL}^{-\rho}[(p-q)(1-p-\kappa_i q)] - \lambda_{iHL} = 0 \tag{3.2}$$

$$(1-p)p - \mu_i(1-\rho)w_{iLH}^{-\rho}[-(p-q)(p-\kappa_i(1-q))] - \lambda_{iLH} = 0 \tag{3.3}$$

$$(1-p)^2 - \mu_i(1-\rho)w_{iLL}^{-\rho}[-(p-q)(1-p+\kappa_i(1-q))] - \lambda_{iLL} = 0 \tag{3.4}$$

$$\begin{aligned} &w_{iHH}(p+\kappa_i q) + w_{iHL}(1-p-\kappa_i q) \\ &\qquad -w_{iLH}(p-\kappa_i(1-q)) - w_{iLL}((1-p)+\kappa_i(1-q)) \geq \tfrac{c}{p-q} \end{aligned} \tag{3.5}$$

$$\begin{aligned} &\mu_i \{w_{iHH}(p+\kappa_i q) + w_{iHL}(1-p-\kappa_i q) \\ &\qquad -w_{iLH}(p-\kappa_i(1-q)) - w_{iLL}((1-p)+\kappa_i(1-q)) - \tfrac{c}{p-q}\} = 0 \end{aligned} \tag{3.6}$$

$$w_{iHH} \geq 0 \tag{3.7}$$

$$w_{iHL} \geq 0 \tag{3.8}$$

$$w_{iLH} \geq 0 \tag{3.9}$$

$$w_{iLL} \geq 0 \tag{3.10}$$

$$\lambda_{iHH} w_{iHH} = 0 \tag{3.11}$$

$$\lambda_{iHL} w_{iHL} = 0 \tag{3.12}$$

$$\lambda_{iLH} w_{iLH} = 0 \tag{3.13}$$

$$\lambda_{iLL} w_{iLL} = 0 \tag{3.14}$$

$$\lambda_{iHH} \geq 0 \tag{3.15}$$

$$\lambda_{iHL} \geq 0 \tag{3.16}$$

$$\lambda_{iLH} \geq 0 \tag{3.17}$$

$$\lambda_{iLL} \geq 0 \tag{3.18}$$

$$\mu_i \geq 0 \tag{3.19}$$

for all $i \in \{1,2\}$, where μ_i is the Lagrange multiplier associated with the incentive compatibility constraint while λ_is are the ones associated with the nonnegativity constraint.

As was the case under risk aversion, (IC_i) must bind. If that was not the case, $\mu_i = 0$ would imply through equations (3.1) – (3.4) that $\lambda_{iHH}, \lambda_{iHL}, \lambda_{iLH}, \lambda_{iLL} > 0$, and thus, by force of the complementary slackness conditions (3.11)–(3.14), that $w_{iHH} = w_{iHL} = w_{iLH} =$

$w_{iLL} = 0$. However, substituting into (3.5), one obtains $0 \geq \frac{c}{p-q} > 0$, a contradiction[12].

The first three incentive schemes described in the text are obtained by using equation (3.5), the incentive compatibility constraint, with equality and considering each case in turn:

1. Individual incentive scheme: $w_{iHH} = w_{iHL} > 0 = w_{iLH} = w_{iLL}$;

2. Team incentive scheme: $w_{iHH} > 0 = w_{iHL} = w_{iLH} = w_{iLL}$

3. Tournament scheme: $w_{iHL} > 0 = w_{iHH} = w_{iLH} = w_{iLL}$

For the relative performance scheme, assume $1 - p - \kappa_i q > 0$ and compute the ratio of equations (3.1) and (3.2),

$$\begin{aligned}
\tfrac{p^2}{p(1-p)} &= \tfrac{\mu_i(1-\rho)w_{iHH}^{-\rho}(p-q)(p+\kappa_i q)}{\mu_i(1-\rho)w_{iHL}^{-\rho}(p-q)(1-p-\kappa_i q)} && \Leftrightarrow \\
\left(\tfrac{w_{iHL}}{w_{iHH}}\right)^{\rho} &= \tfrac{p(1-p-\kappa_i q)}{(1-p)(p+\kappa_i q)} && \Leftrightarrow \\
w_{iHL} &= w_{iHH} \underbrace{\left(\frac{p(1-p-\kappa_i q)}{(1-p)(p+\kappa_i q)}\right)^{\frac{1}{\rho}}}_{=A(\kappa_i,\rho)}
\end{aligned}$$

Since I assume $1 - p - \kappa_i q > 0$, $\kappa_i \in [0, 1]$ and $0 < q < p < 1$, note that $A(\kappa_i, \rho) > 0$. Moreover, $A(0, \rho) = 1$ and

$$\frac{\partial A(\kappa_i, \rho)}{\partial \kappa_i} \propto -pq(1-p)(p+\kappa_i q) - q(1-p)p(1-p-\kappa_i q) < 0,$$

so that $A(\kappa_i, \rho) \in (0, 1]$ for all $\kappa_i \in [0, 1]$ and $\rho \in (0, 1)$. Plugging w_{iHH}, $w_{iHL} = w_{iHH}A(\kappa_i, \rho)$ and $w_{iLH} = w_{iLL} = 0$ in (3.5) yields the result, taking into consideration the nonnegativity constraint as well. ■

Proof of Lemma 4. Suppose $1-p-\kappa_i q > 0$, so that a tournament is a candidate solution to the principal's problem. For $\rho \in (0, 1)$, the principal prefers a tournament over an individual performance scheme if and only if the expected transfers under the former are smaller than

[12] An argument similar to the one used in the risk-neutral case could be employed here as well, and would fit the more general case of a utility function of wealth satisfying $u' > 0$, $u'' \leq 0$, $u(0) = 0$.

under the latter, that is, iff

$$
\begin{aligned}
&p(1-p)\left(\tfrac{c}{(p-q)(1-p-\kappa_i q)}\right)^{\frac{1}{1-\rho}} < [p^2 + p(1-p)]\left(\tfrac{c}{p-q}\right)^{\frac{1}{1-\rho}} && \Leftrightarrow \\
&(1-p)^{1-\rho}\tfrac{c}{(p-q)(1-p-\kappa_i q)} < \tfrac{c}{p-q} && \Leftrightarrow \\
&\kappa_i < \tfrac{1-p}{q}[1-(1-p)^{-\rho}]
\end{aligned}
$$

Since $\kappa_i \in [0,1]$ by assumption, the inequality above holds only if $1-(1-p)^{-\rho} \geq 0$, which is equivalent to

$$1 \geq \frac{1}{(1-p)^{\rho}} > 1,$$

a contradiction. ■

Proof of Lemma 5. The principal's expected payments under team incentives are smaller than under individual performance iff

$$
\begin{aligned}
&p^2\left(\tfrac{c}{(p-q)(p+\kappa_i q)}\right)^{\frac{1}{1-\rho}} < [p^2 + p(1-p)]\left(\tfrac{c}{p-q}\right)^{\frac{1}{1-\rho}} && \Leftrightarrow \\
&p^{1-\rho}\tfrac{c}{(p-q)(p+\kappa_i q)} < \tfrac{c}{p-q} && \Leftrightarrow \\
&\kappa_i q > p^{1-\rho} - 1 && \Leftrightarrow \\
&\kappa_i > \tfrac{p}{q} \cdot \underbrace{\frac{1-p^{\rho}}{p^{\rho}}}_{=\overline{\kappa}(\rho)}. && \blacksquare
\end{aligned}
$$

Proof of Proposition 2. Using the KKT conditions obtained in the proof of Lemma 3, I will look for the optimal incentive scheme. As argued before, such scheme must satisfy the incentive compatibility constraint with equality (i.e. $\mu_i > 0$ for all $i \in \{1,2\}$). Moreover, it must be such that $w_{iLL} = 0$. Indeed, on equation (3.4), note that $-(p-q)[(1-p)+\kappa_i(1-q)] < 0$ for all $0 < q < p < 1$ and $\kappa_i \in [0,1]$; therefore, if $w_{iLL} > 0$, the complementary slackness condition implies that $\lambda_{iLL} = 0$, and thus the left-hand side of equation (4) is strictly positive, contradicting the first-order condition.

A similar argument can be used on equations (3.2) and (3.3): whenever the term multiplying the wage is negative, a solution must have the nonnegativity constraint binding. Therefore,

$$\kappa_i \geq \tfrac{1-p}{q} \Rightarrow w_{iHL} = 0, \quad (A.1)$$

and

$$\kappa_i \leq \tfrac{p}{1-q} \Rightarrow w_{iLH} = 0. \quad (A.2)$$

One can easily check that

$$\frac{1-p}{q} < 1 < \frac{p}{1-q} \Leftrightarrow p+q > 1, \quad \frac{1-p}{q} \geq 1 \geq \frac{p}{1-q} \Leftrightarrow p+q \leq 1,$$

so the analysis can be conveniently divided in two cases, namely $p+q > 1$ and $p+q \leq 1$.

Suppose first that $p+q > 1$. If $\kappa_i \in \left[\frac{1-p}{q}, 1\right]$, conditions $(A.1)$ and $(A.2)$ imply that $w_{iHL} = w_{iLH} = 0$, and the only solution candidate is the team incentive scheme described in Lemma 3. On the other hand, for $\kappa_i \in \left[0, \frac{1-p}{q}\right)$, the two conditions above imply that $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$, so the four incentive schemes in Lemma 3 are candidate solutions.

It is easy to see that the relative performance scheme performs at least as good as any of the other three schemes in this case. Indeed, let $\mathcal{C} = \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH}, w_{iHL} \geq 0, w_{iLH} = w_{iLL} = 0\right\}$ denote the set of contracts than can be offered if $p+q > 1$ and $\kappa_i \in \left[0, \frac{1-p}{q}\right)$. In a similar fashion, let

$$\begin{aligned}
\mathcal{C}^{Team} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH} \geq 0, w_{iHL} = w_{iLH} = w_{iLL} = 0\right\} \\
\mathcal{C}^{Ind} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH} = w_{iHL} \geq 0, w_{iLH} = w_{iLL} = 0\right\} \\
\mathcal{C}^{Tour} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHL} \geq 0, w_{iHH} = w_{iLH} = w_{iLL} = 0\right\} \\
\mathcal{C}^{Rel} &= \left\{\mathbf{w} \in \mathbb{R}^4_+ : w_{iHH}, w_{iHL} \geq 0, w_{iLH} = w_{iLL} = 0\right\},
\end{aligned}$$

denote the set of contracts satisfying the conditions for the performance schemes described in Lemma 3. One can readily note that $\mathcal{C}^{Team}, \mathcal{C}^{Ind}, \mathcal{C}^{Tour} \subset \mathcal{C}$ and $\mathcal{C}^{Rel} = \mathcal{C}$. Therefore, a team, individual or tournament schemes add more constraints to the set of contracts under which the principal can maximize his profits, and must not yield a strictly higher profit than the one obtained under the more relaxed constraint set $\mathcal{C}$.

If $p+q \leq 1$ and $\kappa < \frac{p}{1-q}$, the optimal scheme is the same as in the previous paragraph, i.e. the relative performance scheme with $w_{iHH}, w_{iHL} \geq 0$ and $w_{iLH} = w_{iLL} = 0$. However, if $p+q \leq 1$ and $\kappa \in \left[\frac{p}{1-q}, 1\right]$, the principal can maximize over the set $\widetilde{\mathcal{C}} =$

$\left\{\mathbf{w} \in \mathbb{R}_+^4 : w_{iHH}, w_{iHL}, w_{iLH} \geq 0, w_{iLL} = 0\right\}$. Now, the contract sets defined by the four schemes presented above are strict subsets of $\widetilde{\mathcal{C}}$ and cannot, thus, yield a strictly higher payoff to the principal. ■

Proof of Proposition 3. Follows directly from the argument in the main text. ■

Proof of Corollary 3. Follows from the observation that the proposed incentive schemes satisfy the static incentive compatibility constraint and, thus, the dynamic version considered in Proposition 3. ■

Proof of Proposition 4. The proof follows closely the argument developed in Lemma 3 and Proposition 2. Suppose that $U_i(\mathbf{w}^*, 0, 1; \kappa_i) > U_i(\mathbf{w}^*, 0, 0; \kappa_i)$. The principal's problem becomes

$$\begin{aligned}
\min_{\mathbf{w}} \quad & p^2 w_{iHH} + p(1-p)(w_{iHL} + w_{iLH}) + (1-p)^2 w_{iLL} \\
s.t. \quad & p^2 w_{iHH}^{1-\rho} + p(1-p)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-p)^2 w_{iLL}^{1-\rho} \geq \\
& \quad (1-\delta)\left[(1-\kappa_i)(qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho})\right. \\
& \quad \left. +\kappa_i(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\right] \\
& \quad +\delta(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho}) && (DIC_i) \\
& w_{iHH}, w_{iHL}, w_{iLH} w_{iLL} \geq 0 && (LL_i)
\end{aligned}$$

whose KKT conditions are given by

$$
\begin{aligned}
& p^2 - \lambda_{iHH} \\
& \quad -\mu_i(1-\rho)w_{iHH}^{-\rho}[p^2 - (1-\delta)((1-\kappa_i)pq + \kappa_i q^2) - \delta q^2] = 0 && (4.1)\\
& p(1-p) - \lambda_{iHL} \\
& \quad -\mu_i(1-\rho)w_{iHL}^{-\rho}[p(1-p) - (1-\delta)((1-\kappa_i)q(1-p) + \kappa_i q(1-q)) - \delta q(1-q)] = 0 && (4.2)\\
& (1-p)p - \lambda_{iLH} \\
& \quad -\mu_i(1-\rho)w_{iLH}^{-\rho}[p(1-p) - (1-\delta)((1-\kappa_i)(1-q)p + \kappa_i q(1-q)) - \delta q(1-q)] = 0 && (4.3)\\
& (1-p)^2 - \lambda_{iLL} \\
& \quad -\mu_i(1-\rho)w_{iLL}^{-\rho}[(1-p)^2 - (1-\delta)((1-\kappa_i)(1-q)(1-p) + \kappa_i(1-q)^2) - \delta(1-q)^2] = 0 && (4.4)\\
& p^2 w_{iHH}^{1-\rho} + p(1-p)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-p)^2 w_{iLL}^{1-\rho} \geq \\
& \quad (1-\delta)\left[(1-\kappa_i)(qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho})\right. \\
& \quad \left.+\kappa_i(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\right] \\
& \quad +\delta(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho}) && (4.5)\\
& \mu_i \left\{p^2 w_{iHH}^{1-\rho} + p(1-p)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-p)^2 w_{iLL}^{1-\rho}\right. \\
& \quad (1-\delta)\left[(1-\kappa_i)(qpw_{iHH}^{1-\rho} + q(1-p)w_{iHL}^{1-\rho} + (1-q)pw_{iLH}^{1-\rho} + (1-q)(1-p)w_{iLL}^{1-\rho})\right. \\
& \quad \left.+\kappa_i(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\right] \\
& \quad \left.+\delta(q^2 w_{iHH}^{1-\rho} + q(1-q)(w_{iHL}^{1-\rho} + w_{iLH}^{1-\rho}) + (1-q)^2 w_{iLL}^{1-\rho})\right\} = 0 && (4.6)\\
& w_{iHH} \geq 0 && (4.7)\\
& w_{iHL} \geq 0 && (4.8)\\
& w_{iLH} \geq 0 && (4.9)\\
& w_{iLL} \geq 0 && (4.10)\\
& \lambda_{iHH} w_{iHH} = 0 && (4.11)\\
& \lambda_{iHL} w_{iHL} = 0 && (4.12)\\
& \lambda_{iLH} w_{iLH} = 0 && (4.13)\\
& \lambda_{iLL} w_{iLL} = 0 && (4.14)\\
& \lambda_{iHH} \geq 0 && (4.15)\\
& \lambda_{iHL} \geq 0 && (4.16)\\
& \lambda_{iLH} \geq 0 && (4.17)\\
& \lambda_{iLL} \geq 0 && (4.18)\\
& \mu_i \geq 0 && (4.19)
\end{aligned}
$$

By assumption, $1 > p > q > 0$, and thus

$$\begin{aligned}
&(1-\delta)\left[(1-\kappa_i)(1-q)(1-p)+\kappa_i(1-q)^2\right]+\delta(1-q)^2 \\
&\quad > (1-\delta)\left[(1-\kappa_i)(1-q)(1-q)+\kappa_i(1-q)^2\right]+\delta(1-q)^2 \\
&\quad = (1-q)^2 \\
&\quad > (1-p)^2,
\end{aligned}$$

so equation (4.4) can only be satisfied if $w_{iLL} = 0$. Otherwise, the complementary slackness condition (4.14) would imply $\lambda_{iLL} = 0$ and equation (4.4) would be violated for any $\mu_i \geq 0$. Moreover, there exists no solution such that $\lambda_{iHH}, \lambda_{iHL}, \lambda_{iLH} > 0$: if that was true, then $w_{iHH} = w_{iHL} = w_{iLH} = w_{iLL} = 0$, and (4.5) would be reduced to $-c \geq 0$, a contradiction.

Notice that $w_{iHH} > 0$ or $w_{iHL} > 0$ or $w_{iLH} > 0$ only if $\mu_i > 0$ and the terms in brackets in equations (4.1) – (4.3), respectively, are strictly positive. Thus, in any solution, the dynamic incentive compatibility constraint must be binding.

On equation (4.1) it is easy to see that $(1-\delta)((1-\kappa_i)pq+\kappa_i q^2)+\delta q^2 < (1-\delta)((1-\kappa_i)pp+\kappa_i q^2)+\delta q^2 < (1-\delta)((1-\kappa_i)pq+\kappa_i p^2)+\delta p^2 = p^2$, so that $w_{iHH} > 0$ for any values of δ and κ_i. On equation (4.2), $p(1-p) > (1-\delta)((1-\kappa_i)q(1-p)+\kappa_i q(1-q))+\delta q(1-q)$ iff

$$\kappa_i < \overline{\kappa}(\delta) = \frac{p(1-p)-\delta q(1-q)}{(1-\delta)q(p-q)} - \frac{1-p}{p-q},$$

and, on equation (4.3), $p(1-p) > (1-\delta)((1-\kappa_i)(1-q)p+\kappa_i q(1-q))+\delta q(1-q)$ iff

$$\kappa_i > \underline{\kappa}(\delta) = \frac{\delta q(1-q)-p(1-p)}{(1-\delta)(1-q)(p-q)} + \frac{p}{p-q}.$$

Notice that

$$\overline{\kappa}(0) = \frac{1-p}{q}, \quad \underline{\kappa}(0) = \frac{p}{1-q},$$

and

$$\frac{\partial\overline{\kappa}(\delta)}{\partial\delta}\begin{cases} >0 & \text{if } p+q<1 \\ =0 & \text{if } p+q=1 \\ <0 & \text{if } p+q>1 \end{cases}, \frac{\partial\underline{\kappa}(\delta)}{\partial\delta}\begin{cases} <0 & \text{if } p+q<1 \\ =0 & \text{if } p+q=1 \\ >0 & \text{if } p+q>1 \end{cases}.$$

Moreover,

$$\lim_{\delta\to 1}\overline{\kappa}(\delta) = \begin{cases} +\infty & \text{if } p+q<1 \\ -\infty & \text{if } p+q>1 \end{cases}, \lim_{\delta\to 1}\underline{\kappa}(\delta) = \begin{cases} -\infty & \text{if } p+q<1 \\ +\infty & \text{if } p+q>1 \end{cases}.$$

As was the case in Proposition 2, if $p+q>1$, then $\underline{\kappa}(0)>1>\overline{\kappa}(0)>0$. Thus, for $\kappa_i \geq \overline{\kappa}(\delta)$, a team performance scheme $\mathbf{w}_i^{Team} = (w_{iHH}^{Team}, 0, 0, 0)$ such that

$$w_{iHH}^* = \left(\frac{c}{p^2-(1-\delta)q[(1-\kappa_i)q+\kappa_i q]-\delta q^2}\right)^{\frac{1}{1-\rho}}$$

is optimal. For $\kappa_i < \overline{\kappa}(\delta)$, the relative performance scheme $\mathbf{w}_i^{Rel} = (w_{iHH}^{Rel}, w_{iHL}^{Rel}, 0, 0)$ is optimal, with

$$w_{iHL}^{Rel} = w_{iHH}^{Rel} \times \underbrace{\left(\frac{p}{1-p}\cdot\frac{p(1-p)-(1-\delta)q[(1-\kappa_i)(1-p)+\kappa_i(1-q)]-\delta q(1-q)}{p^2-(1-\delta)q[(1-\kappa_i)p+\kappa_i q]-\delta q^2}\right)^{\frac{1}{\rho}}}_{=\mathcal{A}(\kappa_i,\delta,\rho)},$$

$$w_{iHH}^{Rel} = \left(\frac{c}{p[p+(1-p)\mathcal{A}]-(1-\delta)q\{(1-\kappa_i)[p+(1-p)\mathcal{A}]+\kappa_i[q+(1-q)\mathcal{A}]\}-\delta q[q+(1-q)\mathcal{A}]}\right)^{\frac{1}{1-\rho}}.$$

If $p+q<1$, then $0<\underline{\kappa}(0)<1<\overline{\kappa}(0)$. For $\kappa_i \leq \underline{\kappa}(\delta)$, the optimal incentive scheme is the relative performance described in the last paragraph. On the other hand, for $\kappa_i > \underline{\kappa}(\delta)$, the optimal incentive scheme is $\mathbf{w}_i^{Comp} = (w_{iHH}^{Comp}, w_{iHL}^{Comp}, w_{iLH}^{Comp}, 0)$ such that

$$w_{iLH}^{Comp} = w_{iHH}^{Comp} \times \underbrace{\left(\frac{p}{1-p}\cdot\frac{p(1-p)-(1-\delta)(1-q)[(1-\kappa_i)(1-p)+\kappa_i(1-q)]-\delta(1-q)^2}{p^2-(1-\delta)q[(1-\kappa_i)p+\kappa_i q]-\delta q^2}\right)^{\frac{1}{\rho}}}_{=\mathcal{B}(\kappa_i,\delta,\rho)},$$

$$w_{iHL}^{Comp} = w_{iHH}^{Comp} \times \underbrace{\left(\frac{p}{1-p}\cdot\frac{p(1-p)-(1-\delta)q[(1-\kappa_i)(1-p)+\kappa_i(1-q)]-\delta q(1-q)}{p^2-(1-\delta)q[(1-\kappa_i)q+\kappa_i q]-\delta q^2}\right)^{\frac{1}{\rho}}}_{=\mathcal{A}(\kappa_i,\delta,\rho)},$$

$$w_{iHH}^{Comp} = \left(\frac{c}{D(\kappa_i,\delta,\rho)}\right)^{\frac{1}{1-\rho}},$$

where

$$\begin{aligned} D(\kappa_i,\delta,\rho) = {} & p[p+(1-p)(\mathcal{A}+\mathcal{B})] \\ & -(1-\delta)\left\{(1-\kappa_i)[pq+q(1-p)\mathcal{A}+(1-q)p\mathcal{B}]+\kappa_i q[q+(1-q)(\mathcal{A}+\mathcal{B})]\right\} \\ & -\delta q[q+(1-q)(\mathcal{A}+\mathcal{B})]. \end{aligned}$$

■

References

Alchian, A. A. and H. Demsetz (1972, December). Production , information costs, and economic organization. *American Economic Review 62*(5), 777–95.

Alger, I. and J. W. Weibull (2013, November). Homo moralis - preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2302.

Alger, I. and J. W. Weibull (2016). Evolution and kantian morality. *Games and Economic Behavior 98*(C), 56–67.

Becker, G. S. (1974, Nov.-Dec.). A theory of social interactions. *Journal of Political Economy 82*(6), 1063–93.

Bénabou, R. and J. Tirole (2006, December). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Besley, T. and M. Ghatak (2005, June). Competition and incentives with motivated agents. *American Economic Review 95*(3), 616–636.

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

Bowles, S. and S. Polania-Reyes (2012, June). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature 50*(2), 368–425.

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

Crawford, V. P. and E. M. Knoer (1981, March). Job matching with heterogeneous firms and workers. *Econometrica 49*(2), 437–50.

de la Rosa, L. E. (2011). Overconfidence and Moral Hazard. *Games and Economic Behavior 73*(2), 429–451.

Demougin, D. and C. Fluet (2003). Inequity aversion in tournaments. Cahiers de recherche 0322, CIRPEE.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2009, 03). Homo reciprocans: Survey evidence on behavioural outcomes. *Economic Journal 119*(536), 592–612.

Dur, R. and J. Sol (2010, July). Social interaction, co-worker altruism, and incentives. *Games and Economic Behavior 69*(2), 293–301.

Dur, R. and J. Tichem (2015, 09). Altruism and relational incentives in the workplace. *Journal of Economics & Management Strategy 24*(3), 485–500.

Englmaier, F. and A. Wambach (2010, July). Optimal incentive contracts under inequity aversion. *Games and Economic Behavior 69*(2), 312–328.

Fehr, E., A. Klein, and K. M. Schmidt (2007, 01). Fairness and contract design. *Econometrica 75*(1), 121–154.

Fehr, E. and K. M. Schmidt (2000, May). Fairness, incentives, and contractual choices. *European Economic Review 44*(4-6), 1057–1068.

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

Itoh, H. (2004). Moral hazard and other-regarding preferences. *The Japanese Economic Review 55*(1), 18–45.

Kolm, S. and J. M. Ythier (Eds.) (2006a). *Handbook of the Economics of Giving, Altruism and Reciprocity* (1 ed.), Volume 1. Elsevier.

Kolm, S. and J. M. Ythier (Eds.) (2006b). *Handbook of the Economics of Giving, Altruism and Reciprocity* (1 ed.), Volume 2. Elsevier.

Laffont, J.-J. (1975, November). Macroeconomic constraints, economic efficiency and ethics: An introduction to kantian economics. *Economica 42*(168), 430–37.

Laffont, J.-J. and J. Tirole (1986, June). Using cost observation to regulate firms. *Journal of Political Economy 94*(3), 614–41.

Livio, L. (2015). Friend or foes? optimal incentives for reciprocal agents. *Mimeo*.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and L. J. Taylor (2002, September). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review 92*(4), 850–873.

Rabin, M. (1993, December). Incorporating fairness into game theory and economics. *American Economic Review 83*(5), 1281–1302.

Rey-Biel, P. (2008, 06). Inequity aversion and team incentives. *Scandinavian Journal of Economics 110*(2), 297–320.

Formal and Informal Risk-Sharing under Altruistic Preferences*

Roberto Sarkisian†

May 5, 2020

Abstract

Can informal risk-sharing crowd out formal insurance policies? I consider an insurance model where altruistic agents can buy insurance, self-protect against a loss and cross-insure by means of bilateral transfers. Such altruism-driven transfers may lead the agents to free-ride on each others' choices of effort and demand for formal insurance, and hence hinder the development of formal insurance markets. Absent any information asymmetries, I show that an actuarially fair insurance policy providing full coverage can be crowded out by the agents' informal risk-sharing arrangements. A similar result holds when the agents' self-protection efforts cannot be contracted upon by the insurer. These findings suggest a novel source of inefficiency in insurance markets, namely prosociality among the insurees.

*This work is prepared under the supervision of Ingela Alger and François Salanié, as part of my Ph.D. thesis at Toulouse School of Economics. I thank Andrea Attar, Renee Bowen, Marco Celentani, Victor Gay, Renato Gomes, Matthew Jackson, Natalia Kovaleva, Ignacio Ortuño-Ortín, and Takuro Yamashita for very valuable comments and feedback. I also thank seminar audiences at the Africa Meeting of the Econometric Society (2019), Stanford University (June, 2018), Toulouse School of Economics, Universidad Carlos III Madrid (Sept., 2019), LUISS Guido Carli (Jan., 2020), and University of Rome Tor Vergata (April, 2020).

†Toulouse School of Economics. *E-mail:* roberto.sarkisian@tse-fr.eu

Keywords: Moral hazard in teams, risk-sharing, insurance, altruism.

JEL Classification: D02,D64,D82,D86.

1 Introduction

It is not uncommon to observe individuals engaging in market trades and nonmarket arrangements for risk-sharing purposes. Individuals rely not only on insurance contracts to protect themselves against occasional losses but also on family members and friends to help in adverse situations. In most cases, such reliance on nonmarket arrangements is more pronounced in less developed societies or those where kinship ties among members is higher, while more developed societies are associated with stronger formal institutions, such as the rule of law and well-established banking and credit markets (Cox et al., 2006; Cox and Fafchamps, 2008).

Figure 1 displays total non-life insurance premiums by private companies as a percentage of GDP. While average insurance penetration in the poorer countries, displayed on the top two panels of Figure 1, is less than half of the one displayed by the richer countries in the bottom two panels, no clear pattern can be observed within each group. Moreover, countries with lower insurance penetration rates, such as Brazil, Bolivia, India, and Indonesia, display strong family ties, while countries with higher insurance penetration, such as France, Germany, Netherlands, and Switzerland display weaker family ties (Alesina and Giuliano, 2014).[1] Among European countries, Costa-Font (2010) estimates a negative effect of family ties in the demand for long-term care insurance.[2] While a common interpretation of these patterns is that the lack of formal insurance fosters informal family insurance (Coate and

[1]The United States display the second highest insurance penetration rates in the sample, but also has strong family ties. This result is biased due to the presence of poor and tightly-knit minorities communities, according to Alesina and Giuliano (2014).

[2]While Alesina and Giuliano (2014) construct their measure of family ties using only the World Value Survey, Costa-Font (2010) constructs an index of *familism* using both the World Values Survey and Eurobarometer survey questions on the importance of family.

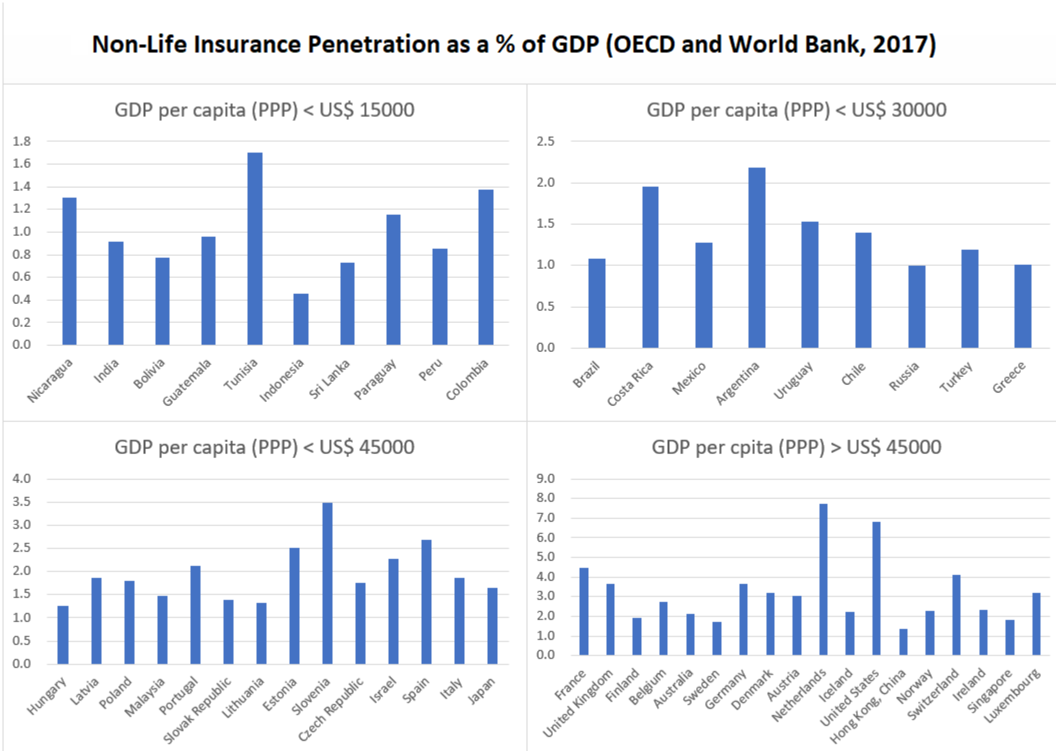


Figure 1: Non-life insurance penetration as a percentage of GDP.
Source for insurance penetration: OECD. URL: $https://stats.oecd.org/Index.aspx?DataSetCode=INSIND$. This includes health, vehicles and housing insurance policies. Source for GPD: World Bank. URL: $https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?end=2017\&start=2017$. The values are the most recent available, PPP, current international USD.

Ravallion, 1993; Udry, 1994; Besley, 1995), I ask whether a causal relationship in the opposite direction can be at work as well, i.e. whether formal insurance contracts offered by a firm can be crowded out by informal risk-sharing arrangements among individuals.

In what follows, I examine the demand for formal private insurance policies when agents can also engage in informal risk-sharing arrangements with other individuals, possibly because of family ties. In accordance with the empirical evidence mentioned above, I will let the strength of altruism be a key driver of insurance demand. I consider a model in which altruistic agents engage in three actions: (i) buy *market insurance* by individually engaging in a contractual relationship with the insurance provider; (ii) *self-protect* by exerting effort to reduce the probability of a loss taking place; and (iii) *cross-insure* by transferring wealth to one another after observing the realized outcomes.

While the usual incentive problem due to moral hazard is present in my model, altruism introduces new channels through which formal insurance can be, a priori, either hindered or fostered. The main channel through which altruism affects the outcome is cross-insurance transfers. Intuitively, an agent's cross-insurance transfers become larger as he becomes more altruistic, and they induce better risk-sharing among the agents. However, larger cross-insurance transfers generate a substitution effect: one agent can choose to rely on another's transfers instead of purchasing an insurance policy himself. Additionally, a free-riding effect on self-protection effort also takes place, since cross-insurance transfers reduce the risk faced by an agent and, thus, his incentives to exert a costly effort to avoid a loss. Therefore, a firm's insurance policy can be crowded out by altruism if the substitution and free-riding effects generated by cross-insurance transfers are large.

A second channel also exists which can affect the agents' demand for formal insurance contracts. As his altruism increases, an agent exerts more effort for two reasons. First, he wishes to avoid a loss from taking place so that he does not burden the other agent with cross-insurance transfers. Second, he works harder to avoid a loss in order to be able to help his partner if the other agent does suffer a loss. A priori, this empathy effect[3] may thus help alleviate the free-riding effect induced by a reduction in the risk due to formal insurance.

Last, formal insurance contracts affect both cross-insurance transfers and self-protection efforts directly. The latter corresponds to the usual free-riding effect in moral hazard problems: as risk is reduced by the formal insurance policy, the agents' incentives to exert a costly effort to avoid a loss are also reduced. For the former, larger coverage against a loss decreases the need for cross-insurance transfers between agents. Overall, the combination of these substitution, free-riding, and empathy effects may lead to the crowding-out of formal insurance contracts.

My main results are twofold. First, I show that insurance contracts can be crowded out when a self-interested individual substitutes a formal insurance contract for an altruistic agent's cross-insurance transfers. Absent moral-hazard between insurer and insurees, I show

[3] As discussed in Alger and Weibull (2010).

that the substitution effect can be strong enough that a self-interested agent rejects an actuarially fair insurance policy providing him with full coverage.

Under moral-hazard, the contract that would maximize the firm's expected profit when trading with a purely self-interested agent will be rejected by sufficiently altruistic agents, since it fails to take into account the cross-insurance transfers made by these prosocial individuals. In this case, the offered contract sets too high a premium vis-à-vis the coverage and the risk such altruistic agents face when cross-insurance transfers are present. Therefore, individuals are better off not buying the insurer's contract.

Second, the firm can induce altruistic agents to buy insurance contracts with small premium and coverage. Intuitively, the insurer anticipates the agents' cross-insurance transfers complementing his contract and offers a contract with partial coverage that doesn't induce too low an effort on the insurees' part. Since the agents are risk-averse, this contract may have higher than actuarially fair premium, thus providing the insurer with positive expected profits at the same time that it prompts the agents to buy it.

This paper proceeds as follows. Section 2 introduces the notation and the single-agent model that serves as a benchmark for the multi-agent setting presented later in that section. Section 3 analyzes insurance demand when self-protection is not available to the agents, while Section 4 does the same in the presence of self-protection. Section 5 introduces the insurance firm and evaluates the offers it makes to the agents. Section 6 discusses testable implications, extensions and future research, before the concluding remarks that appear in Section 7. For ease of exposition, all proofs are relegated to the Appendix.

1.1 Related Literature

The literature has identified several factors that may prevent the emergence of formal markets. A large body of literature has devoted attention to the information asymmetries plaguing the relationship between insurers and insurees since the pioneering works of Ehrlich and Becker (1972), Pauly (1974), and Rothschild and Stiglitz (1976). More recently, Hendren (2013) and Attar et al. (2019) point to adverse selection as being a contributing factor to market breakdown, in the same spirit as Akerlof (1970), while Einav and Finkelstein (2018) underline the role of moral hazard. Another possibility for the underdevelopment of formal, private insurance markets is the presence of additional players, such as the government, that offer a substitute to the contracts proposed by private insurance companies. However, Brown and Finkelstein (2011) and Gruber and Simon (2008)empirically show that the crowding-out of private insurance by expansions in public health insurance is not that large. While these articles focus on supply-side issues pertaining to insurance markets, I focus on factors affecting the demand side that hinder the emergence of formal insurance markets. The novelty of my results rests on bringing forth the link between altruism and the underdevelopment of formal insurance markets.

My model studies the interaction of a formal market, captured by a principal offering insurance contracts to a pair of agents, and informal arrangements, represented by transfers made by this pair of agents. There are a small number of theoretical articles that analyze such an interaction. Arnott and Stiglitz (1991) provide an early modeling attempt at understanding the interaction between market and nonmarket insurance, but where the agents have to choose exclusively between formal and informal trade. My model does not have the same limitation: agents can participate simultaneously in formal and informal trades and therefore choose when to trade in both, one or none of them. Kranton (1996) and Jain (1999) also consider the interaction between formal and informal arrangements, but focus, respectively, on monetary exchange and credit rather than insurance as in my setting.

A large body of literature studying informal risk-sharing agreements in poor regions has also been developed. Udry (1990, 1994) and Townsend (1994) were among the first to study the behavior of rural villagers engaging in *quasi-credit* transactions with each other in order to smooth consumption. The main observations of these analyses are that households in these rural areas mostly engage in trades with other individuals who are close to them, even when they have the possibility to sign a credit contract with a bank or interact with local and itinerary merchants that could also provide them with credit. Coate and Ravallion (1993), Kocherlakota (1996), Ligon et al. (2002), Fafchamps and Lund (2003) and, more recently, Dubois et al. (2008), pay attention to the issue of limited commitment present in these interactions by considering infinitely repeated structures, where the threat of exclusion from future trades sustains these informal arrangements. In contrast to these papers, my model posits altruism as the channel through which informal risk-sharing is sustained in a one-shot interaction. Foster and Rosenzweig (2001) show that altruism plays an important role in sustaining informal risk-sharing among individuals, but do not evaluate the impact of such arrangements in the demand for formal risk-sharing instruments as in this paper.

Motivated by the findings of Udry (1990, 1994), Townsend (1994), and Foster and Rosenzweig (2001) that most risk-sharing agreements take place among individuals who are closely related to one another, a recent literature has explored the role of altruism, transfers and risk-sharing in networks. Fafchamps and Lund (2003), Bramoullé and Kranton (2007), Cox and Fafchamps (2008), and Di Falco and Bulte (2013) empirically explore altruism in economic networks as a means to diversify risk, and their results show that stronger kinship ties lead to more risk-sharing among agents in the same network. Bourlès et al. (2017) and Bourlès et al. (2018) provide formal models of social networks where agents care about each other and may transfer funds to one another to share risk. Their focus, however, lies in identifying conditions in the network structure that induce positive transfers among its members. I study a simpler network structure, namely, one with only two individuals, but focus on the agents' equilibrium choices of risk-sharing, self-protection and demand for insurance policies, where the last two elements are absent in Bourlès et al. (2017) and Bourlès et al. (2018).

While this paper can be seen as evaluating the role of altruistic preferences in canonical insurance settings of asymmetric information, it also provides an independent contribution on modelling prosocial behaviors. There is by now a large literature on the effects on outcomes in otherwise standard economic models of other-regarding preferences - such as altruism (Becker, 1974, 1976), warm-glow (Andreoni, 1990), inequity aversion (Fehr and Schmidt, 1999) and morality (Alger and Weibull, 2013, 2016). Recent papers exploring the role of other-regarding behavior in contracting situations are Itoh (2004), Rotemberg (2006), Rey-Biel (2008), von Siemens (2011), Sarkisian (2017) and Biener et al. (2018). Most of these papers explore the effects of other-regarding preferences in alleviating contracting constraints, either in moral hazard (i.e., reducing the incentives to free-ride or to slack) or screening (e.g., individuals self-selecting to job propositions according to their perception of the firms' missions), while my main concern is how prosociality influences the choices of the agents to engage in trades either with a formal institution or informally among themselves. Closer to my main message, Bernheim and Stark (1988), Lindbeck and Weibull (1988), and Alger and Weibull (2017) also study situations in which altruism has adverse effects on economic outcomes but do not study how altruism affects market structures. My model extends the stage game in Alger and Weibull (2010) by introducing an insurer and her contractual offers, but while I focus on how altruism will affect the demand for formal insurance, Alger and Weibull (2010) ask the reverse question of how the environment affects the evolution of altruism itself.[4]

The papers by Costa-Font (2010), Costa-Font and Courbage (2015), Cremer et al. (2013, 2017), Cremer and Roeder (2014, 2017), Cremer et al. (2016), De Donder and Pestieau (2017) and Klimaviciute et al. (2019) study the effect of altruism and family help in long-term care (LTC)[5] and LTC insurance. They consider the interaction between parents and siblings, where the former may become dependent on care in their old age, while the latter can decide either to care for their parents directly, pay for care, buy insurance or not collaborate at all. Most of these studies assume that parents are pure altruists towards their children, while children display either impure altruism or care primarily about the bequest left to them by the parents. Moreover, it is also usually assumed that parents and children obtain an exogenous income stream, and that parents face an exogenous probability of becoming dependent, assumptions I relax in my model.[6]

[4]I will explore this point in more detail in a later section.

[5]See Cremer (2014) and Einav and Finkelstein (2018) for a more detailed discussion on long-term care and LTC insurance.

[6]There are two more strands of literature that are related to this paper. The first one is the literature on moral hazard in teams, firstly analyzed by Alchian and Demsetz (1972) and Holmström (1982) and more recently by Che and Yoo (2001). The second one is the literature on *nonexclusive contracting* analyzed by Attar et al. (2011, 2014, 2017) in the adverse selection case, and by Bisin and Guaitoli (2004) and Attar

2 The Model

I consider an interaction between a principal (or insurer) and two agents over three periods. First, an insurance contract is proposed by the insurer, and each agent decides whether to buy or to reject the contract. After observing each other's decisions, agents simultaneously and noncooperatively choose how much self-protection effort they will exert to prevent a loss from happening. Losses are realized, wealth levels are observed and contractual terms are executed, thus prompting each agent to decide whether to make a cross-insure transfer to the other and, if so, how much. The timing is synthesized below.

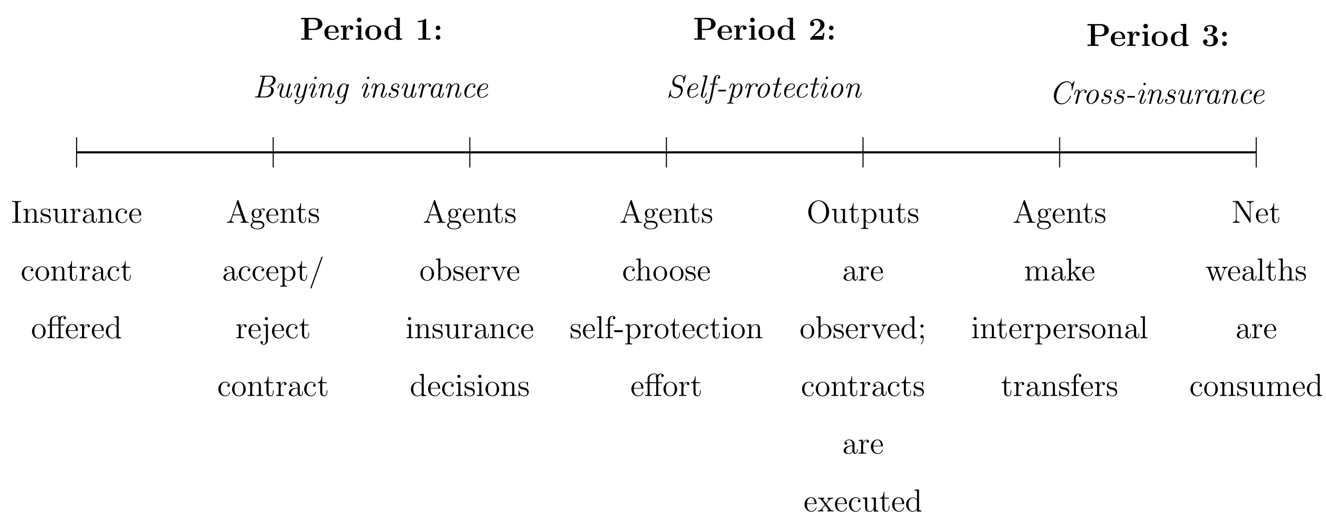


Figure 2: Timing of the game.

I must stress that the interaction between the two agents constitutes a game of complete information: at each moment they are called to take an action, they know (*i*) if the other has bought insurance or not; and (*ii*) the realized output of the other. This assumption is in line with the observation in the informal risk-sharing literature that agents mostly engage in trades with close friends and families, even in small villages (Udry, 1990, 1994; Townsend,

and Chassagnon (2009) under moral hazard. Indeed, one can take the view that in my model an agent will choose to buy the market insurance to *complement* the risk-sharing arrangement he already has with his partner, hence the link with nonexclusive contracting.

1994), and, therefore, that information asymmetries are small between agents.

On the other hand, I will analyze the firm's contract offers under two polar assumptions. First, I will assume that the insurer cannot observe or contract upon the transfers between the agents, but she observes the realized outputs at the end of period 2. This is a natural assumption, given that insurance companies typically cannot observe informal arrangements among potential insurees. Second, I assume the firm anticipates the cross-insurance transfers between agents, and offer contracts that internalize the substitution and free-riding effects generated by such transfers.

The analysis is divided into three parts. The first part studies the demand for insurance when the agents cannot self-protect, while the second part of the analysis considers insurance demand under self-protection. The third and last part of the analysis focuses on the insurance policies offered by the principal given the agents' demand. To introduce the notation, the next subsection presents the model with a single agent.

2.1 Benchmark: A Single Agent

An agent faces a loss with probability $(1-p) \in [0, 1]$. The wealth of this agent is either high, w^H, when he suffers no losses, or low, $w^L < w^H$, when losses take place. Let $L = w^H - w^L > 0$ denote the agent's loss.

Let $u(w)$ denote the utility of consuming wealth $w \geq 0$ and $\psi(p)$ the disutility of choosing the probability of not suffering a loss p. I assume that both u and ψ are twice continuously differentiable, with $u' > 0 > u''$, $\psi' \geq 0$, $\psi'' > 0$, and $\psi'(0) = 0$. The expected payoff of an agent under autarky is

$$U(p) = pu(w^H) + (1-p)u(w^L) - \psi(p). \tag{1}$$

The optimal choice of effort for the agent is given by the first-order condition

$$\psi'(p^{Aut}) = u(w^H) - u(w^L) > 0. \tag{2}$$

As expected, the higher the loss $L = w^H - w^L$ the agent can potentially suffer, the higher his effort to prevent it from happening will be. Let

$$U^{Aut} \equiv p^{Aut}u(w^H) + (1 - p^{Aut})u(w^L) - \psi(p^{Aut}) \tag{3}$$

denote the agent's expected utility in autarky.

The agent can buy an insurance policy $C = (q, t)$, where q denotes the coverage of the policy, while t is the insurance premium, from the set of contracts

$$\mathcal{C} = \{(q, t) : 0 \leq t \leq q \leq L\}. \tag{4}$$

Notice that I impose two conditions on the set contracts. The first condition is that the premium of an insurance policy cannot exceed the coverage. Indeed, any policy with a premium larger than the coverage would be rejected by the agent, since it would reduce his wealth after any realized output. The second condition is that the coverage does not exceed the loss the agent is subject to. This inequality can be justified by the fact that an insurance company would not offer a coverage larger than the loss, since such a contract would always incur a loss[7].

If the agent buys the insurance policy $C \in \mathcal{C}$, his expected utility becomes

$$U(p; C) = pu(w^H - t) + (1 - p)u(w^L - t + q) - \psi(p). \tag{5}$$

The agent's rejection of such a contract is equivalent to trading the null contract $C_0 = (0, 0)$, which makes (5) identical to the expected utility of the agent under autarky (1), and thus leads to the same choice of effort as when no transaction between insurer and agent takes place. The agent's choice of effort for $C \neq (0, 0)$ is given by

$$\psi'(p^{MH}) = u(w^H - t) - u(w^L - t + q), \tag{6}$$

where the right-hand side of (6) is positive for any $q \leq L$, and strictly positive for $q < L$.

[7]Such restrictions may be relaxed if one considers an environment where the government subsidizes insurance policies. Since my model seeks to explain the emergence of formal insurance companies, these assumptions improve the likelihood of a firm making a positive profit, and therefore they fit the model well.

As can be seen from comparing equations (2) and (6), the equilibrium effort made by the agent when accepting the insurance policy is smaller than the one he would exert in autarky. This is a consequence of the agent's risk aversion together with the fact that the contract reduces the overall risk he faces[8].

In Appendix B, I derive the contract that maximizes the principal's expected profit when dealing with this single agent. Such contract is characterized by partial coverage that induces the agent to exert positive effort to avoid the loss, and by a higher than actuarially fair premium.

2.2 My Setup: Two Altruistic Agents

I now turn to the model studied hereafter. There are two agents, 1 and 2, independently facing a loss $L_i = w_i^H - w_i^L$ with probability $(1 - p_i) \in [0, 1]$. The agents are altruistic towards each other, such that if only one of them suffers a loss, the rich agent is inclined to transfer part of his wealth to the poor agent. These transfers between individuals can be thought of as an informal risk-sharing device complementing the contracts $C_i = (q_i, t_i)$, for $i = 1, 2$.

To formalize this notion, let $d_i \in \{a, r\}$ be agent i's decision to buy (a) or not (r) the contract $C_i = (q_i, t_i)$, while $\omega = (w_1, w_2) \in \Omega = \{w_1^H, w_1^L\} \times \{w_2^H, w_2^L\}$ denotes the realized outputs. Let $y_i(w_i, d_i, C_i)$ denote the pretransfer wealth available to agent i, i.e.

$$y_i(w_i, d_i; C_i) = \begin{cases} w_i^H - t_i & \text{if } w_i = w_i^H \text{ and } d_i = a, \\ w_i^L - t_i + q_i & \text{if } w_i = w_i^L \text{ and } d_i = a, \\ w_i^H & \text{if } w_i = w_i^H \text{ and } d_i = r, \\ w_i^L & \text{if } w_i = w_i^L \text{ and } d_i = r. \end{cases} \tag{7}$$

The wealth consumed by agent i at the end of the game is equal to the pretransfer wealth

[8]Formally, one has that $\psi'(p^{Aut}) = u(w^H) - u(w^L) \geq u(w^H - t) - u(w^L - t + q) = \psi'(p^{MH})$ due to concavity of u and $0 \leq t \leq q \leq L$. Then, since $\psi(\cdot)$ is a convex function, I conclude that $p^{Aut} \geq p^{MH}$ for any $(q, t) \in \mathcal{C}$.

$y_i(\cdot)$ minus any transfer he makes to his pair, denoted by τ_i, plus any transfer τ_j received from his pair.

Define a strategy for agent i in the three-stage game as the triple $s_i = (d_i, p_i, T_i)$, where $d_i : \mathbf{C} \to \{a, r\}$, $p_i : \mathbf{C} \times \mathbf{d} \to [0, 1]$ and $T_i : \mathbf{C} \times \mathbf{d} \times [0, 1] \to \mathbb{R}_+$. Each strategy profile $\mathbf{s} = (s_A, s_B)$ determines the utility to each agent i and output ω, conditional on contracts $\mathbf{C} = (C_1, C_2)$:

$$U_i(\mathbf{s}, \mathbf{C}) = V_i(\mathbf{s}, \mathbf{C}) + \alpha_i V_j(\mathbf{s}, \mathbf{C}), \tag{8}$$

where $j \neq i$ and V_i denotes agent i's expected material payoff

$$V_i(\mathbf{s}, \mathbf{C}) = \mathbb{E}_{\mathbf{p}} \left[u(y_i(\cdot) - T_i(\cdot) + T_j(\cdot)) | p_i, p_j \right] - \psi(p_i(\cdot)), \tag{9}$$

and $\alpha_i \in [0, 1]$ represents i's degree of altruism towards his partner.[9,10] In the next two sections I will characterize subgame perfect equilibria of this game, assuming first in Section 3 that the probability of obtaining the high output, p_i, is fixed, and then relaxing this assumption in Section 4.

3 Insurance Demand Without Self-Protection

The analysis in this section is divided in two parts, following backwards induction analysis of the game absent self-protection. The timing is presented in Figure 3.

The first part focuses on cross-insurance, representing the last stage of the game (Period 2). I derive the conditions under which cross-insurance transfers take place and show that cross-insurance is independent of the probabilities of suffering a loss and in the receiver's

[9]If $\alpha_i = 0$, agent i is said to be selfish, while for $\alpha_i = 1$ the agent is called fully altruistic.

[10]Alternatively, one can think of the agents having preferences that take into consideration the internalization of the collective benefits generated by their actions. Indeed, let x_i and x_j denote the strategies for two different agents, and suppose that agent i's utility is given by $U_i(x_i, x_j, \alpha_i) = (1 - \alpha_i)\pi_i(x_i, x_j) + \alpha_i[\pi_i(x_i, x_j) + \pi_j(x_j, x_i)]$, where $\pi_i(x, y)$ denotes the material payoff of the game played between agents i and j. This specification is also aligned with Bergström (1995), where $U_i = U(\pi, \alpha_i, U_j) = \pi(x_i, x_j) + \alpha_i U_j$ when $\alpha_i \alpha_j < 1$.

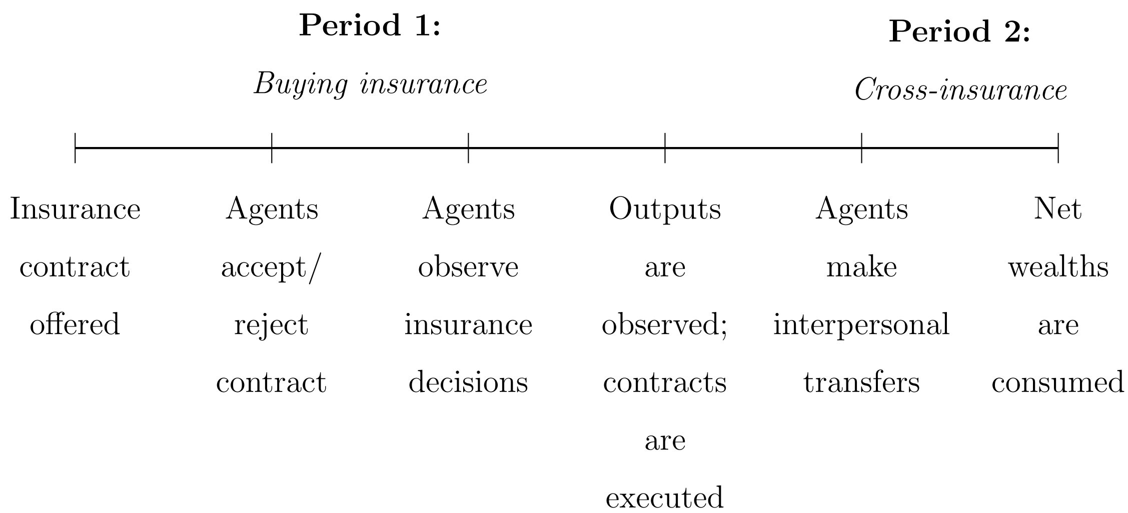


Figure 3: Timing of the game without self-protection.

degree of altruism. The second part focuses on the agents' insurance demand (Period 1), given equilibrium transfers functions. I show that an actuarially fair full coverage insurance contract is crowded out when one agent is very altruistic and the other is not. Moreover, if only one agent buys the insurance policy, he is not willing to buy a coverage larger than his loss at the actuarially fair price to cover for his partner's lack of insurance coverage.

3.1 Cross-Insurance for Given Contracts (C_1, C_2)

At the beginning of the last stage of the game without self-protection, agents 1 and 2 are aware of each other's decision to purchase or not their insurance policies C_1 and C_2, as well as the realized outputs $\omega = (w_1, w_2)$. For any pair of pre-transfer wealths (y_1, y_2) defined in (7), and given τ_j, agent i chooses to transfer τ_i to agent j to maximize

$$u(y_i - \tau_i + \tau_j) + \alpha_i u(y_j - \tau_j + \tau_i). \tag{10}$$

Suppose that agent i believes his pair will give him zero transfers, i.e. $\tau_j = 0$. Then, i would choose τ_i to maximize

$$u(y_i - \tau_i) + \alpha_i u(y_j + \tau_i), \tag{11}$$

leading to the first-order condition

$$u'(y_i - \tau_i) = \alpha_i u'(y_j + \tau_i). \tag{12}$$

Consider what happens when the left-hand side of (12) is smaller than the right-hand side: the marginal utility of agent i's own wealth is smaller than the marginal utility of his pair j, weighted by the degree of altruism α_i. To equalize both, i's wealth should be reduced while j's wealth should increase, which is achieved by a positive transfer τ_i from i to j. On the other hand, if the left-hand side of (12) is greater than the right-hand side, agent i making a positive transfer will only further drive the difference between the two marginal utilities, so he must choose $\tau_i = 0$ in this case.

This argument also implies a minimum degree of altruism for which agent i makes a positive transfer. Indeed, consider the pre-transfer wealths y_i and y_j, and let $\widehat{\alpha}_i$ be such that $u'(y_i) = \widehat{\alpha}_i u'(y_j)$. Then, no transfers are made by agent i, since the weighted marginal utilities of wealth are equalized. If, however, i's degree of altruism increases to some $\alpha_i > \widehat{\alpha}_i$, then $u'(y_i) < \alpha_i u'(y_j)$, and agent i would be inclined to make a positive transfer to equalize marginal utilities.[11] Therefore, I can conclude that, for any given vector of pretransfer wealths (y_i, y_j), agent i will make a positive transfer to agent j only if α_i is greater than

$$\widehat{\alpha}_i(y_i, y_j) \equiv \frac{u'(y_i)}{u'(y_j)}. \tag{13}$$

There are two important remarks to be made about these transfers. The first one is that they never flow from a poorer agent to a richer one. Indeed, suppose that $y_j > y_i$; then, because $u(\cdot)$ is strictly increasing and strictly concave, $u'(y_j) < u'(y_i)$ which in turns implies that (13) is greater than 1. Therefore, since $\alpha_i \in [0, 1]$ by assumption, agent i does not transfer a positive amount to his richer partner j. On the other hand, when $y_i > y_j$, $\widehat{\alpha}_i(y_i, y_j) < 1$ and positive transfers thus take place for sufficiently high degrees of altruism. If $y_i = y_j$, then $\widehat{\alpha}_i(y_i, y_j) = 1$ and (12) implies that $\tau_i = 0$[12].

[11] Appendix A computes the degree of risk-aversion for altruistic agents. Interestingly, I show that altruism decreases an agent's risk-aversion for fixed wealth levels.

[12] Alternatively, note that $\widehat{\alpha}_j(y_i, y_j) \equiv \frac{u'(y_j)}{u'(y_i)} = \frac{1}{\widehat{\alpha}_i(y_i, y_j)}$. Therefore, for any (y_i, y_j) such that $\widehat{\alpha}_i \leq 1$, it must be that $\widehat{\alpha}_j \geq 1$, thus implying that $T_j(\cdot) = 0$.

The second remark is that each agent's equilibrium transfer depends neither on the probabilities with which each agent faces a loss, nor on the other agent's degree of altruism. This happens because each individual is choosing how much to transfer only after any uncertainty about outputs has been realized, and noting that equation (12) does not depend on the recipient's degree of altruism.

Proposition 1: *For each (y_1, y_2), there exists at least one Nash equilibrium of the transfers subgame. If $\alpha_1\alpha_2 < 1$, then this equilibrium is unique, and at most one agent makes a transfer, which is never made from the poorer to the richer agent, and does not depend on the poorer agent's degree of altruism. If $\alpha_1\alpha_2 = 1$, there is a continuum of Nash equilibria, but a unique equilibrium value for final incomes, namely $\frac{y_1+y_2}{2}$.*

In the following lemma, I further show that an agent's transfer increases in the agent's degree of altruism and in the difference between pretransfer wealths of the rich and poor agents. Naturally, as one individual becomes more altruistic, he is more inclined to help an agent with a lower wealth, even if the latter agent did not suffer a loss. If such an individual values his partner's material payoff as much as he values his own, then transfers are chosen to equalize final incomes. Additionally, for any given degree of altruism, a higher gap in outputs requires a larger transfer to satisfy the weighted marginal utilities condition in equation (12).

Two other comparative statics results must be made with respect to the equilibrium transfers defined in (12), related to the terms of the insurance policies C_1 and C_2. In a broad sense, any terms of the policies that make agent i richer than agent j for a given output increases the transfers from the former to the latter, while the reverse holds if j becomes relatively richer. Thus, agent i will be aiding agent j less if i has to bear a larger premium in his contract or if his coverage decreases, while the opposite happens when it is j's contract that is subject to a raise in premium or diminishing coverage. These results are summarized in the following lemma.

Lemma 1: *The equilibrium transfer function $T_i : \mathbf{C} \times \mathbf{d} \times [0, 1] \to \mathbb{R}_+$ is continuous, positive* *if $\alpha_i > \widehat{\alpha}_i(\cdot)$, and zero otherwise. Moreover, T_i is differentiable for all $\alpha_i \neq \widehat{\alpha}_i(\cdot)$, increasing in α_i and y_i, and decreasing in y_j.*

The most important point to be made about equilibrium transfers is that they locally alter the individuals' behavior towards risk. Indeed, this can be seen in Figure 4, where the dashed line represents an individual's utility absent any transfers, while the solid line represents the same agent's utility for different levels of wealth, holding constant the other agent's wealth and the degrees of altruism. When one agent is making the transfers, his disposable wealth is reduced in comparison to his pre-transfer wealth, thus resulting in a locally more risk-averse behavior around the threshold $y_H(\alpha_1)$. On the other hand, when an agent is the one receiving a transfers, his disposable wealth is larger than the pre-transfer one, and locally said agent becomes *risk-loving* around $y_L(\alpha_2)$. This induced change in the individuals' risk behavior together with the substitution effect generated by the presence of cross-insurance transfers will determine the purchasing decision of insurance contracts, as I will discuss in more detail in the next section.

3.2 Insurance Contract Purchasing Decision

Given the equilibrium transfers characterized in Proposition 1 and Lemma 1 above, I can proceed by backwards induction in the game played between agents 1 and 2 absent moral-hazard. Since I am now focusing on the case where the agents cannot self-protect (i.e., $\mathbf{p} = (p_1, p_2)$ is exogenously given), I can compute each agent's expected utility $U_i(\mathbf{d})$ for arbitrary tuples of contracts, outputs, probabilities of suffering losses and degrees of altruism. Therefore, the agents' decisions to buy or not buy the insurance policy being offered to them can be summarized in the normal form game $\Gamma(\mathbf{C}, \mathbf{p}, \alpha)$, represented by the payoff matrix

I am now in a position to state my first main result. Absent any transfers, efficiency would require a full coverage actuarially fair insurance policy $C_i^{af} = (L_i, (1 - p_i)L_i)$ for each agent $i = 1, 2$. Such a contract would eliminate the risk faced by an individual agent by equalizing all outputs to the expected wealth $\overline{w}_i = p_i w_i^H + (1 - p_i) w_i^L$ and, thus, would be strictly

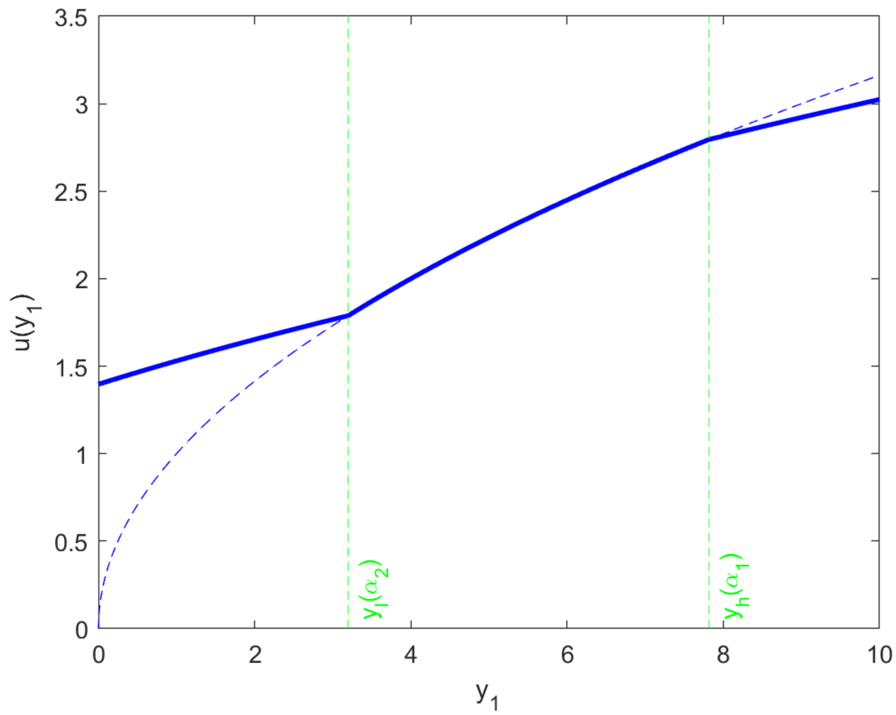


Figure 4: Material payoff under equilibrium transfers.

		Agent 2	
		a	r
Agent 1	a	$U_1(a,a), U_2(a,a)$	$U_1(a,r), U_2(a,r)$
	r	$U_1(r,a), U_2(r,a)$	$U_1(r,r), U_2(r,r)$

preferred by the risk-averse agents over no insurance. The introduction of cross-insurance transfers, however, may change this result, as shown next.

Proposition 2: *Suppose that* $(w_1^H, w_1^L, p_1) = (w_2^H, w_2^L, p_2)$, $\alpha_1 = 1$, *and* α_2 *is small enough. Then, buying the full coverage actuarially fair insurance policy* $C^{af} = (L, (1-p)L)$ *is a strictly dominant strategy for agent* 1. *Agent* 2 *will prefer not to buy* C^{af} *if he is not too risk averse.*

The main insight here is that a purely selfish, risk-averse, and rational agent may reject an actuarially fair insurance contract that eliminates all the risk he faces in favor of a small, but free of charge, help from his altruistic partner.

The intuition behind Proposition 2 is that rejecting the actuarially fair full coverage insurance policy yields a lottery of payoffs to the altruistic agent 1 that is a mean-preserving spread of the lottery induced by purchasing such a contract, irrespective of the decision made by the self-interested agent 2, since he receives no transfers from agent 2. Then, because agents are assumed to be risk averse with respect to wealth, I obtain that agent 1 strictly prefers to insure himself.

The same, however, is not true for agent 2 since, given that agent 1 buys the insurance policy, the expected wealth he obtains when rejecting C^{af} given that 1 buys the policy is higher than the expected wealth he would obtain when purchasing C^{af}. Therefore, a second-order stochastic dominance argument cannot be applied here, and for some parameter configurations the best response of agent 2 to $d_1 = a$ might be either to accept or to reject C^{af}.

Proposition 2 implies that altruism may lead to an inefficient outcome due to the substitution effect. In particular, one agent's altruism may induce the other to free-ride on cross-insurance transfers and not to purchase an insurance policy that would eliminate all the risk faced by the agents.

Example: *I will use the following example to illustrate my results. Suppose that $u(w) = \sqrt{w}$ and that both agents have identical wealth and suffer a loss with the same probability. More precisely, let $w_1^H = w_2^H = 3$, $w_1^L = w_2^L = 1$ and $p_1 = p_2 = 0.6$. In this case, the full coverage actuarially fair insurance policy is identical for both agents and is given by $C^{af} = (2, 0.8)$. Figure 5 represents the equilibrium decisions for agents 1 and 2 to accept or reject the policy C^{af} for different values of the degrees of altruism $(\alpha_1, \alpha_2) \in [0,1]^2$. Notice first that, for low degrees of altruism, both agents will buy the insurance policy. This is not surprising, since no transfers are made when agents care very little about each other and must therefore rely* *solely on the insurance contract to protect themselves against a loss.*

If both agents display high degrees of altruism, buying insurance is the unique equilibrium strategy for both of them: recall that equilibrium transfers cannot completely eliminate the risk the agent faces, and thus the improved risk-sharing offered by the insurance policy is the most preferred alternative to the agents when both display high degrees of altruism. These two cases are captured in the central area in Figure 5.

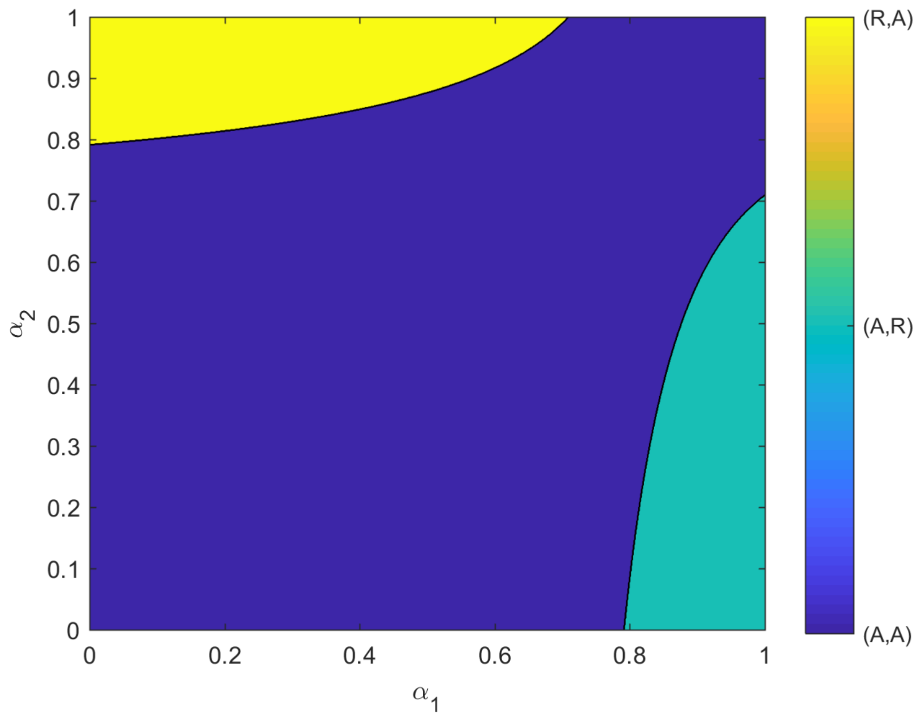


Figure 5: Equilibrium demand for insurance.

Figure 5 also exhibits two regions (the top left and bottom right corners, respectively) with asymmetric equilibria, in which only one agent accepts the insurance offer. As pointed out in Proposition 2*, the self-interested agent rejects his insurance policy and free-rides on the transfers provided by his altruistic partner to dissipate risk. The altruistic agent, on the other hand, always prefers to buy the insurance policy both to protect himself against a loss and to help the other agent in case of need.*

Meanwhile, Figure 6 exhibits insurance demand behavior for different values of p, and associated actuarially fair full coverage contracts $C^{af}(p)$. The top left panel has the highest probability of suffering a loss, namely $1 - p = 0.75$, while the bottom right panel has the smallest probability of suffering a loss at $1 - p = 0.1$. As the probability of suffering a loss decreases, the gains of a deviation from (a, a) for a self-interested individual becomes larger, and so do the regions associated with asymmetric equilibria (a, r) and (r, a). ■

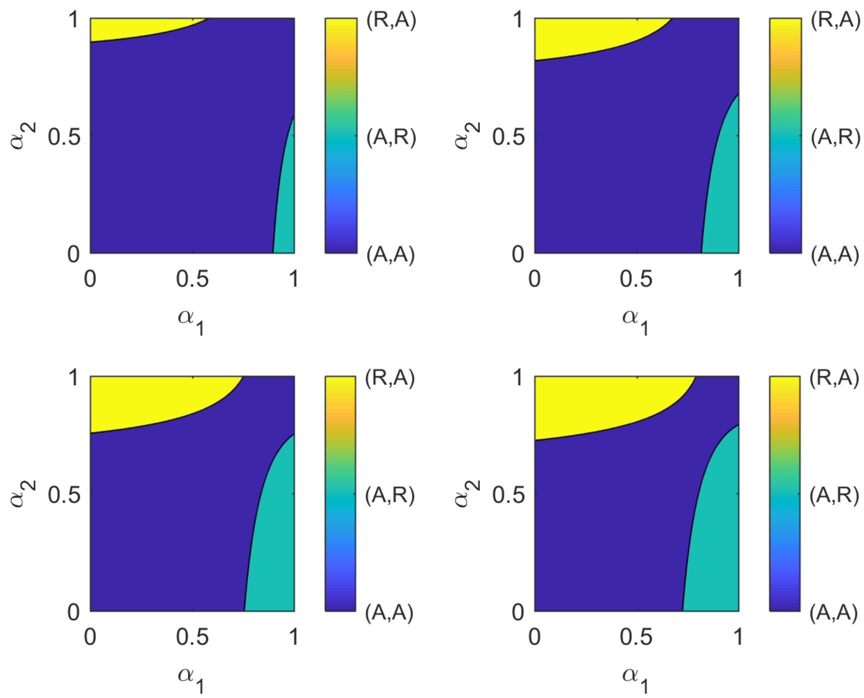


Figure 6: Equilibrium demand for insurance without self-protection, for $p \in \{0.25, 0.5, 0.75, 0.9\}$.

One interesting question that arises from Proposition 2 and Figures 5 and 6 is the following: knowing that his partner will be uninsured, would the altruistic agent be willing to buy an actuarially fair insurance policy with a coverage larger than his own loss in order to protect both himself and his partner? Equilibrium transfers imply that the altruistic agent will make a positive transfer whenever his partner suffers a loss if the latter is uninsured, and when his partner suffers a loss but he himself doesn't if both are insured. In particular for the former case, the altruistic agent could potentially buy larger coverage to insure both himself and his partner in the case both suffer a loss. Proposition 3 below shows this is not the case.

Proposition 3: *Suppose that* $(w_1^H, w_1^L, p_1) = (w_2^H, w_2^L, p_2)$ *and* $(\alpha_1, \alpha_2) = (1, 0)$*. If* $(d_1, d_2) = (a, \cdot)$ *and agent* 1 *can buy any actuarially fair insurance policy* $C = (q, (1-p)q)$*, he buys the full coverage* ($q = L$) *policy.*

The intuition is simple: although both agents would benefit from sharing a larger coverage if both suffer a loss, the increased premium associated with such a coverage would ultimately reduced the altruistic agent's utility. On the other hand, a smaller coverage at an actuarially fair price would increase the risk faced by the altruistic agent, thus reducing expected utility due to risk aversion.

4 Insurance Demand with Self-Protection

Let us focus now on the more general game between the agents, first presented in Section 2 and reproduced in Figure 7 below, where the agents choose how much effort they will exert to avoid a loss from happening (Period 2). For any output vector $\mathbf{w}$ and contracts $\mathbf{C}$, agents 1 and 2 play a three-stage sequential game, where they must first choose whether to buy insurance, then choose the effort to avoid the loss, and finally choose how much to transfer to one another. In the previous section, I have shown that one agent can free-ride on another's cross-insurance transfers and not buy an insurance policy due to the substitution effect generated by cross-insurance transfers. In this section, I will focus on the free-riding and empathy effects, namely the effects that altruism has on self-protection effort through equilibrium transfers.

In the third period of the game described in Figure 7, when agents choose equilibrium transfers, the disposable wealth is the same as in equation (7), the only difference being

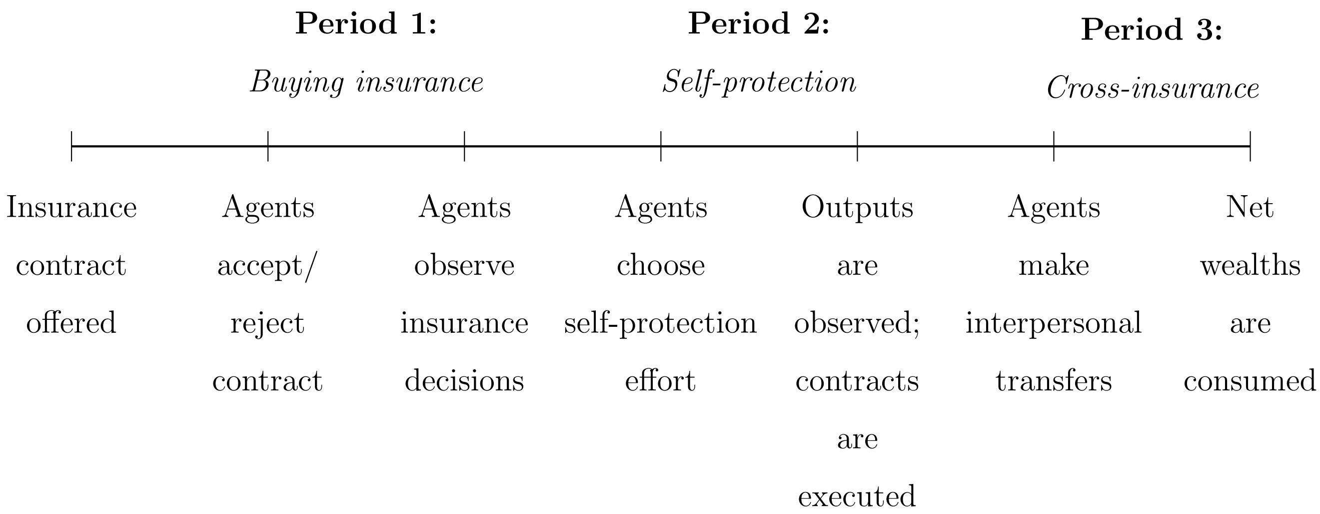


Figure 7: Timing of the game.

that the probabilities with which each output is realized are now endogenous. Therefore, equilibrium transfers are given by equation (12) exactly as in the benchmark case with exogenous probabilities, and the results in Proposition 1 and Lemma 1 still hold.

Given equilibrium transfers, each agent must now choose his level of effort to maximize expected utility, i.e. each agent $i = 1, 2$ now solves[13]

$$
\begin{aligned}
&\max_{p_i} U_i(d_i, d_j, p_i, p_j, T_i, T_j; C_i, C_j, \alpha_i, \alpha_j, \mathbf{w}) \\
&\qquad = p_i p_j \left[u(w_i^H - t_i - T_i + T_j) + \alpha_i u(w_j^H - t_j + T_i - T_j) \right] \\
&\qquad\quad + p_i(1 - p_j) \left[u(w_i^H - t_i - T_i + T_j) + \alpha_i u(w_j^L - t_j + q_j + T_i - T_j) \right] \\
&\qquad\quad + (1 - p_i) p_j \left[u(w_i^L - t_i + q_i - T_i + T_j) + \alpha_i u(w_j^H - t_j + T_i - T_j) \right] \\
&\qquad\quad + (1 - p_i)(1 - p_j) \left[u(w_i^L - t_i + q_i - T_i + T_j) + \alpha_i u(w_j^L - t_j + q_j + T_i - T_j) \right] \\
&\qquad\quad - \psi(p_i) - \alpha_i \psi(p_j). \qquad (14)
\end{aligned}
$$

Equilibrium effort for agents 1 and 2 are given as the solution to the system of first-order

[13]For ease of exposition, I omit the dependence of T_i on $\mathbf{d}, \mathbf{C}, \omega, \alpha_1, \alpha_2$. Also, if $d_i = r$ for some $i = 1, 2$, setting $t_i = q_i = 0$ appropriately adjusts equation (14).

conditions

$$\psi'(p_1) = a_1(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_1) + p_2\left[b_1(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_1) - a_1(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_1)\right] \tag{15}$$

$$\psi'(p_2) = a_2(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_2) + p_1\left[b_2(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_2) - a_2(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_2)\right], \tag{16}$$

where the constants $a_i(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_i)$ and $b_i(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_i)$ are given, respectively, by

$$\begin{aligned} a_i(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_i) = {} & u(y_i(w_i^H, C_i, d_i) - T_i((w_i^H, w_j^L), \mathbf{C}, \mathbf{d}) + T_j((w_i^H, w_j^L), \mathbf{C}, \mathbf{d})) \\ & - u(y_i(w_i^L, C_i, d_i) - T_i((w_i^L, w_j^L), \mathbf{C}, \mathbf{d}) + T_j((w_i^L, w_j^L), \mathbf{C}, \mathbf{d})) \\ & + \alpha_i \big[u(y_j(w_j^L, C_j, d_j) + T_i((w_i^H, w_j^L), \mathbf{C}, \mathbf{d}) - T_j((w_i^H, w_j^L), \mathbf{C}, \mathbf{d})) \\ & -u(y_j(w_j^L, C_j, d_j) + T_i((w_i^L, w_j^L), \mathbf{C}, \mathbf{d}) - T_j((w_i^L, w_j^L), \mathbf{C}, \mathbf{d}))\big] \end{aligned} \tag{17}$$

and

$$\begin{aligned} b_i(\mathbf{w}, \mathbf{C}, \mathbf{d}, \alpha_i) = {} & u(y_i(w_i^H, C_i, d_i) - T_i((w_i^H, w_j^H), \mathbf{C}, \mathbf{d}) + T_j((w_i^H, w_j^H), \mathbf{C}, \mathbf{d})) \\ & - u(y_i(w_i^L, C_i, d_i) - T_i((w_i^L, w_j^H), \mathbf{C}, \mathbf{d}) + T_j((w_i^L, w_j^H), \mathbf{C}, \mathbf{d})) \\ & + \alpha_i \big[u(y_j(w_j^H, C_j, d_j) + T_i((w_i^H, w_j^H), \mathbf{C}, \mathbf{d}) - T_j((w_i^H, w_j^H), \mathbf{C}, \mathbf{d})) \\ & -u(y_j(w_j^H, C_j, d_j) + T_i((w_i^L, w_j^H), \mathbf{C}, \mathbf{d}) - T_j((w_i^L, w_j^H), \mathbf{C}, \mathbf{d}))\big]. \end{aligned} \tag{18}$$

I will focus on the case where the agents are symmetric in terms of wealth and altruism, that is, $(w_1^H, w_1^L, \alpha_1) = (w_2^H, w_2^L, \alpha_2)$. Moreover, because agents are identical both in their preferences and in the risk they face, I will assume that a single insurance policy $C = (q, t) \in \mathcal{C}$ is offered to both individuals.

Suppose that both agents accept the policy C. The first consequence of such assumptions is that, for any $\alpha < 1$, equilibrium transfers only take place when one agent suffers a loss and the other does not, and it is the last one who makes a positive transfer if the degree of altruism is sufficiently high. Formally, for $\alpha > \widehat{\alpha}(C) \equiv \frac{u'(w^H - t)}{u'(w^L - t + q)}$, $T(\alpha, C) > 0$ is given by the equilibrium condition

$$u'(w^H - t - T(\alpha, C)) = \alpha u'(w^L - t + q + T(\alpha, C)). \tag{19}$$

Each agent chooses his individual success probability to maximize his *ex ante* expected utility

$$\begin{aligned}
U_i(p_i, p_j) = {} & p_i p_j (1+\alpha) u(w^H - t) \\
& + (1-p_i)(1-p_j)(1+\alpha) u(w^L - t + q) \\
& + p_i(1-p_j)[u(w^H - t - T(\alpha, C)) + \alpha u(w^L - t + q + T(\alpha, C))] \\
& + (1-p_i) p_j [u(w^L - t + q + T(\alpha, C)) + \alpha u(w^H - t - T(\alpha, C))] \\
& - \psi(p_i) - \alpha \psi(p_j)
\end{aligned} \tag{20}$$

for $i, j = 1, 2$ and $j \neq i$, and thus, a necessary and sufficient condition[14] for the pair $(p_1, p_2) \in (0,1)^2$ to be a Nash equilibrium of $G^*(C)$ is that each of them satisfy the first-order condition

$$\begin{aligned}
\psi'(p_i) = {} & u(w^H - t - T(\alpha, C)) - u(w^L - t + q) \\
& + \alpha[u(w^L - t + q + T(\alpha, C)) - u(w^L - t + q)] \\
& - p_j(1+\alpha)\big[(u(w^L - t + q + T(\alpha, C)) - u(w^L - t + q)) \\
& -(u(w^H - t) - u(w^H - t - T(\alpha, C)))\big]
\end{aligned} \tag{21}$$

for $j \neq i$.

Notice that the right-hand side of equation (21) is an affine function of p_j. For $\alpha \leq \widehat{\alpha}(C)$, $T(\alpha, C) = 0$ and the slope is equal to zero, while the intercept is $u(w^H - t) - u(w^L - t + q) \geq 0$ since $q \leq L$ and $u' > 0$ by assumption. This leads to exactly the same first-order condition that determined the self-protection of a single agent faced with an insurance policy C in equation (6). In other words, if altruism is not high enough to induce transfers between the agents, each will ignore the presence of the other and choose the level of effort that maximizes expected utility under the insurance contract $C = (q, t)$. Because $u'' < 0$ and $\alpha \in (\widehat{\alpha}(C, d), 1]$ by assumption, I find that $w^H - t - T(\alpha, C) \geq w^L - t + q + T(\alpha, C) > w^L - t + q$, and thus the slope is strictly negative. Furthermore, given the assumptions about the disutility of effort, the right-hand side of equation (21) is strictly increasing in p_i. These observations lead to the following result.

[14] The second-order condition is given by $-\psi''(p) < 0$ by assumption for all values of p.

Lemma 2: *If $\alpha_1 = \alpha_2 = \alpha$ and $d_1 = d_2 = d$, then $G^*(C)$ has a unique symmetric equilibrium (p^*, p^*). If $p^*(\alpha, C, d) > 0$, then it solves the equation*

$$\begin{aligned}\psi'(p^*) = {} & u(w^H - t - T(\alpha, C)) - u(w^L - t + q) \\ & + \alpha[u(w^L - t + q + T(\alpha, C)) - u(w^L - t + q)] \\ & - p^*(1+\alpha)\left[(u(w^L - t + q + T(\alpha, C)) - u(w^L - t + q)) \right. \\ & \left. -(u(w^H - t) - u(w^H - t - T(\alpha, C)))\right] \end{aligned} \tag{22}$$

if $d = a$ and

$$\begin{aligned}\psi'(p^*) = {} & u(w^H - T(\alpha, C)) - u(w^L) \\ & + \alpha[u(w^L + T(\alpha, C)) - u(w^L)] \\ & - p^*(1+\alpha)\left[(u(w^L + T(\alpha, C)) - u(w^L)) \right. \\ & \left. -(u(w^H) - u(w^H - T(\alpha, C)))\right] \end{aligned} \tag{23}$$

if $d = r$.

Contrary to equilibrium transfers, equilibrium efforts are not monotonic in the degree of altruism. Indeed, for low degrees of altruism ($\alpha \leq \widehat{\alpha}(C, d)$), agents cannot affect each other's material payoff because no transfers are made in equilibrium and the occurrence of a loss for one of them is independent of the other's choice of effort. Thus, agents 1 and 2 behave as if they were in an autarky relation with the insurer. In particular, the right-hand sides of equations (22) and (23) become identical to the right-hand sides of the first-order conditions for a single agent who buys (equation (6)) and who does not buy (equation (2)) the insurance contract C. Moreover, for any $C \in \mathcal{C}$, the equilibrium effort of the uninsured agents is greater than the equilibrium effort of the insured ones, as is the case when analyzing the single agent problem.

For degrees of altruism larger than but close to $\widehat{\alpha}(C, d)$, the positive transfers between agents reduce the expected loss they face, and thus a *free-riding* effect appears: agents reduce their equilibrium effort in the vicinity of $\widehat{\alpha}(C, d)$ due to the decrease in the risk each of them faces because of the equilibrium transfers. However, as α approaches 1, the problem faced by each agent in $G^*(C)$ becomes ever more similar to one that would be solved by a social planner seeking to maximize total utility, and thus the free-riding problem is mitigated by the empathy effect and a higher equilibrium effort is exerted.[15]

Proposition 4: *Consider the unique symmetric Nash equilibrium (p^*, p^*) of $G^*(C)$ when $d_1 = d_2$ and $\alpha_1 = \alpha_2 = \alpha$. If $p^*(\widehat{\alpha}(C), C, d) > 0$ and $p^*(1, C, d) > 0$, then there is an $\overline{\varepsilon}(C) > 0$ such that $p^*(\widehat{\alpha}(C) + \varepsilon, C, d) < p^*(\widehat{\alpha}(C), C, d)$ and $p^*(1 - \varepsilon, C, d) < p^*(1, C, d)$ for all $\varepsilon \in (0, \overline{\varepsilon}(C, d))$ for any $C \in \mathcal{C}$.*

Example (continued): *The results in Proposition 4 can be observed in Figure 8 for the quadratic cost function $\psi(p) = \frac{p^2}{2}$. Firstly, equilibrium effort when both agents reject an insurance contract is higher than the effort they would exert had the policy been accepted, due to the higher risk the agents are facing. For low degrees of altruism such that no transfers are made, equilibrium efforts are constant with respect to the common degree of altruism α, but become smaller as cross-insurance transfers become positive, reflecting the free-riding effect. As α increases, so does the empathy effect, and thus equilibrium effort also increases for sufficiently large values of α.* ■

After computing transfers and effort in the unique symmetric equilibrium of $G^*(C)$, the equilibrium expected material payoff of each agent is given by

$$\begin{aligned} V^*(\alpha, C) = {} & [p^*(\alpha, C)]^2 u(w^H - t) + [1 - p^*(\alpha, C)]^2 u(w^L - t + q) \\ & + p^*(\alpha, C)[1 - p^*(\alpha, C)][u(w^H - t - T(\alpha, C)) + u(w^L - t + q + T(\alpha, C))] \\ & - \psi(p^*(\alpha, C)), \end{aligned} \tag{24}$$

while symmetry implies that the utility function can be written as

$$U^*(\alpha, C) = (1 + \alpha) V^*(\alpha, C). \tag{25}$$

[15] For $\alpha = 1$, the agents fully internalize the effects of their choices on each other's payoffs, and therefore the free-riding problem disappears.

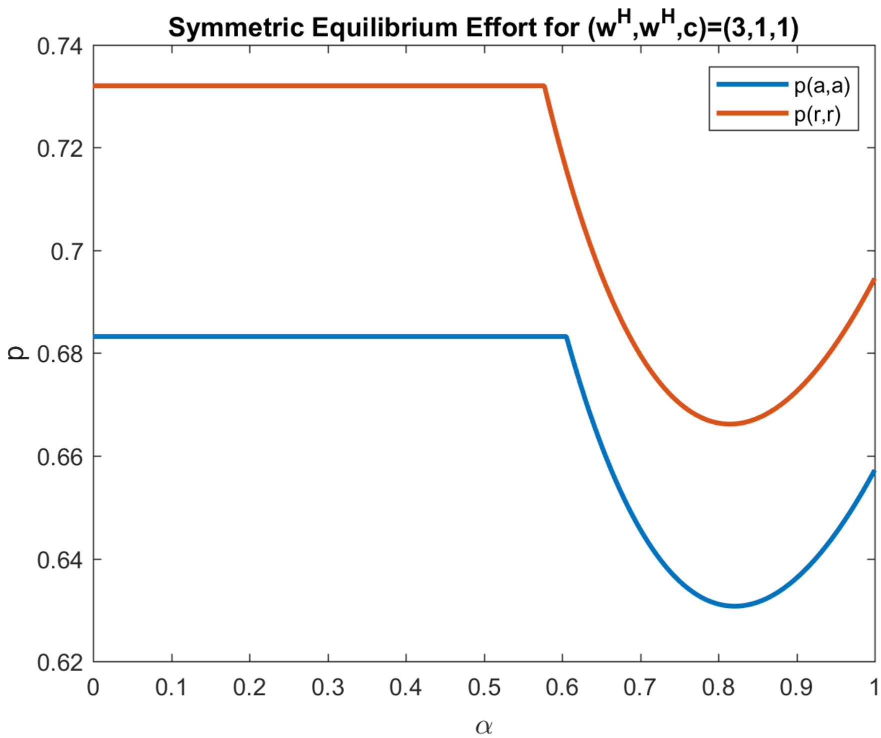


Figure 8: Symmetric equilibrium effort.

Similar expressions can be derived for the case where both agents reject the insurance policy.

Two important comparative statics results about the equilibrium expected material payoff (and consequently expected utility) can be made. First, the highest expected material payoff is reached at full altruism, that is, when $\alpha = 1$. In that case, each agent attaches the same weight to own and the other's material payoff, and therefore the free-riding effect is minimized. Secondly, the expected material payoff is increasing in the degree of altruism even in the region where equilibrium effort decreases with α. These statements are collected below.

Proposition 5: *Fix any insurance contract $C = (q, t) \in \mathbb{R}^2_+$, and take $d_1 = d_2$ and $\alpha_1 = \alpha_2 = \alpha$. Then,*

1. $V^*(1, C, d) \geq V^*(\alpha, C, d)$ *for all* $\alpha \in [0, 1]$*;*

2. *If* $p^*(\widehat{\alpha}(C), C, d) > 0$*, there is an* $\overline{\varepsilon}(C, d) > 0$ *such that* $V^*(\widehat{\alpha}(C)+\varepsilon, C, d) > V^*(\widehat{\alpha}(C), C, d)$ *for all* $\varepsilon \in (0, \overline{\varepsilon}(C, d))$.

So far, the analysis with self-protection has examined the symmetric insurance demand profiles (a, a) and (r, r). The equilibrium conditions for effort described in equations (15)-(18) suggest that even when the agents are identical, their choices of effort under asymmetric insurance purchase decisions are not trivial due to equilibrium transfers and their effects on effort.

Example (continued): *To illustrate this point, consider Figure 9, which depicts the equilibrium effort for agents* 1 *and* 2 *with symmetric outputs* $(w^H, w^L) = (3, 1)$ *and quadratic cost function* $\psi(p) = \frac{p^2}{2}$*. In this example, I assume that agent* 1 *buys an insurance policy* $C \in \mathcal{C}$*, while agent* 2 *remains uninsured. As was the case when the purchase decision was identical for the two agents, equilibrium efforts are constant when both agents display a low degree of altruism, since the absence of any transfers induces them to behave as in an autarky. Moreover, one can also observe that the equilibrium effort for the uninsured agent* $(p_2(a, r))$ *is larger than the effort for the insured agent* $(p_1(a, r))$ *for the same degrees of altruism.*

One interesting observation coming from inspection of Figure 9 is that, for a given degree of altruism for agent 2*, an increase in agent* 1*'s degree of altruism has opposite effects in equilibrium efforts:* p_1 *decreases, while* p_2 *increases. For agent* 1*, who purchases the insurance, the intuition is that he shirks in his self-protection to avoid making a large positive transfer to the uninsured agent and therefore mitigate any free-riding the latter would enjoy by not purchasing the policy* C*. On the other hand, agent* 2 *must engage in higher effort to avoid a loss due to correctly anticipating the lower probability of receiving a transfer from agent* 1*.*

Focusing on the cases where $\alpha_1 = \alpha_2$ *in both level plots of Figure 9 suggest that* $p_1(a, r)$ *and* $p_2(a, r)$ *will also exhibit the nonmonotonic behavior on a common degree of altruism* $\alpha \in [0, 1]$ *as the symmetric equilibrium efforts* $p^*(C, a)$ *and* $p^*(C, r)$ *derived in the previous*

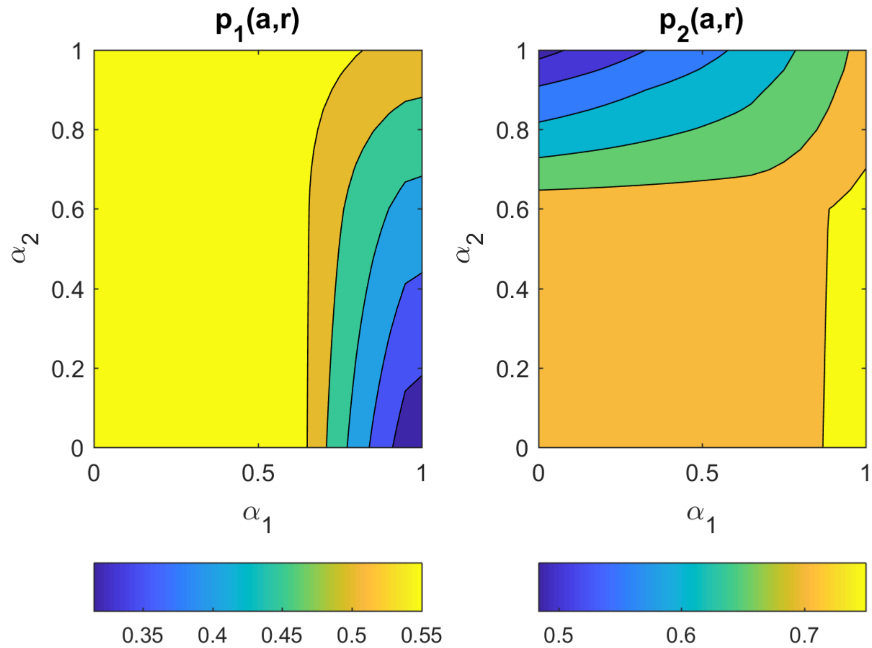


Figure 9: Equilibrium efforts for asymmetric purchase decisions $(d_1, d_2) = (a, r)$.

section. Such pattern is more evident for $p_2(a, r)$ *than for* $p_1(a, r)$. ■

As was the case without self-protection, for $\alpha_1\alpha_2 < 1$ equilibrium transfers are uniquely determined for any profiles of actions regarding the purchase of insurance and disposable wealth. Proceeding by backwards induction, suppose that the system of equations (15)-(16) admits a unique solution[16]. Then, the choice of buying insurance or not boils down to the normal form game $\widetilde{\Gamma}(\mathbf{C}, \mathbf{w}, \alpha)$, with associated payoff matrix

As before, each entry in the payoff matrix corresponds to an agent's expected utility following an action profile $(d_1, d_2) \in \{a, r\}^2$, with the difference that now the probabilities of suffering a loss for each agent are endogenous objects and the functions U_i also take into consideration the costs of effort for both agents whenever $\alpha_1\alpha_2 > 0$.

[16]Which can be guaranteed by an appropriate choice of the cost function of effort $\psi(p)$.

Agent 2

		a	r
Agent 1	a	$U_1(a,a), U_2(a,a)$	$U_1(a,r), U_2(a,r)$
	r	$U_1(r,a), U_2(r,a)$	$U_1(r,r), U_2(r,r)$

Example (continued:) *Figure 10 below extends the previous example with* $(w^H, w^L) = (3,1)$ *and* $u(w) = \sqrt{w}$ *by assuming the quadratic cost function* $\psi(p) = \frac{cp^2}{2}$ *for the contract* $C = (1.5, 0.85) \in \mathcal{C}$*, different from the actuarially fair full coverage one. For this particular parametrization, the introduction of self-protection effort does not qualitatively alter the results: a self-interested agent would free-ride on his altruistic partner's transfers and not purchase the insurance policy.*

However, increasing the cost parameter c reduces the regions under which an agent rejects the insurance policy. Moving from the top left panel to the bottom right, one can observe that the regions in which the asymmetric equilibria (a,r) *and* (r,a) *prevail diminish. This is due to the increase in the cost parameter associated with the increase in the equilibrium effort as seen in Figure 9. For sufficiently high values of c, a unique equilibrium exists for all values of* α_1 *and* α_2 *in this parametrization:* (a,a). ■

5 The Insurer's Contractual Offers

Until now, the insurance policy $C \in \mathcal{C}$ offered to the agents has been exogenous. Absent the possibility of self-protection, the natural candidate to consider was the actuarially fair full coverage insurance policy $C^{af} = (L, (1-p)L)$. Such a contract maximizes the agents' expected utility given a nonnegative profit condition for the principal. As I have shown, even such a contract is crowded out by informal risk-sharing when dispersion in the degrees of altruism is high.

On the other hand, when the agents can affect the probability of a loss taking place,

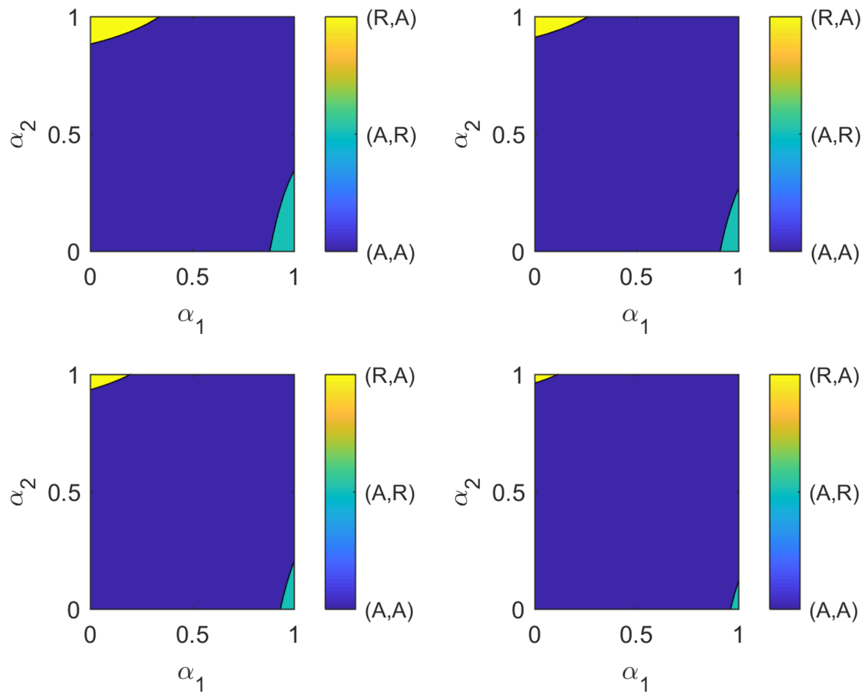


Figure 10: Equilibrium demand for insurance with self-protection for different $c \in \{0.98, 1, 1.02, 1.05\}$.

such a contract is no longer a natural candidate. The problem facing a monopolistic insurer, then, is to choose a menu of policies $(C_1, C_2) \in \mathcal{C}^2$ to be offered to the agents that maximizes the firm's expected profits given the equilibrium behavior it will induce in the insurance demand game played by agents 1 and 2. However, given transfers and self-protection effort, the insurance company could find it not profitable to offer any contract at all to the agents.[17]

I will divide the analysis of the offers made by the insurance firm in two parts, reflecting two polar assumptions about the information the firm has about the agents' preferences and, consequently, transfers.

First, I assume that the firm cannot observe the agents' degrees of altruism and thus offers contracts as if the agents are purely self-interested. This assumption reflects the ideas

[17] Alternatively, offer the null contract $C_0 = (0, 0)$ to both agents.

that there may be heterogeneity in altruism in the population, and this heterogeneity is unobserved, and that some agents do not have friends or family with whom they can engage in informal risk-sharing. In that case, each agent's equilibrium behavior is identical to the one an agent would have had he been alone. The next subsection thus argues that while such a contract would be accepted by both agents if altruism is low, insurance demand goes to zero when agents care sufficiently about one another.

Second, I consider the symmetric case where both agents share a common degree of altruism that is observed by the firm, capturing the idea that the insurance company faces a more homogenous population. Then, focusing on symmetric equilibria of the game played between agents, I show that equilibrium transfers and self-protection cannot crowd out formal insurance for any $\alpha \in [0, 1]$, but can substantially reduce the insurer's profits.

5.1 Demand for the Candidate Insurance Contract for the Naïve Insurer

In the symmetric case, under the assumption that both agents accept the insurance policy $C = (q, t)$, the introduction of moral hazard does not allow using the full coverage actuarially fair insurance policy $C^{af} = (L, (1-p)L)$ as the benchmark contract to analyze the agents' decisions to buy insurance or not. The reason for that is the lack of incentives to exert any effort when the agent has full coverage. Indeed, closer inspection of equation (19) shows that all transfers are equal to zero under full coverage, for any specified premium, and thus the right-hand side of first-order condition (22) also becomes zero, leading to no effort by the agents due to the assumption that $\psi'(0) = 0$. Finally, any firm offering such a contract would make losses equal to L per agent.

To circumvent this problem, I will consider a policy $C^B = (q^B, t^B)$ offered to agent i when the firm assumes that both agents are purely selfish (i.e., no transfers are made). This assumption reflects an informational asymmetry between the firm and the agents, in the sense that the firm cannot observe and/or contract upon the degree of prosociality one agent has for the other.

As shown in Appendix B, such insurance contract must satisfy two conditions

$$V(q^B, t^B) = V(0,0), \tag{26}$$

$$\left.\frac{dt}{dq}\right|_{V=V(0,0)} = \left.\frac{dt}{dq}\right|_{\pi=k}. \tag{27}$$

The first condition is the participation constraint, which extracts all the surplus of the agent through the appropriate choice of premium t^B, while the second condition is a akin to the incentive compatibility constraint, selecting the coverage q^B that maximizes profits given the equilibrium effort to be made by the agent.

Finally, the agents' equilibrium demand decision regarding the contract offered by the principal can be analyzed. By construction, the benchmark insurance contract $C^B = (q^B, t^B)$ is such that a purely self-interested agent is indifferent between purchasing it or not[18]. The same is true if the agents are not very altruistic: if α_1, α_2 are low enough such that no positive equilibrium transfers are made, then effectively not very altruistic agents will also prefer to buy the policy C^B.

However, as the degree of altruism increases and equilibrium transfers become positive, the benchmark policy C^B will be rejected by symmetric agents. This is due to two crucial features of C^B. First, it can provide very low coverage: for instance, in the example below coverage is less than 25% of the loss an agent faces. Second, C^B is expensive: it has a higher than actuarially fair premium.

Proposition 6: *Suppose that* $(w_1^H, w_1^L, \alpha_1) = (w_2^H, w_2^L, \alpha_2)$ *and that the firm offers* $C^B = (q^B, t^B)$. *Then, the set of Nash equilibria of the insurance demand game between agents* 1 *and* 2 *is*

1. $\{(a,a)\}$ if $\alpha < \frac{u'(w^H)}{u'(w^L)}$;
2. $\{(a,a),(r,r)\}$ if $\alpha \in \left[\frac{u'(w^H)}{u'(w^L)}, \min\left\{\frac{u'(w^H - t^B)}{u'(w^L)}, \frac{u'(w^H)}{u'(w^L - t^B + q^B)}\right\}\right]$;

[18] In line with the literature in mechanism design, I assume that in case of such indifference the agent will accept the principal's offered mechanism.

3. $\{(r, r)\}$ otherwise.

The suboptimality of C^B is related to the informational disadvantage of the principal. When altruism is relatively low, the agents' equilibrium behavior in terms of transfers and self-protection is identical to the one they would exhibit in autarky, and therefore the principal's informational disadvantage has no bite. On the other hand, for high degrees of altruism, the insurance company misjudges the actual risk the agents face and therefore offers an inefficient contract that is rejected by the agents.

Example (continued): *For the example symmetric wealths* $(w^H, w^L) = (3, 1)$ *and assuming a quadratic cost function* $\psi(p) = \frac{p^2}{2}$*, the contract satisfying equations (26)-(27) is given by* $C^B = (0.4485, 0.2108)$*, while equilibrium effort is* $p = 0.5576$*. Figure 11 exhibits the equilibrium insurance demand for the agents, with a more elaborate pattern than before.*

Focusing on symmetric degrees of altruism, Proposition 6 *is clear: for low degrees of altruism, such that no cross-insurance transfers are made, agents* 1 *and* 2 *purchase the insurance contract due to the absence of substitution effect. For the region in which only transfers between uninsured agents exist, represented by the dark yellow region in the middle of Figure 11, two symmetric equilibria exist:* (a, a) *and* (r, r)*. The former is present because no cross-insurance transfers are made when a unilateral deviation from* (a, a) *occurs, and therefore the agents are better off with the insurance contract. The latter is due to the opposite: given that an agent chooses to reject the insurance contract, the best-response for the other is to also reject it and engage in symmetric cross-insurance transfers.*

Last, for degrees of altruism such that cross-insurance transfers are positive in asymmetric purchase decision profiles, unilateral deviations from (a, a) *are now profitable. On the other hand, the best-response to* $d_j = r$ *still is to reject* C^B*, and therefore the unique Nash equilibrium is* (r, r)*.* ■

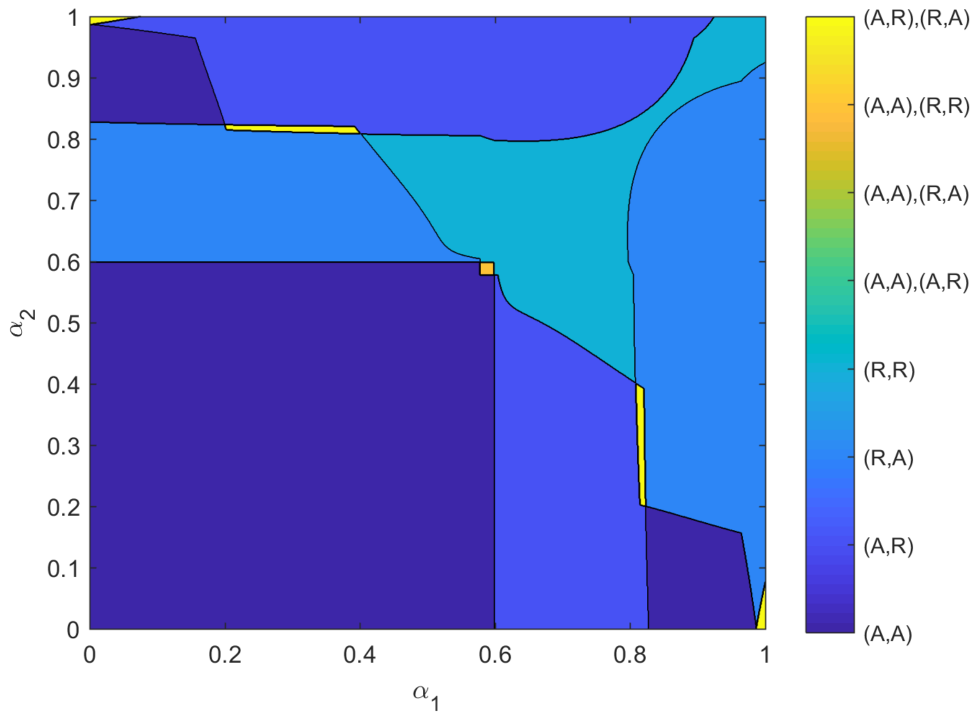


Figure 11: Equilibrium demand for insurance with self-protection, for $c = 1$ and $C^B = (0.4485, 0.2108)$.

5.2 Insurance Supply and Demand for a Fully Informed Principal

Now let us study the polar opposite case, where the principal is fully informed about the agents' common degree of altruism $\alpha \in [0, 1]$ and therefore correctly anticipates transfers and self-protection effort produced by 1 and 2 in a symmetric equilibrium. I show that the principal can then always find a contract that will yield him positive profits and induce both agents to purchase the contract.

Consider first low degrees of altruism, such that no positive transfers are made. Then, the benchmark contract C^B derived above yields positive profits, since its premium is larger than the actuarially fair one, at the same time that it makes the agents indifferent between accepting and rejecting it[19].

[19]A contract $C^{\varepsilon} = (q^B, t^B - \varepsilon)$ for $\varepsilon \approx 0$ would still yield positive profits for the principal and make the

To show the existence of a contract that generates gains from trades for all parties for high degrees of altruism, I extend the idea used to compute the benchmark contract C^B. In particular, I compute the indifference curve for the agents and the principal's zero profit line in the (q, t)-plane for any degree of altruism, and show that, at $(q, t) = (0, 0)$, the agent's indifference curve is steeper than the ZPL for all $\alpha \in [0, 1]$. Therefore, there exist partial coverage insurance contracts with larger than actuarially fair premiums that yield positive profits to the firm and are accepted by the agents. Proposition 7 synthesizes the result.

Proposition 7: *Consider the symmetric case where $(w_1^H, w_1^L, \alpha_1) = (w_2^H, w_2^L, \alpha_2)$, and consider symmetric strategies for the agents. Then, there exist contracts $C \in \mathcal{C}/\{(0,0)\}$ that the insurer can profitably offer to the agents for any $\alpha \in [0, 1]$.*

Example (continued): *Figure 12 and Figure 13 depict the firm's optimal contractual offer to the symmetric agents for different values of α and c. In both cases, the optimal contract exhibits a downward shift when cross-insurance transfers become positive. Such a shift reflects the principal's desire to minimize the free-riding effect, as well as the fact that the agents' outside option, given by the expected utility they would obtain from only cross-insuring with one another, is increasing in the common degree of altruism, as established in Proposition* 5.

Indeed, had the contract be kept constant, the cross-insurance transfers would lead both agents to drastically reduce equilibrium effort due to the free-riding effect, and therefore the firm would suffer losses. By reducing the coverage, the firm ensures that equilibrium effort remains high, and she accordingly adjusts the premium to extract as much surplus as possible, while taking into consideration the higher outside option obtained by the agents through cross-insurance transfers.

Such reduction in the contract's terms ultimately decreases the firm's expected profits: for either low or high cost parameter, the firm's expected profit is reduced close to zero after cross-insurance transfers become positive. Thus, in this example, if the firm faces a fixed entry cost or has a loading factor, it may choose not to trade at all with highly altruistic

agents strictly prefer it to being uninsured.

agents. ■

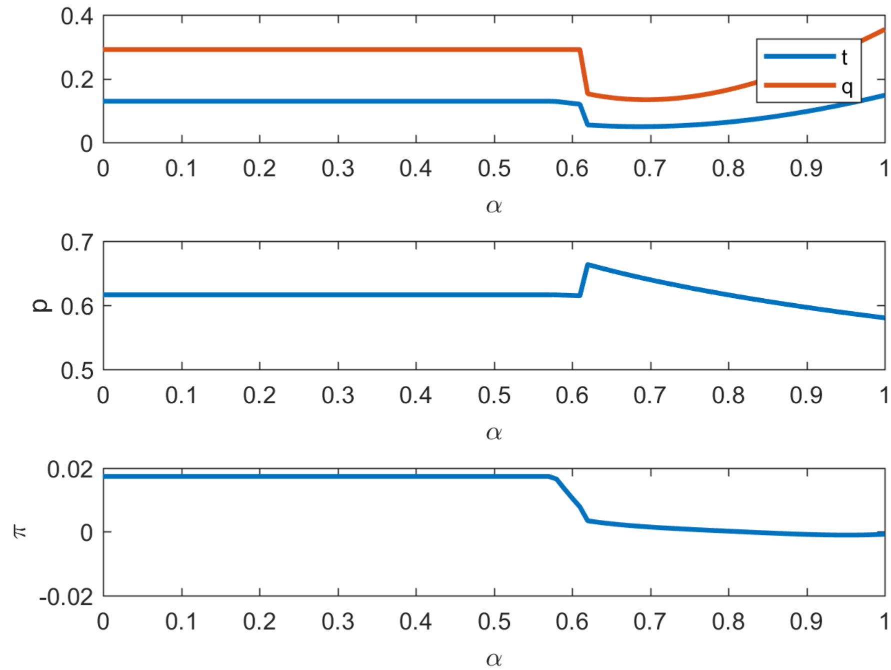


Figure 12: Optimal contract offer, equilibrium effort and firm's profit for $c = 1$.

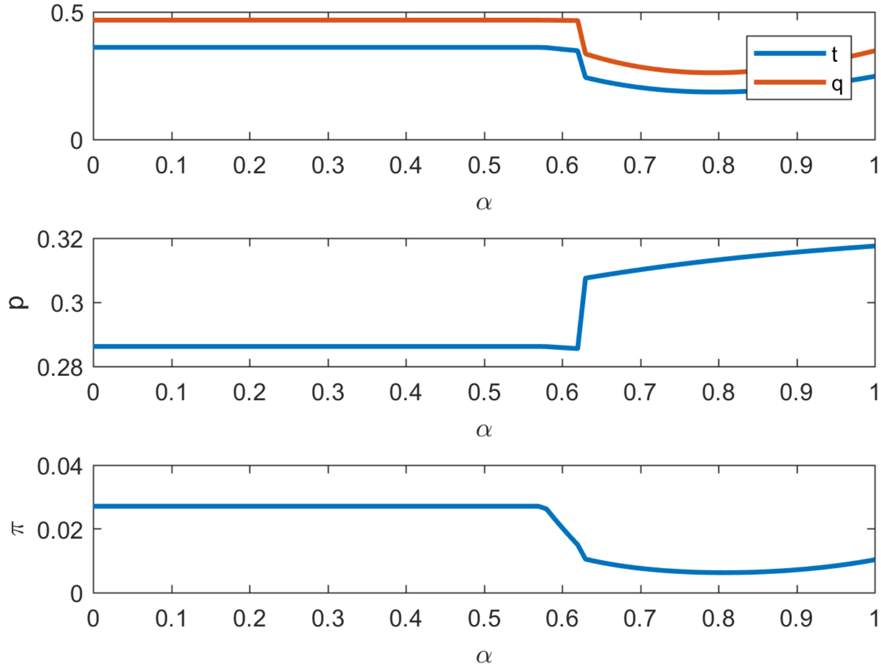


Figure 13: Optimal contract offer, equilibrium effort and firm's profit for $c = 2$.

6 Discussion

The presentation has mostly focused on the symmetric case where both agents share the same wealth and independent losses and, in the case without self-protection, the same probability of suffering a loss. In this section, I would like to briefly discuss some extensions and other issues related to the model.

1. *Estimating altruism:* The first testable implication of my model is related to informal transfers. For low degrees of altruism, no transfers should be observed, while they should become larger and more frequent as the degree of altruism increases. Moreover, if losses and insurance policies are known, one can use the threshold degree of altruism $\widehat{\alpha}$ to indirectly estimate the functional form of the utility of wealth by observing when transfers become positive.

The second testable implication is related to self-protection efforts. Due to the free-riding effect, one should observe an increase in the ratio of claims repayments to premiums when transfers are small in comparison to the same ratio when no cross-insurance transfers are observed if both agents are purchasing insurance. Also, for fixed losses and insurance terms, the ratio of claims repayments to premiums should behave non-monotonically as transfers increase.

While the two implications above can test the overall degree of altruism, the model can also be used to test dispersion of altruism by looking at insurance purchase decisions. In particular, if agents have similar wealths and face similar losses, observing cross-insurance transfers from an insured agent to an uninsured one is suggestive of the latter free-riding on the former's insurance policy, which should happen, according to my model, when the insured agent is very altruistic towards the uninsured one, whilst the uninsured agent is more self-interested.

Last, but not least, the main testable implication is that insurance penetration must be smaller in countries with higher degrees of altruism, due to the crowding-out of formal insurance demand by cross-insurance transfers. Costa-Font (2010) finds a similar result when studying LTC insurance demand in European countries, but using an constructed index of family ties instead of altruism.

2. *Adverse selection:* While I have focused on moral hazard as the main information asymmetry between insurer and insurees, an extensive body of literature has focused on adverse selection as the main information friction between the two parties. For instance, Hendren (2013) develops a model of adverse selection on insurance provision and tests it with data on life, disability and long-term care (LTC) insurance. In his setting, a unit mass of agents have the same wealth $w > 0$ and suffer a loss $L > 0$ with privately known (and exogenously given) probability $p \in [0, 1]$. My analysis of the case without self-protection is very much the same, save for the assumption that the principal knows p in my model.

The important remark that has to be made here is that introducing privately known probabilities does not affect the agents' transfers decision, since those are made *ex post*, i.e. only after realizing each other's output. Moreover, as long as the agents can compute their respective expected utilities, any contract offered by the principal will induce a lottery that can be compared in much the same way as in Proposition 2. Therefore, if the agents are privately informed about their probability of suffering a loss, their equilibrium decisions about transfers and insurance demand are unchanged from my analysis.

However, the same will not be true for the insurer. As stated by Proposition 2, even an actuarially fair full coverage insurance policy may be crowded out by one agent's altruism towards his pair. Introducing private information on the agents' part would just worsen the outlook for the principal. Indeed, suppose for instance $p = \{p_r, p_s\}$ such that $p_s > p_r$ and $Prob(p = p_s) = \lambda \in (0, 1)$. A pooling full coverage contract would induce losses for the principal, and so would a menu $\mathbf{C} = ((L, (1-p_s)L), (L, (1-p_r)L))$, since it would either attract only a altruistic but risky type r at the correct contract $(L, (1 - p_r)L)$ or attract a risky agent on the low premium contract $(L, (1 - p_s)L)$.

3. *The "chicken-and-egg" problem:* I have studied whether altruism, by means of the informal risk-sharing between agents it sustains, can hinder the emergence of formal insurance markets in the sense that a monopolistic insurer would have no demand for its policies when trying to enter the market populated by altruistic agents.

   The other side of the coin is how the presence of a formal insurance market would affect the degree of altruism of the agents populating such a market. In essence, what would be the evolutionarily stable degree of altruism in the society? Such a question would extend the analysis of Alger and Weibull (2010), who briefly discusses the effect of mandatory public insurance in their model. In particular, they argue that the introduction of public insurance may lead to a higher degree of altruism, unless transfers between agents are *completely* crowded out.

By fixing an insurance policy C in the game between agents of my model, one can follow the same steps as Alger and Weibull (2010) to determine the evolutionarily stable degree of altruism arising from the interaction between agents whenever a symmetric equilibrium is considered. Alternatively, one can study the evolutionarily stable *strategies* of such game, in particular the agents' decisions about insurance purchases. Last, but not least, determining the evolutionarily stable triplet of agents' strategies, principal's contract offers and degrees of altruism could potentially be done numerically and provide a first glimpse at how all these factors coevolve.

4. *Remittances:* The attentive reader will notice that the setup of my model allows for asymmetric levels of wealth. One particularly interesting case that can be analyzed by such extension is the effects of remittances on the development of local insurance markets.[20] For instance, suppose that agent 1 is a parent who lives in a poor region, while sibling 2 is a child who moves to a richer region or country.[21] In particular, assume that $w_2^H > w_2^L > w_1^H > w_1^L$, and suppose that each party has access to a local insurance market offering policies satisfying $0 \leq t_i \leq q_i \leq L_i$ for $i = 1, 2$. Then, one can see that transfer will always happen from the child to the parent, whenever the former's degree of altruism is not too low. In that case, if the difference in wealth is sufficiently large, the child may transfer amounts large enough to render the parent's demand for insurance null, thus effectively hindering the emergence of the local insurance market in the poor region.

   A similar result could be obtained for $w_2^H > w_1^H \geq w_2^L > w_1^L$, if w_1^H and w_2^L are sufficiently close. In this case, transfers always flow from agent 2 to agent 1, with one exception: when the former has suffered a loss but the latter hasn't. Then, once again, if α_2 is sufficiently high and so are the differences in wealth when agent 2 makes the transfers, then agent 1's demand for the local formal insurance market may be crowded out by the help he receives from agent 1.

[20] See, for instance, Azam and Gubert (2006) for a more detailed discussion of remittances.

[21] An alternative story could be a well-off parent and a child who goes away to study.

5. *Savings and self-insurance:* In the classical work of Ehrlich and Becker (1972), agents can *self-insure*, i.e., affect the size of the loss each faces. If self-insurance is costless, then formal insurance rejection is going to happen more often. In particular, if agents are able to eliminate the risk through self-insurance, not only formal insurance but also cross-insurance will be crowded-out.

   On the other hand, if self-insurance is costly, the interaction among all risk-sharing mechanisms is not obvious. While equilibrium transfers will still be computed *ex post* and exhibit a monotonic behavior with respect to the degree of altruism, self-protection and self-insurance may exhibit even more nonmonotonic behavior. I must point out that if such actions are not too costly, the crowding-out effect on formal insurance should become stronger.

   A similar result should be observed if agents are allowed to save (or redistribute wealth between outputs in any other way): the additional channel through which agents can share risk would ultimately reduce the need for formal insurance policies.

6. *Public vs. private insurance:* The exposition above has focused on the case of privately provided insurance policies. Another possibility is that the government provides the risk-sharing mechanism, either in place of the firm or in addition to it.

   If the government posts an insurance policy comprised only of a premium and a coverage, the demand for such policy can be studied in exactly the same way as above. The interesting question then would be how the private firm would design its policies to compete with the government's offer, a point that is beyond the scope of this paper.

   On the other hand, the public insurance can be compulsory, and the agents' choice is simply how to complement such policies.[22] Such a design can be incorporated in the model by rewriting the agents' levels of wealth to reflect the payments they make and

[22]This is the case for automobile insurance in many countries, where all individuals purchasing a car or motorcycle must pay the mandatory social liability policy and then choose to complement it with privately provided coverage policies.

receive for each possible output.

Overall, since the presence of a publicly provided insurance policy reduces the risk faced by agents, equilibrium transfers, effort and private insurance demand should decrease.

7 Concluding Remarks

I have studied why private insurance markets may fail to develop. While there are well-known issues on the supply side, such as information asymmetries and the presence of other players (like the government), this paper shows that issues also arise on the demand side. In particular, I have asked whether altruism and the informal risk-sharing sustained by it can be one factor hindering the emergence of formal insurance markets, and the analysis has provided a positive answer for this question. In an insurance setting with two altruistic agents, I consider the demand for insurance policies both when the agents can self-protect from a loss by exerting effort to change the probability of such loss taking place and when the agents become poorer with an exogenous probability.

Without self-protection, I have shown that even an actuarially fair full coverage insurance policy can be crowded out by informal risk-sharing between the agents when the dispersion of prosociality is high. More precisely, a self-interested agent may choose not to buy such a contract when he is paired with a highly altruistic agent, who makes large transfers in case the uninsured agent suffers a loss. This result is in line with the literature of long-term care insurance, in which formal insurance policies are rejected in favor of familial care.

Under certain parameterizations, a similar result holds for the case in which the agents can affect the probability of suffering a loss. However, due to the presence of informal risk-sharing, equilibrium levels of effort are not necessarily monotonic in the agents degrees of altruism, and such effort must balance two effects: a *free-riding* effect, in which an agent reduces his effort due to an increase in his partner's transfers to himself, and an *empathy* effect, which makes an agent increase his own effort in order to reduce the burden of his own loss in his partner.

Last, I considered a monopolistic insurer designing the contract to be offered to the agents. I have shown that, if the insurer cannot observe the agents' degree of altruism and thus offers the contract that would maximize her profits had she been facing a single self-interested agent, then both agents will reject such an offer for a sufficiently high common degrees of altruism. This is due to the principal's misinformation about the true preferences of the agents, who informally insure one another by means of transfers when they care enough about one another.

I also show that if the insurance firm can offer contracts contingent on the agents' common degree of altruism, gains from trade always exist. However, the insurer's quest to mitigate the substitution and free-riding effect may lead to reduced profitability for the firm, which may induce the firm to choose not to trade with altruistic agents if fixed costs or loading factors are present.

A Altruistic Agent's Risk-Aversion

Recall that an agent $i = 1, 2$ with degree of altruism $\alpha_i \in [0, 1]$ chooses transfers T_i to solve

$$\mathbb{E}\max_{T_i} u(y_i - T_i) + \alpha_i u(y_j + T_i) \tag{28}$$

for $j = 1, 2$, $j \neq i$. By the Envelope Theorem,

$$\frac{\partial V_i}{\partial y_i} = \mathbb{E}u'(y_i - T_i) \tag{29}$$

and thus

$$\frac{\partial^2 V_i}{\partial y_i^2} = \mathbb{E}u''(y_i - T_i) \cdot \left(1 - \frac{\partial T_i}{\partial y_i}\right), \tag{30}$$

where

$$\frac{\partial T_i}{\partial y_i} = \frac{u''(y_i - T_i)}{u''(y_i - T_i)^{+}\alpha_i u''(y_j + T_i)} \in (0, 1] \tag{31}$$

for all $\alpha_i \geq \widehat{\alpha}_i = \frac{u'(y_i)}{u'(y_j)}$. Therefore, the coefficient of absolute risk-aversion is given by

$$A_i(y_i, y_j, \alpha_i) = -\frac{\frac{\partial^2 V_i}{\partial y_i^2}}{\frac{\partial V_i}{\partial y_i}} = -\frac{\mathbb{E}u''(y_i - T_i)}{\mathbb{E}u'(y_i - T_i)} \cdot \underbrace{\left(\frac{\alpha_i u''(y_j + T_i)}{u''(y_i - T_i) + \alpha_i u''(y_j + T_i)}\right)}_{\in(0,1)}. \tag{32}$$

Inspection of (32) indicates that the coefficient of risk-aversion for an altruistic individual is equal to the coefficient for a selfish agent multiplied by a factor smaller than one, that is to say, for the same wealth, an altruistic agent is less risk-averse than a selfish counterpart. A second remark is that the altruistic agent's risk-aversion also depends on his partner's wealth, y_j, something absent when considering a selfish agent.

B Equilibrium Contract Offer for a Single Agent

Suppose that the principal offers a policy $C = (q, t) \in \mathcal{C}$ to a single agent. Such a policy assumes that agent i chooses effort to solve

$$\max_p pu(w^H - t) + (1 - p)u(w^L - t + q) - \psi(p), \tag{33}$$

leading to the first-order condition

$$u(w^H - t) - u(w^L - t + q) - \psi'(p^B) = 0 \tag{34}$$

where $p^B : (q, t) \to \mathbb{R}$ is a continuous function satisfying

$$\frac{dp^B}{dq} = -\frac{u'(w^L - t + q)}{\psi''(p^B)} < 0 \tag{35}$$

$$\frac{dp^B}{dt} = \frac{-u'(w^H - t) + u'(w^L - t + q)}{\psi''(p^B)} \geq 0 \tag{36}$$

for all $C = (q, t)$ such that $0 \leq t \leq q \leq L$. Writing

$$V(q, t) \equiv p^B(q, t)u(w^H - t) + (1 - p^B(q, t))u(w^L - t + q) - \psi(p^B(q, t)), \tag{37}$$

the marginal rate of substitution between premium and coverage is

$$\left.\frac{dt}{dq}\right|_{V=\overline{V}} = -\frac{\frac{\partial V}{\partial q}}{\frac{\partial V}{\partial t}} = \frac{(1 - p^B(q, t))u'(w^L - t + q)}{p^B(q, t)u'(w^H - t) + (1 - p^B(q, t))u'(w^L - t + q)}, \tag{38}$$

or alternatively,

$$\left.\frac{dt}{dq}\right|_{V=\overline{V}} = \frac{1}{1 + \frac{p}{1-p}\frac{u'(w^H - t)}{u'(w^L - t + q)}} \in (0, 1). \tag{39}$$

Notice that

$$\begin{aligned}\left.\frac{d^2 t}{dq^2}\right|_{V=\overline{V}} = {} & p^B(q, t)u'(w^H - t)\left[(1 - p^B(q, t))u''(w^L - t + q)\left(1 - \frac{dt}{dq}\right)\right. \\ & \left. - u'(w^L - t + q)\left(\frac{dp^B}{dq} + \frac{dp^B}{dt}\frac{dt}{dq}\right)\right] \\ & + (1 - p^B(q, t))u'(w^L - t + q) \\ & \quad \times \left[p^B(q, t)u''(w^H - t)\frac{dt}{dq} - u'(w^H - t)\left(\frac{dp^B}{dq} + \frac{dp^B}{dt}\frac{dt}{dq}\right)\right]\end{aligned} \tag{40}$$

which is negative for sufficiently large $\psi''(\cdot)$[23] since

$$\frac{dp^B}{dq} + \frac{dp^B}{dt}\frac{dt}{dq} = -\frac{u'(w^H - t)u'(w^L - t + q)}{\psi''(p^B(q,t))\left[p^B(q,t)u'(w^H - t) + (1 - p^B(q,t))u'(w^L - t + q)\right]}. \tag{41}$$

On the other hand, the expected profit made by contract C is given by

$$\pi(q,t) = p^B(q,t) \cdot t + (1 - p^B(q,t)) \cdot (t - q) = t - (1 - p^B(q,t))q. \tag{42}$$

Then, denote by $\mathcal{F} = \left\{(q,t) \in \mathbb{R}^2_+ : 0 \leq t \leq q \leq L, \pi(q,t) \geq 0\right\}$ the set of nonnegative expected profit insurance policies. The boundary of this set is the Zero Profit Line[24], defined as[25]

$$ZPL = \left\{(q,t) \in \mathbb{R}^2_+ : 0 \leq t \leq q \leq L, \pi(q,t) = 0\right\}. \tag{43}$$

For any arbitrary expected profit $k \in \mathbb{R}$, I can define the isoprofit curve

$$\overline{\pi}(k) = \left\{(q,t) \in \mathbb{R}^2_+ : 0 \leq t \leq q \leq L, t - (1-p)q = k\right\}. \tag{44}$$

As was the case with the agents' expected utility, the isoprofit $\overline{\pi}(k)$ implicitly defines the premium t as a function of coverage q, subject to 34 determining the agents' equilibrium effort. Therefore, by applying the implicit function theorem to the system of equations

$$F_1 = t - (1-p)q - k \tag{45}$$

$$F_2 = \psi'(p) - \left[u(w^H - t) - u(w^L - t + q)\right], \tag{46}$$

[23]One might find these results at odds with the seminal work by Arnott and Stiglitz (1991), but I must highlight two crucial differences from our approaches in modelling insurance. While I explicitly model premium and coverage separately, Arnott and Stiglitz (1991) consider the net payments after each output, namely $\beta = t$ and $\alpha = q - t$. Secondly, I map the choice of effort x into the choice of probability of not suffering a loss p and analyze the model in terms of the latter as well as allowing for a generical convex cost function for x, while Arnott and Stiglitz (1991) considers a linear cost of effort and does not simplify the model to a single effort/probability choice variable.

[24]This is exactly the Zero Profit Locus described in Arnott and Stiglitz (1991).

[25]Two remarks: First, note that $C_0 = (0,0)$ belongs to the ZPL by construction. Second, $C^L = (L,L)$ also belongs to the ZPL; indeed, at $q = L$, effort is zero following (34), and thus $L - (1-0)L = 0$.

the slope of the isoprofit curve $\overline{\pi}(k)$ is given by[26]

$$\left.\frac{dt}{dq}\right|_{\pi=k} = \frac{(1-p)\psi''(p) + qu'(w^L - t + q)}{\psi''(p) + q[u'(w^L - t + q) - u'(w^H - t)]} > 0. \tag{47}$$

Notice that at $(q,t) = (0,0)$ and for $k = 0$,

$$\left.\frac{dt}{dq}\right|_{\pi=0}(0,0) = 1 - p < \frac{(1-p)u'(w^L)}{pu'(w^H) + (1-p)u'(w^L)} = \left.\frac{dt}{dq}\right|_{V=V(0,0)}(0,0), \tag{48}$$

so that there exists some contract $\widetilde{C} \neq (0,0)$ that yields a higher expected utility to the agent and a positive expected profit to the principal, i.e., there are gains from trade to be had in the interaction between the firm and a single agent. Thus, the principal chooses (q^B, t^B) to maximize expected profits, and the benchmark insurance policy must satisfy the conditions

$$V(q^B, t^B) = V(0,0), \tag{49}$$

$$\left.\frac{dt}{dq}\right|_{V=V(0,0)} = \left.\frac{dt}{dq}\right|_{\pi=k}, \tag{50}$$

where the first condition, (49), implies that the principal chooses the policy that makes the agent indifferent between purchasing it and remaining uninsured, while (50) is the tangency condition for maximization. Notice that (50) can be written as

$$-\frac{q}{1-p}u'(w^H - t) = p\psi''(p)\left[\frac{u'(w^H - t) - u'(w^L - t + q)}{u'(w^L - t + q)}\right], \tag{51}$$

where the left-hand side is strictly positive for $q = L$ while the right-hand side is equal to zero for full coverage. Therefore, as I have argued before, the benchmark contract under self-protection must offer only partial insurance, so the agents have incentives to exert positive effort. Additionally, surplus extraction by the principal implies that such contract is not actuarially fair.

[26]Imposing the isoprofit condition $t = k + (1-p)q$, one can see that $\left.\frac{d^2t}{dq^2}\right|_{\pi=k} > 0$, i.e. the isoprofit curve is a convex function in the (q,t)-plane.

C Proofs

C.1 Proof of Proposition 1

Suppose, by contradiction, that $\alpha_1\alpha_2 < 1$ and that $(b_1, b_2) \in \mathbb{R}^2_{++}$ is a Nash equilibrium of $G(\omega)$. The first-order conditions for the maximization problem of the agents in (10) are

$$u'(y_1 - b_1 + b_2) = \alpha_1 u'(y_2 - b_2 + b_1) \tag{52}$$

$$u'(y_2 - b_2 + b_1) = \alpha_2 u'(y_1 - b_1 + b_2). \tag{53}$$

Substituting (52) into (53) yields $u'(y_2 - b_2 + b_1) = \alpha_1\alpha_2 u'(y_2 - b_2 + b_1)$, which can only hold if $\alpha_1\alpha_2 = 1$ since $u' > 0$ by assumption, a contradiction. Thus, if $\alpha_1\alpha_2 < 1$, at most one transfer is positive.

Let $\widehat{\tau}_i : \omega \to [0, w^H]$ be the transfer agent i would give to his pair if the latter makes no transfer to i. Then, if $u'(y_i) \geq \alpha_i u'(y_j)$, agent j is already richer than i, and thus i makes no transfers, i.e. $\widehat{\tau}_i(\omega) = 0$. Otherwise, $\widehat{\tau}_i(\omega)$ is positive and determined by the first-order condition $u'(y_i - \widehat{\tau}_i) = \alpha_i u'(y_j + \widehat{\tau}_i)$, which is uniquely defined.

Thus, if $\alpha_1\alpha_2 < 1$, the unique Nash equilibrium of $G(\omega)$ is

- $(b_1, b_2) = (0, 0))$ when $y_1 = y_2$;
- $(b_1, b_2) = (\widehat{\tau}_1(\omega), 0)$ when $y_1 > y_2$;
- $(b_1, b_2) = (0, \widehat{\tau}_2(\omega))$ when $y_1 < y_2$.

Finally, if $\alpha_1 = \alpha_2 = 1$, then

- if $y_1 > y_2$, any $(b_1, b_2) = (\widehat{\tau}_1(\omega) + \varepsilon, \varepsilon)$ is a Nash equilibrium of $G(\omega)$ for all $\varepsilon \in (0, y_1 - \widehat{\tau}_1(\omega))$;
- if $y_1 < y_2$, any $(b_1, b_2) = (\varepsilon, \widehat{\tau}_2(\omega) + \varepsilon)$ is a Nash equilibrium of $G(\omega)$ for all $\varepsilon \in (0, y_2 - \widehat{\tau}_2(\omega))$;
- if $y_1 = y_2$, any $(b_1, b_2) = (\varepsilon, \varepsilon)$ is a Nash equilibrium of $G(\omega)$ for any $\varepsilon \in [0, y_1]$.

C.2 Proof of Proposition 2

Fix $(\alpha_1, \alpha_2) = (1, 0)$ and $C^{af} = (L, (1-p)L)$, while wealth and probabilities of losses are symmetric and given by (w^H, w^L, p). Since $\alpha_2 = 0$, $T_2(d, \omega, C) = 0$ for any triple (d, ω, C). On the other hand,

$$T_1(d, \omega, C^{af}) = \begin{cases} \frac{w^H - \overline{w}}{2} & \text{if } (d_1, d_2) = (r, a),\ \omega = (w^H, \cdot) \text{ and } \alpha \geq \frac{u'(w^H)}{u'(\overline{w})}, \\ \frac{\overline{w} - w^L}{2} & \text{if } (d_1, d_2) = (a, r),\ \omega = (\cdot, w^L) \text{ and } \alpha \geq \frac{u'(\overline{w})}{u'(w^L)}, \\ \frac{w^H - w^L}{2} & \text{if } (d_1, d_2) = (r, r),\ \omega = (w^H, w^L) \text{ and } \alpha \geq \frac{u'(w^H)}{u'(w^L)}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\overline{w} = pw^H + (1-p)w^L$.

Given transfers, the expected utilities for the altruistic agent 1 are

$$U_1(a, a) = 2u(\overline{w}), \tag{54}$$

$$U_1(r, a) = \left[pu\left(\frac{w^H + \overline{w}}{2}\right) + (1-p)u(\overline{w})\right] + \left[pu\left(\frac{w^H + \overline{w}}{2}\right) + (1-p)u(w^L)\right], \tag{55}$$

$$U_1(a, r) = p[u(\overline{w}) + u(w^H)] + 2(1-p)u\left(\frac{\overline{w} + w^L}{2}\right), \tag{56}$$

$$U_1(r, r) = 2p^2u(w^H) + 2(1-p)^2u(w^L) + p(1-p)\left[2u\left(\frac{w^H + w^L}{2}\right) + u(w^H) + u(w^L)\right] \tag{57}$$

for each profile of insurance purchase decisions (d_1, d_2). Dividing each expression by 2, I can compute the lotteries induced by the decision profiles and, in particular, I can show that $\mathbb{E}[w] = \overline{w}$ for every one of them. It is then trivial to check that the lottery induced by (r, a) is a mean-preserving spread of the lottery induced by (a, a), while the same is true for the lottery induced by (a, r) with respect to the one from (r, r). Because the agents are assumed to be risk-averse, second-order stochastic dominance then implies that $U_1(a, a) > U_1(r, a)$ and $U_1(a, r) > U_1(r, r)$, and therefore, buying the policy C^{af} is a strictly dominant strategy for the altruistic agent.

Meanwhile, the expected utilities for the purely self-interested agent 2 are

$$U_2(a,a) = u(\overline{w}), \tag{58}$$

$$U_2(a,r) = pu(w^H) + (1-p)u\left(\frac{w^L + \overline{w}}{2}\right), \tag{59}$$

$$U_2(r,a) = pu\left(\frac{\overline{w} + w^H}{2}\right) + (1-p)u(\overline{w}), \tag{60}$$

$$U_2(r,r) = p^2u(w^H) + p(1-p)u\left(\frac{w^H + w^L}{2}\right) + (1-p)^2u(w^L). \tag{61}$$

First, notice that the lotteries induced by (r,a) and (r,r) have the same mean, but the later can be constructed as a mean-preserving spread of the former, and therefore I have that $U_2(r,a) > U_2(r,r)$, i.e. the best response for the self-interested agent when the altruistic one rejects insurance is to buy it. On the other hand, notice that

$$\mathbb{E}_{(a,r)}[w] = pw^H + (1-p)\frac{w^L + \overline{w}}{2} > pw^H + (1-p)w^L = \overline{w} = \mathbb{E}_{(a,a)}[w] \tag{62}$$

for any $p \in (0,1)$. Thus, if $i = 2$ is risk-neutral, he strictly prefers not to buy C^{af} when $i = 1$ does so, and, by continuity, the same holds true if the agents are not too risk-averse.

C.3 Proof of Proposition 3

Fix $(w_1^H, w_1^L, p_1) = (w_2^H, w_2^L, p_2)$, $(\alpha_1, \alpha_2) = (1,0)$ and $(d_1, d_2) = (a,r)$. Then, equilibrium transfers from agent 1 to agent 2 must satisfy

$$T(\omega, C, (a,r)) = \begin{cases} 0 & \text{if } \omega = (w^H, w^H), \\ \frac{w^H - w^L - t}{2} & \text{if } \omega = (w^H, w^L), \\ 0 & \text{if } \omega = (w^L, w^H), \\ \frac{q-t}{2} & \text{if } \omega = (w^L, w^L), \end{cases}$$

so that the expected utility of agent 1 is given by

$$\begin{aligned}
\mathbb{E}\left[U_1(a,r;C)\right] &= p^2\left[u(w^H-t)+u(w^H)\right]+(1-p)p\left[u(w^L-t+q)+u(w^H)\right] \\
&\quad + p(1-p)\left[u(w^H-t-T(w^H,w^L))+u(w^L+T(w^H,w^L)\right] \\
&\quad + (1-p)^2\left[u(w^L-t+q-T(w^L,w^L))+u(w^L+T(w^L,w^L))\right] \\
&= p^2\left[u(w^H-(1-p)q)+u(w^H)\right]+(1-p)p\left[u(w^L+pq)+u(w^H)\right] \\
&\quad + p(1-p)2u\left(\frac{w^H+w^L-(1-p)q}{2}\right)+(1-p)^2 2u\left(w^L+\frac{pq}{2}\right) \qquad (63)
\end{aligned}$$

when the actuarially fair policy $C=(q,(1-p)q)$ is considered.

Then, I can show that

$$\begin{aligned}
\frac{\partial\mathbb{E}\left[U_1(a,r;C)\right]}{\partial q} &= p(1-p)\left\{p\left[u'(w^L+pq)-u'(w^H-q+pq)\right]\right. \\
&\quad \left. +(1-p)\left[u'\left(w^L+\frac{pq}{2}\right)-u'\left(\frac{w^H+w^L-q}{2}+\frac{pq}{2}\right)\right]\right\} \qquad (64)
\end{aligned}$$

so that

$$\frac{\partial\mathbb{E}\left[U_1(a,r;C)\right]}{\partial q}\begin{cases} >0 & \text{if } q<L, \\ =0 & \text{if } q=L, \\ <0 & \text{if } q>L, \end{cases}$$

and

$$\begin{aligned}
\frac{\partial^2\mathbb{E}\left[U_1(a,r;C)\right]}{\partial q^2} &\propto p^2u''(w^L+pq)+p(1-p)u''(w^H-(1-p)q) \\
&\quad \frac{p(1-p)}{2}u''\left(w^L+\frac{pq}{2}\right)+\frac{(1-p)^2}{2}u''\left(\frac{w^H+w^L-(1-p)q}{2}\right) \\
&< 0 \qquad (65)
\end{aligned}$$

since $u'>0>u''$ by assumption. Therefore, $q=L$ is a global maximum when (d_1,d_2).

Now, suppose that $(d_1,d_2)=(a,a)$, so that equilibrium transfers become

$$T(\omega,C,(a,r))=\begin{cases} 0 & \text{if } \omega=(w^H,w^H), \\ \frac{w^H-w^L-q}{2} & \text{if } \omega=(w^H,w^L), \\ 0 & \text{if } \omega=(w^L,w^H), \\ 0 & \text{if } \omega=(w^L,w^L), \end{cases}$$

while expected utility is

$$\begin{aligned}
\mathbb{E}\left[U_1(a,a;C)\right] = {} & 2p^2u(w^H-(1-p)q)+2(1-p)^2u(w^L+pq) \\
& +2p(1-p)u\left(\frac{w^H+w^L+pq-(1-p)q}{2}\right) \\
& +(1-p)p\left[u(w^L+pq)+u(w^H-(1-p)q)\right]. \qquad (66)
\end{aligned}$$

Taking the partial derivative with respect to coverage yields

$$\begin{aligned}
\frac{\partial\mathbb{E}\left[U_1(a,a;C)\right]}{\partial q} = {} & p(1-p)\left\{(1+p)\left[u'(w^L+pq)-u'(w^H-(1-p)q)\right]\right. \\
& \left. +(1-2p)\left[u'(w^L+pq)-u'\left(\frac{w^H+w^L+pq-(1-p)q}{2}\right)\right]\right\} \qquad (67)
\end{aligned}$$

which is positive for $q<L$ and equal to zero if $q=L$, while the second derivative is

$$\begin{aligned}
\frac{\partial^2\mathbb{E}\left[U_1(a,a;C)\right]}{\partial q^2} \propto {} & (1-p^2)u''(w^H-(1-p)q)+p(2-p)u''(w^L+pq) \\
& +\frac{(1-2p)^2}{2}u''\left(\frac{w^H+w^L+pq-(1-p)q}{2}\right) \\
< {} & 0. \qquad (68)
\end{aligned}$$

C.4 Proof of Lemma 2

Fix $C=(q,t)\in\mathcal{C}$, and assuming a symmetric equilibrium, clearly (21) becomes (22), which can be rewritten as

$$\psi'(p)=a(\omega,C,\alpha)-(1+\alpha)pb(\omega,C,\alpha) \qquad (69)$$

for

$$a(\cdot)=u(w^H-t-T(\alpha,C))-u(w^L-t+q)+\alpha[u(w^L-t+q+T(\alpha,C))-u(w^L-t+q)]\geq 0 \qquad (70)$$

and

$$b(\cdot)=u(w^L-t+q+T(\alpha,C))-u(w^L-t+q)+\left[u(w^H-t)-u(w^H-t-T(\alpha,C))\right]\geq 0, \qquad (71)$$

with strict inequalities holding for $q<L$. Therefore, the left-hand side of (69) is a continuous and increasing function from zero to plus infinity by assumption, while the right-hand side is a decreasing affine function with positive intercept, which establishes the uniqueness claim.

C.5 Proof of Proposition 4

Let $y^L = w^L - t + q$ and $y^H = w^H - t$ for any $C = (q, t) \in \mathcal{C}$. Applying the Implicit Function Theorem on (22) yields

$$\begin{aligned}\frac{dp^*}{d\alpha} &= \frac{1-p^*}{A}\left[u(y^L + T(\cdot,\alpha)) - u(y^L)\right] \\ &\quad + \frac{p^*}{A}\left[u(y^H) - u(y^H - T(\cdot,\alpha))\right] \\ &\quad - \frac{p^*(1-\alpha^2)}{A}u'(y^L + T(\cdot,\alpha))\frac{dT}{d\alpha} \end{aligned} \tag{72}$$

where

$$A = \psi''(p^*) + (1+\alpha)b(\omega, C, \alpha) > 0 \tag{73}$$

and $b(\cdot)$ is the same given in (70). Recall that $\frac{dT}{d\alpha} > 0$ whenever $\alpha \geq \widehat{\alpha}(C)$. As $\alpha \downarrow \widehat{\alpha}(C)$, a point in which p^* is not differentiable, the first two terms in (72) tend to zero while the last term is negative, thus implying that $\frac{dp^*}{d\alpha} < 0$ for all $\alpha > \widehat{\alpha}(C)$ close to $\widehat{\alpha}(C)$. On the other hand, as $\alpha \uparrow 1$, the third term in (72) goes to zero whie the first two terms remain positive, thus implying that $\frac{dp^*}{d\alpha} > 0$ for large degrees of altruism.

C.6 Proof of Proposition 5

For item (1), I first characterize the socially optimal probability p and transfer τ to be given from the richer to the poor under a social welfare function, and then verify that these coincide with equilibrium probabilities p^* and transfers $T(\cdot)$ if and only if $\alpha = 1$.

Let $y^H = w^H - t$ and $y^L = w^L - t + q$ for any $C \in \mathcal{C}$. A hypothetical social planer must then choose p and τ to maximize the expected material payoff of one individual (due to the symmetry assumption), i.e. choose (p, τ) to maximize

$$W(p,\tau;C) = p^2u(y^H) + (1-p)^2u(y^L) + p(1-p)\left[u(y^H - \tau) + u(y^L + t)\right] - \psi(p). \tag{74}$$

The necessary first-order condition for an interior solution for p is

$$\psi'(p) = 2pu(y^H) - 2(1-p)u(y^L) + (1-2p)\left[u(y^H - \tau) + u(y^L + \tau)\right], \tag{75}$$

while, for any value of p, full risk-sharing maximizes $W(\cdot)$, i.e. transfers are such that $y^H - \tau = y^L + \tau$ for every output.

Looking back at the equilibrium condition for transfers in (12), $u''(\cdot) < 0$ implies that $y^H - T(\cdot, \alpha) = y^L + T(\cdot, \alpha)$ if and only if $\alpha = 1$. Moreover, for $\alpha = 1$, the symmetric equilibrium effort condition in (22) coincides with (75), and therefore $\alpha = 1$ is a necesary and sufficient condition for the equilibrium output to coincide with the welfare maimizing result.

For item (2), let $V(\alpha, \beta; C)$ denote the expected material payoff obtained by one individual with degree of altruism $\alpha \in [0, 1]$ in equilibrium play with another agent characterized by the degree of altruism $\beta \in [0, 1]$, i.e.

$$\begin{aligned} V(\alpha, \beta; C) = {} & p(\alpha, \beta)p(\beta, \alpha)u(y^H) \\ & + [1 - p(\alpha, \beta)][1 - p(\beta, \alpha)]u(y^L) \\ & + p(\alpha, \beta)[1 - p(\beta, \alpha)]u(y^H - T(\cdot, \alpha)) \\ & + [1 - p(\alpha, \beta)]p(\beta, \alpha)u(y^L + T(\cdot, \alpha)) \\ & - \psi(p(\alpha, \beta)). \end{aligned} \tag{76}$$

The claim in the proposition holds if

$$\lim_{\alpha \downarrow \widehat{\alpha}(C)} \left[\frac{\partial V(\alpha, \beta; C)}{\partial \alpha} + \frac{\partial V(\alpha, \beta; C)}{\partial \beta} \right] \Bigg|_{\beta = \alpha} > 0. \tag{77}$$

For the remainder of the proof, I will omit the conditioning on policy C and denote partial derivatives of a function $f(x_1, x_2)$ with respect to argument x_i by $f_i(x_1, x_2)$.

From (76), the corresponding probabilities of not suffering a loss, $p(\alpha, \beta)$ and $p(\beta, \alpha)$, satisfy the system of first-order conditions

$$\begin{aligned} \psi'(p(\alpha, \beta)) = {} & u(y^H) - u(y^L) \\ & + [1 - p(\beta, \alpha)] \left[u(y^H - T(\alpha)) + \alpha u(y^L + T(\alpha)) - (u(y^H) + \alpha u(y^L)) \right] \\ & - p(\beta, \alpha) \left[u(y^L + T(\beta)) + \alpha u(y^H - T(\beta)) - (u(y^L) + \alpha u(y^H)) \right], \end{aligned} \tag{78}$$

$$\psi'(p(\beta,\alpha)) = u(y^H) - u(y^L)$$
$$+ [1 - p(\alpha,\beta)] \left[u(y^H - T(\beta)) + \beta u(y^L + T(\beta)) - (u(y^H) + \beta u(y^L))\right]$$
$$- p(\alpha,\beta) \left[u(y^L + T(\alpha)) + \beta u(y^H - T(\alpha)) - (u(y^L) + \beta u(y^H))\right]. \qquad (79)$$

Taking the partial derivatives of V with respect to the degrees of altruism α and β lead, respectively, to

$$V_1(\alpha,\beta) = [p_1(\alpha,\beta)p(\beta,\alpha) + p(\alpha,\beta)p_2(\beta,\alpha)]\, u(y^H)$$
$$- [p_1(\alpha,\beta)[1 - p(\beta,\alpha)] + [1 - p(\alpha,\beta)]p_2(\beta,\alpha)]\, u(y^L)$$
$$+ [p_1(\alpha,\beta)[1 - p(\beta,\alpha)] - p(\alpha,\beta)p_2(\beta,\alpha)]\, u(y^H - T(\alpha))$$
$$- [p_1(\alpha,\beta)p(\beta,\alpha) - [1 - p(\alpha,\beta)]p_2(\beta,\alpha)]\, u(y^L + T(\beta))$$
$$- p(\alpha,\beta)[1 - p(\beta,\alpha)]u'(y^H - T(\alpha))T'(\alpha)$$
$$- \psi'(p(\alpha,\beta))p_1(\alpha,\beta) \qquad (80)$$

and

$$V_2(\alpha,\beta) = [p_2(\alpha,\beta)p(\beta,\alpha) + p(\alpha,\beta)p_1(\beta,\alpha)]\, u(y^H)$$
$$- [p_2(\alpha,\beta)[1 - p(\beta,\alpha)] + [1 - p(\alpha,\beta)]p_1(\beta,\alpha)]\, u(y^L)$$
$$+ [p_2(\alpha,\beta)[1 - p(\beta,\alpha)] - p(\alpha,\beta)p_1(\beta,\alpha)]\, u(y^H - T(\alpha))$$
$$- [p_2(\alpha,\beta)p(\beta,\alpha) - [1 - p(\alpha,\beta)]p_1(\beta,\alpha)]\, u(y^L + T(\beta))$$
$$+ p(\beta,\alpha)[1 - p(\alpha,\beta)]u'(y^L + T(\beta))T'(\beta)$$
$$- \psi'(p(\alpha,\beta))p_2(\alpha,\beta). \qquad (81)$$

From the system of equations (78)-(79), one can write

$$\psi'(p(\alpha,\beta)) = p(\beta,\alpha)(1+\alpha)u(y^H)$$
$$- [1 - p(\beta,\alpha)](1+\alpha)u(y^L)$$
$$+ [1 - p(\beta,\alpha)] \left[u(y^H - T(\alpha)) + \alpha u(y^L + T(\alpha))\right]$$
$$- p(\beta,\alpha) \left[u(y^L + T(\beta)) + \alpha u(y^H - T(\beta))\right], \qquad (82)$$

which I use to replace the last terms in equations (80)-(81) and simplify to obtain

$$\begin{aligned}
V_1(\alpha,\beta) = {} & [p(\alpha,\beta)p_2(\beta,\alpha) - \alpha p(\beta,\alpha)p_1(\alpha,\beta)]\, u(y^H) \\
& - [[1-p(\alpha,\beta)]p_2(\beta,\alpha) - \alpha[1-p(\beta,\alpha)]p_1(\alpha,\beta)]\, u(y^L) \\
& - p(\alpha,\beta)p_2(\beta,\alpha)u(y^H - T(\alpha)) \\
& + p_1(\alpha,\beta)p(\beta,\alpha)\alpha u(y^H - T(\beta)) \\
& + [1-p(\alpha,\beta)]p_2(\beta,\alpha)u(y^L + T(\beta)) \\
& - [1-p(\beta,\alpha)]p_1(\alpha,\beta)\alpha u(y^L + T(\alpha)) \\
& - p(\alpha,\beta)[1-p(\beta,\alpha)]u'(y^H - T(\alpha))T'(\alpha)
\end{aligned} \tag{83}$$

and

$$\begin{aligned}
V_2(\alpha,\beta) = {} & [p(\alpha,\beta)p_1(\beta,\alpha) - \alpha p(\beta,\alpha)p_2(\alpha,\beta)]\, u(y^H) \\
& - [[1-p(\alpha,\beta)]p_1(\beta,\alpha) - \alpha[1-p(\beta,\alpha)]p_2(\alpha,\beta)]\, u(y^L) \\
& - p(\alpha,\beta)p_1(\beta,\alpha)u(y^H - T(\alpha)) \\
& + p_2(\alpha,\beta)p(\beta,\alpha)\alpha u(y^H - T(\beta)) \\
& + [1-p(\alpha,\beta)]p_1(\beta,\alpha)u(y^L + T(\beta)) \\
& - [1-p(\beta,\alpha)]p_2(\alpha,\beta)\alpha u(y^L + T(\alpha)) \\
& + p(\beta,\alpha)[1-p(\alpha,\beta)]u'(y^L + T(\beta))T'(\beta).
\end{aligned} \tag{84}$$

Rearranging the expressions after evaluating them at $(\alpha,\beta) = (\alpha,\alpha)$ yields

$$\begin{aligned}
V_1(\alpha,\alpha) = {} & p(\alpha,\alpha)\,[p_2(\alpha,\alpha) - \alpha p_1(\alpha,\alpha)]\,\left[u(y^H) - u(y^H - T(\alpha))\right] \\
& + [1-p(\alpha,\alpha)]\,[p_2(\alpha,\alpha) - \alpha p_1(\alpha,\alpha)]\,\left[u(y^L + T(\alpha)) - u(y^L)\right] \\
& - p(\alpha,\alpha)[1-p(\alpha,\alpha)]u'(y^H - T(\alpha))T'(\alpha)
\end{aligned} \tag{85}$$

and

$$\begin{aligned}
V_2(\alpha,\alpha) = {} & p(\alpha,\alpha)\,[p_1(\alpha,\alpha) - \alpha p_2(\alpha,\alpha)]\,\left[u(y^H) - u(y^H - T(\alpha))\right] \\
& + [1-p(\alpha,\alpha)]\,[p_1(\alpha,\alpha) - \alpha p_2(\alpha,\alpha)]\,\left[u(y^L + T(\alpha)) - u(y^L)\right] \\
& + p(\alpha,\alpha)[1-p(\alpha,\alpha)]u'(y^L + T(\alpha))T'(\alpha).
\end{aligned} \tag{86}$$

By employing the first-order condition determining equilibrium transfers (12) for any degree of altruism above the threshold $\alpha > \widehat{\alpha}(C)$, and rearranging terms, I obtain

$$\begin{aligned} V_1(\alpha,\alpha) + V_2(\alpha,\alpha) = &(1-\alpha)[p_1(\alpha,\alpha)+p_2(\alpha,\alpha)]p(\alpha,\alpha)\left[u(y^H) - u(y^H - T(\alpha))\right] \\ &+ (1-\alpha)[p_1(\alpha,\alpha)+p_2(\alpha,\alpha)][1-p(\alpha,\alpha)]\left[u(y^L + T(\alpha)) - u(y^L)\right] \\ &+ (1-\alpha)p(\alpha,\alpha)[1-p(\alpha,\alpha)]u'(y^L + T(\alpha))T'(\alpha). \end{aligned} \tag{87}$$

As $\alpha \downarrow \widehat{\alpha}(C)$, $T(\alpha) \to 0$ and the first two terms tend to zero. Meanwhile, the third term tends to a positive number, and therefore the condition in (77) is satisfied.

C.7 Proof of Proposition 6

By construction, the benchmark policy $C^B = (q^B, t^B)$ is a solution to the principal's problem

$$\begin{aligned} \max_{(q,t)} \quad & \pi(q,t) = t - (1-p)q \\ s.t. \quad & \psi(p) = u(w^H - t) - u(w^L - t + q) \\ & pu(w^H - t) + (1-p)u(w^L - t + q) - \psi(p) \geq p^{Aut}u(w^H) + (1-p^{Aut})u(w^L) - \psi(p^{Aut}) \end{aligned}$$

where p^{Aut} satisfies $\psi(p^{Aut}) = u(w^H) - u(w^L)$. Standard arguments imply that the t^B is such that the agent is indifferent between accepting and rejecting the principal's offer, i.e. the last inequality is satisfied with equality. Assuming, as in common in mechanism design, that the indifference will be broken in favor of the principal, the agent accepts C^B.

Let us now turn to the agents' demand for C^B. First, suppose that α is small enough so that no transfers take place, i.e.

$$\alpha \leq \underline{\alpha} \equiv \frac{u'(w^H)}{u'(w^L)}. \tag{88}$$

Indeed, notice that $\underline{\alpha}$ is the threshold for transfers when both agents reject the insurance contract, and since I focus on $C \in \mathcal{C}$, $\underline{\alpha}$ is the lowest threshold for transfers in any possible equilibrium of the game between agents 1 and 2. In this case, the each agent's problem is

$$\max_p \quad pu(w^H - t) + (1-p)u(w^L - t + q) - \psi(p) \tag{89}$$

for $(q,t) \in \{(0,0),(q^B,t^B)\}$, and thus, by construction of (q^B,t^B), the each prefers to buy the benchmark policy C^B. Therefore, (a,a) is the unique Nash equilibrium of the insurance demand game for symmetric $\alpha \leq \underline{\alpha}$ and $C = C^B$.

Now, suppose that

$$\alpha \in \left(\underline{\alpha}, \min\left\{\frac{u'(w^H - t^B)}{u'(w^L)}, \frac{u'(w^H)}{u'(w^L - t^B + q^B)}\right\}\right]. \tag{90}$$

In this case, positive transfers take place only if the strategy profile of the agents involves (r,r). By construction of C^B, the unilateral deviation of an agent from (a,a) to (r,a) is not profitable, since it is equivalent to the problem (88). Thus, (a,a) is a Nash equilibrium of the game. On the other hand, it must be the case that $U_i(r,r;C^B,\alpha) > U_i(a,r;C^B,\alpha)$: for this range of degrees of altruism, the utility obtained by the agents in a symmetric equilibrium with positive transfers is necessarily larger than the utilities they would obtain absent transfers due to the optimality of the functions T and p, while by construction of C^B the utility each agent receives for an asymmetric profile $d \in \{(a,r),(r,a)\}$ is identical to the one they would obtain absent any transfers. Therefore, (r,r) is also a Nash equilibrium of the insurance demand game for the interval in (90).

For

$$\alpha \in \left(\min\left\{\frac{u'(w^H - t^B)}{u'(w^L)}, \frac{u'(w^H)}{u'(w^L - t^B + q^B)}\right\}, \frac{u'(w^H - t^B)}{u'(w^L - t^B + q^B)}\right], \tag{91}$$

the agents can now make positive transfers to one another under the asymmetric profiles (a,r) and (r,a), while the same is not possible for the symmetric profile (a,a), which induces a larger payoff than when transfers are zero. Therefore, (a,a) is not an equilibrium, since the agents can profitably unilaterally deviate from (a,a). Now, I must show that the best-response for agent i when agent j rejects C^B is r. If that was not the case, then the lottery induced by (a,r) has a mean no smaller than the lottery induced by (r,r), since the last has the widest possible range. But this is not possible since $t^B > (1-p^B)q^B$ by construction of the benchmark policy. Thus, $BR_i(r) = r$ and (r,r) is the Nash equilibrium for the interval in (91).

Last, but not least, suppose that

$$\alpha \geq \frac{u'(w^H - t^B)}{u'(w^L - t^B + q^B)}, \tag{92}$$

so that positive transfers take place for any action profile (d_1, d_2). By the same argument as in the previous case, (r, r) is a Nash equilibrium, since $BR_i(r) = r$. Now, I must show that $BR_i(a) = r$, i.e. (a, a) is not an equilibrium, but as before, since t^B is inefficiently high from the perspective of an individual agent, i can profitably deviate to $d_i = r$ when $d_j = a$. Thus, (r, r) is the unique Nash equilibrium for the interval in (92) when C^B is offered by the principal.

C.8 Proof of Proposition 7

This section generalizes the argument made when deriving the optimal benchmark contract, where I have shown that a risk-neutral principal can offer an insurance policy to a single self-interested agent that will yield that principal positive expected profit, and expected utility above the autarky one for the agent.

I will impose symmetry in the wealths of the agents, their respective degrees of altruism, the contract offered by the firm, and, finally, in the equilibrium behavior of the agents. Let $\overline{U}$ the agents' reservation utility. There are two system of equations implicitly defining premium as a function of coverage (and degree of altruism) to be considered: $\widehat{F}(t, p, T; q, \alpha)$ for the agents and $\widetilde{F}(t, p, T; q, \alpha)$ for the principal. Each system is composed by three equations:

$$\begin{aligned}
\widehat{F}_1 &= (1+\alpha)\left\{p^2 u(w^H - t) + (1-p)^2 u(w^L - t + q)\right. \\
&\qquad \left. + p(1-p)\left[u(w^H - t - T) + u(w^L - t + q + T)\right]\right\} - \overline{U}, \qquad (93) \\
\widetilde{F}_1 &= 2\left[t - (1-p)q\right], \qquad (94) \\
F_2 &= \psi'(p) - \left\{u(w^H - t - T) - u(w^L - t + q) + \alpha\left[u(w^L - t + q + T) - u(w^L - t + q)\right]\right\} \\
&\qquad + p(1+\alpha)\left[u(w^L - t + q + T) - u(w^L - t + q) - u(w^H - t) + u(w^H - t - T)\right], \qquad (95) \\
F_3 &= u'(w^H - t - T) - \alpha u'(w^L - t + q + T), \qquad (96)
\end{aligned}$$

where $\widehat{F}_1$ is the agents' indifference condition between accepting C or rejecting it, $\widetilde{F}_1$ is the principal's zero profit condition, F_2 determines the agents' symmetric equilibrium effort while F_3 determines transfer for any $\alpha \geq \widehat{\alpha}(\omega, C) \equiv \frac{u'(w^H-t)}{u'(w^L-t+q)}$. If $\alpha < \widehat{\alpha}(\omega, C)$, equation (96) is ignored in the systems and all transfers in equations (93)-(95) are set to zero.

Define the Jacobian matrix

$$J = \begin{pmatrix} \frac{\partial F_1}{\partial t} & \frac{\partial F_1}{\partial p} & \frac{\partial F_1}{\partial T} \\ \frac{\partial F_2}{\partial t} & \frac{\partial F_2}{\partial p} & \frac{\partial F_2}{\partial T} \\ \frac{\partial F_3}{\partial t} & \frac{\partial F_3}{\partial p} & \frac{\partial F_3}{\partial T} \end{pmatrix} \tag{97}$$

for each system $F(\cdot)$ and notice that

$$\det(\widetilde{J}) > 0 > \det(\widehat{J}) \tag{98}$$

if $\psi''(\cdot) > 0$ is sufficiently large. Therefore, by the Implicit Function Theorem, the implicit functions (t, p, T) are well-defined and continuously differentiable with respect to (q, α), and their derivatives must satisfy

$$\begin{pmatrix} \frac{dt}{dq} \\ \frac{dp}{dq} \\ \frac{dT}{dq} \end{pmatrix} = -J^{-1} \begin{pmatrix} \frac{\partial F_1}{\partial q} \\ \frac{\partial F_2}{\partial q} \\ \frac{\partial F_3}{\partial q} \end{pmatrix}. \tag{99}$$

For any $\alpha \in [0, 1]$, I want to show that

$$\left.\frac{d\widehat{t}}{dq}\right|_{(q,t)=(0,0)} > \left.\frac{d\widetilde{t}}{dq}\right|_{(q,t)=(0,0)}, \tag{100}$$

i.e., that the marginal rate of substitution between premium and coverage is steeper than the slope of the zero profit line for the principal in a neighborhood of zero, and, thus, that there exists gains of trade to be had in the interaction between the equally altruistic agents and the principal.

First, notice that

$$
\begin{aligned}
ZPL'|_{(0,0)} &= -\frac{1}{det(\widetilde{J})}\left[\frac{\partial \widetilde{F}_1}{\partial q}\frac{\partial F_2}{\partial p}\frac{\partial F_3}{\partial T} - \frac{\partial \widetilde{F}_1}{\partial p}\frac{\partial F_2}{\partial q}\frac{\partial F_3}{\partial T} + \left(\frac{\partial \widetilde{F}_1}{\partial p}\frac{\partial F_2}{\partial T} - \frac{\partial \widetilde{F}_1}{\partial T}\frac{\partial F_2}{\partial p}\right)\frac{\partial F_3}{\partial q}\right]\Bigg|_{(0,0)} \\
&= -\frac{\left(\frac{\partial \widetilde{F}_1}{\partial q}\frac{\partial F_2}{\partial p}\frac{\partial F_3}{\partial T}\right)\Big|_{(0,0)}}{2\left(\frac{\partial F_2}{\partial p}\frac{\partial F_3}{\partial T}\right)\Big|_{(0,0)}} \\
&= -\frac{-2(1-p)}{2} \\
&= 1-p, \qquad (101)
\end{aligned}
$$

since $\frac{\partial \widetilde{F}_1}{\partial T} = 0$ and $\left.\frac{\partial \widetilde{F}_1}{\partial p}\right|_{(0,0)} = 2q|_{(0,0)} = 0$.

References

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Alchian, A. A. and H. Demsetz (1972, December). Production , Information Costs, and Economic Organization. *American Economic Review 62*(5), 777–795.

Alesina, A. and P. Giuliano (2014). Family Ties. In *Handbook of Economic Growth*, Volume 2 of *Handbook of Economic Growth*, Chapter 4, pp. 177–215. Elsevier.

Alger, I. and J. W. Weibull (2010, September). Kinship, Incentives, and Evolution. *American Economic Review 100*(4), 1725–1758.

Alger, I. and J. W. Weibull (2013, November). Homo moralis - preference evolution under incomplete information and assortative matching. *Econometrica 81*(6), 2269–2302.

Alger, I. and J. W. Weibull (2016). Evolution and kantian morality. *Games and Economic Behavior 98*(C), 56–67.

Alger, I. and J. W. Weibull (2017, Sep). Strategic behavior of moralists and altruists. *Games 8*(3), 38.

Andreoni, J. (1990, June). Impure altruism and donations to public goods: A theory of warm-glow giving? *Economic Journal 100*(401), 464–77.

Arnott, R. and J. Stiglitz (1991). Moral hazard and nonmarket institutions: Dysfunctional crowding out or peer monitoring? *American Economic Review 81*(1), 179–90.

Attar, A. and A. Chassagnon (2009, September). On moral hazard and nonexclusive contracts. *Journal of Mathematical Economics 45*(9-10), 511–525.

Attar, A., T. Mariotti, and F. Salanié (2011, November). Nonexclusive Competition in the Market for Lemons. *Econometrica 79*(6), 1869–1918.

Attar, A., T. Mariotti, and F. Salanié (2014, January). Nonexclusive competition under adverse selection. *Theoretical Economics 9*(1).

Attar, A., T. Mariotti, and F. Salanié (2017, March). Entry-Proofness and Market Breakdown under Adverse Selection. TSE Working Papers 17-788, Toulouse School of Economics (TSE).

Attar, A., T. Mariotti, and F. Salanié (2019, February). Entry-Proofness and Market Breakdown under Adverse Selection. TSE Working Papers 19-788, Toulouse School of Economics (TSE).

Azam, J.-P. and F. Gubert (2006, December). Migrants' Remittances and the Household in Africa: A Review of Evidence. *Journal of African Economies 15*(2), 426–462.

Becker, G. S. (1974, Nov.-Dec.). A theory of social interactions. *Journal of Political Economy 82*(6), 1063–93.

Becker, G. S. (1976, September). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature 14*(3), 817–26.

Bergström, T. C. (1995, March). On the evolution of altruistic ethical rules for siblings. *American Economic Review 85*(1), 58–81.

Bernheim, B. D. and O. Stark (1988, December). Altruism within the Family Reconsidered: Do Nice Guys Finish Last? *American Economic Review 78*(5), 1034–1045.

Besley, T. (1995, Summer). Nonmarket Institutions for Credit and Risk Sharing in Low-Income Countries. *Journal of Economic Perspectives 9*(3), 115–127.

Biener, C., M. Eling, A. Landmann, and S. Pradhan (2018). Can group incentives alleviate moral hazard? the role of pro-social preferences. *European Economic Review 101*, 230 – 249.

Bisin, A. and D. Guaitoli (2004, Summer). Moral Hazard and Nonexclusive Contracts. *RAND Journal of Economics 35*(2), 306–328.

Bourlès, R., Y. Bramoullé, and E. Perez-Richet (2017). Altruism in networks. *Econometrica 85*(2), 675–689.

Bourlès, R., Y. Bramoullé, and E. Perez-Richet (2018, November). Altruism and Risk Sharing in Networks. Working Papers halshs-01943862, HAL.

Bramoullé, Y. and R. Kranton (2007). Risk-sharing networks. *Journal of Economic Behavior & Organization 64*(3-4), 275–294.

Brown, J. R. and A. Finkelstein (2011, Fall). Insuring Long-Term Care in the United States. *Journal of Economic Perspectives 25*(4), 119–142.

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

Coate, S. and M. Ravallion (1993, February). Reciprocity without commitment : Characterization and performance of informal insurance arrangements. *Journal of Development Economics 40*(1), 1–24.

Costa-Font, J. (2010, Winter). Family ties and the crowding out of long-term care insurance. *Oxford Review of Economic Policy 26*(4), 691–712.

Costa-Font, J. and C. Courbage (2015, March). Crowding Out of LongTerm Care Insurance: Evidence from European Expectations Data. *Health Economics 24*, 74–88.

Cox, D. and M. Fafchamps (2008, January). *Extended Family and Kinship Networks: Economic Insights and Evolutionary Directions*, Volume 4 of *Handbook of Development Economics*, Chapter 58, pp. 3711–3784. Elsevier.

Cox, D., E. Galasso, and E. Jiminez (2006, January). Private Transfers in a Cross Section of Developing Countries. Working Papers, Center for Retirement Research at Boston College wp2006-1, Center for Retirement Research.

Cremer, H. (2014, February). The Economics of Long-Term Care: An Introduction. *The B.E. Journal of Economic Analysis & Policy 14*(2), 339–342.

Cremer, H., F. Gahvari, and P. Pestieau (2013, July). Endogenous Altruism, Redistribution, and Long-Term Care. *The B.E. Journal of Economic Analysis & Policy 14*(2), 499–524.

Cremer, H., F. Gahvari, and P. Pestieau (2017). Uncertain altruism and the provision of long term care. *Journal of Public Economics 151*(C), 12–24.

Cremer, H., P. Pestieau, and K. Roeder (2016). Social long-term care insurance with two-sided altruism. *Research in Economics 70*(1), 101–109.

Cremer, H. and K. Roeder (2014). Transfers within a three generations family: When the rotten kids turn into altruistic parents. *Economics Letters 124*(3), 392–395.

Cremer, H. and K. Roeder (2017, June). Long-term care policy with lazy rotten kids. *Journal of Public Economic Theory 19*(3), 583–602.

De Donder, P. and P. Pestieau (2017, February). Private, Social, and Self-Insurance for Long-Term Care in the Presence of Family Help. *Journal of Public Economic Theory 19*(1), 18–37.

Di Falco, S. and E. Bulte (2013). The impact of kinship networks on the adoption of risk-mitigating strategies in ethiopia. *World Development 43*, 100 – 110.

Dubois, P., B. Jullien, and T. Magnac (2008). Formal and informal risk sharing in ldcs: Theory and empirical evidence. *Econometrica 76*(4), 679–725.

Ehrlich, I. and G. S. Becker (1972, July-Aug.). Market Insurance, Self-Insurance, and Self-Protection. *Journal of Political Economy 80*(4), 623–648.

Einav, L. and A. Finkelstein (2018). Moral Hazard in Health Insurance: What We Know and How We Know It. *Journal of the European Economic Association 16*(4), 957–982.

Fafchamps, M. and S. Lund (2003, August). Risk-sharing networks in rural Philippines. *Journal of Development Economics 71*(2), 261–287.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Foster, A. D. and M. R. Rosenzweig (2001, August). Imperfect Commitment, Altruism, And The Family: Evidence From Transfer Behavior In Low-Income Rural Areas. *The Review of Economics and Statistics 83*(3), 389–407.

Gruber, J. and K. Simon (2008, March). Crowd-out 10 years later: Have recent public insurance expansions crowded out private health insurance? *Journal of Health Economics 27*(2), 201–217.

Hendren, N. (2013, September). Private Information and Insurance Rejections. *Econometrica 81*(5), 1713–1762.

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

Itoh, H. (2004). Moral hazard and other-regarding preferences. *The Japanese Economic Review 55*(1), 18–45.

Jain, S. (1999). Symbiosis vs. crowding-out: the interaction of formal and informal credit markets in developing countries. *Journal of Development Economics 59*(2), 419 – 444.

Klimaviciute, J., P. Pestieau, and J. Schoenmaeckers (2019, April). Family altruism and long-term care insurance. *The Geneva Papers on Risk and Insurance - Issues and Practice 44*(2), 216–230.

Kocherlakota, N. R. (1996). Implications of Efficient Risk Sharing without Commitment. *Review of Economic Studies 63*(4), 595–609.

Kranton, R. E. (1996, September). Reciprocal Exchange: A Self-Sustaining System. *American Economic Review 86*(4), 830–851.

Ligon, E., J. P. Thomas, and T. Worrall (2002). Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies. *Review of Economic Studies 69*(1), 209–244.

Lindbeck, A. and J. W. Weibull (1988, December). Altruism and Time Consistency: The Economics of Fait Accompli. *Journal of Political Economy 96*(6), 1165–1182.

Pauly, M. V. (1974). Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection. *The Quarterly Journal of Economics 88*(1), 44–62.

Rey-Biel, P. (2008, 06). Inequity aversion and team incentives. *Scandinavian Journal of Economics 110*(2), 297–320.

Rotemberg, J. J. (2006). Altruism, reciprocity and cooperation in the workplace. In S. Kolm and J. M. Ythier (Eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Volume 1, Chapter 21, pp. 1371–1407. Elsevier.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics 90*(4), 629–649.

Sarkisian, R. (2017). Team incentives under moral and altruistic preferences: Which team to choose? *Games 8*(3), 37.

Townsend, R. (1994). Risk and insurance in village india. *Econometrica 62*(3), 539–91.

Udry, C. (1990). Credit markets in northern nigeria: Credit as insurance in a rural economy. *World Bank Economic Review 4*(3), 251–69.

Udry, C. (1994). Risk and insurance in a rural credit market: An empirical investigation in northern nigeria. *The Review of Economic Studies 61*(3), 495–526.

von Siemens, F. (2011). Heterogeneous social preferences, screening, and employment contracts. *Oxford Economic Papers 63*(3), 499–522.