Platform-based open source software development under double moral hazard*

Tao Li† Zheng Liu‡ Lei Ma§ Liying Wang¶ Bo Yang‖ Chaonan Yi**

September 17, 2025

Abstract

The Open Source Software (OSS) model has emerged as a dominant digital interface for knowledge creation. However, the OSS project owner and contributors form a relationship which is subject to double-sided 'hidden action' moral hazard. We formalise this situation using a simple principal-agent model containing two key features of decentralised platform-based open-source innovations: double moral hazard and limited individual rationality under which we introduce an additional participation constraint allowing for low-cost input in OSS collaborations. Our findings reveal that an OSS contributor is prone to moral hazard behaviours such as free riding if private incentives faced by the agent are not aligned with the interests of the project. From the perspective of the principal, moral hazard behaviours can also emerge with the incentives arising from several primary activities. Our study provides an important theoretical basis for understanding the source of market failures in open innovation processes and the mechanism that can potentially curb risk-taking incentives. This research contributes nuanced perspectives on policy responses, experiment designs, and the socio-economic implications of open-source innovations.

JEL Classification: L86; M11; O32; O36

Keywords: Open source software; Platforms; Double moral hazard; Limited individual rationality; Principal-agent model

*We acknowledge financial support for this research from the National Social Science Foundation of China (NSSFC) [Grant Number: 19ZDA078]. Declarations of interest: None.

†School of Artificial Intelligence Technology, Guangxi Technological College of Machinery and Electricity; Centre for Innovation and Development, Nanjing University of Science and Technology, China. E-mail: taolee@njust.edu.cn

‡Greenwich Business School, University of Greenwich, United Kingdom. E-mail: z.liu@greenwich.ac.uk

§School of Public Affairs, Nanjing University of Science and Technology; Institute for Digital Economy, Nanjing University of Science and Technology, China. E-mail: maryma2009@njust.edu.cn

¶China Institute for Small and Medium Enterprises, Zhejiang University of Technology, China. E-mail: wangliying@zjut.edu.cn

‖Department of Economics, School of Social Sciences, Swansea University, Singleton Park, Swansea, SA2 8PP, United Kingdom. E-mail: bo.yang@swansea.ac.uk (corresponding author)

**School of Intellectual Property, Nanjing University of Science and Technology, China. E-mail: njlgdx10095@njust.edu.cn

1 Introduction

Recent years have seen a rapid pace of digitalisation which in turn has played a significant role in driving fundamental changes in the industrial and service sectors. Indeed, such transformations can be expounded from a business model perspective on embedding technologies for value creation and enhancing innovative interdependencies through structural ecosystems. However, this new era of digital economy also implies that the recent severe shocks combined with post-COVID disruptions have raised significant challenges for the urgent research need for policy regulations and effective governance that serve in critical and recovery times.

Under suitable circumstances, the Open Source Software (OSS) model has emerged as a dominant digital interface for knowledge incubation and value creation. However, during this collaborative innovation process, the OSS project owner and contributors are prone to double-sided moral hazard behaviours.[1] In the context of global debates surrounding open-source technologies, particularly in the face of increasing uncertainty about decentralised peer-to-peer innovative systems, this paper addresses two critical research questions: How does moral hazard affect the innovation process of OSS platforms in agency relationships? What is the role of governance in realising the benefits of OSS developments while disciplining moral hazard?

On the technological front, an innovator, be it an individual, a firm or a consortium, may freely reveal the source code for the product which becomes a public good, allowing necessary adjustments for programmers' ability and the needs of the users. As for new business models, digital transformations such as OSS in the open innovation paradigm have led to the emergence of commons-based digital platforms. The latter can be regarded as an extension of the two-sided market, serving as an interface among a heterogeneous group of innovators that facilitates transactions and value-creation exchange (Evans, 2003; Rochet and Tirole, 2006; Boudreau, 2010; Gawer, 2010; Iansiti and Lakhani, 2020). Indeed, open innovation in this sense is built on free or selective revealing of product designs or code created and maintained by a (virtual) community of software developers collaboratively through projects and peer production.

Developing OSS entails activities of knowledge production that are made available to the public in order to elicit development collaboration (Henkel, 2006) and innovators making private gains while contributing to the collective good (von Hippel and von Krogh, 2006). During such activities, profit appropriation arises in agency relationships in which the OSS project owner (originator) acts as a principal and crowd-sources the open-source innovation task to one or several contributors (the agent). Coordination failures due to moral hazard in market interactions can be prevalent in such a principal-agent relationship. The literature on principal-agent models typically studies the production process in which the agent provides the costly effort that is unobservable which is therefore prone to the risk of moral hazard (one-sided) in the presence of information asymmetry. However, under the private-collective relationship within an OSS platform, both parties are subject to 'hidden action' moral hazard (double-sided) and

[1] An underlying assumption in our analysis is that monetary rewards are available from a regulator who is interested in the provision of an OSS service and the OSS participation encompasses not just commercial developers (the core team) but also hobbyists (volunteers). In most real-world settings, cooperation failures due to moral hazard in incomplete or relational contracts can arise as OSS communities can be sustained by the value of future relationships but are fraught with uncertainties and contingencies. Codes of conduct may vary from one environment to another (Gibbons, 2005).

tend to exert noncontractible and costly inputs for running the projects.

For example, on the one hand, such openness implies that actions of the contributors such as diversion of inputs or side-selling can cause reputational damages, on the other, the owner (sometimes a consortium) whose costly effort is not fully observable to the contributors maximises their private gains by shirking on their commitments, while both parties abiding by the applicable OSS license. Furthermore, in an online environment, participants can hide their identities, thus, a project owner may face an adverse selection problem, lacking information about the agent's characteristics or type. Unfortunately, there is often a lack of standards pertaining to social norms, ethical guidelines, and the effective governance approach for OSS platforms (Mateos-Garcia and Steinmueller, 2008).

While moral hazard has been investigated regarding the pricing structure of platforms and inter-organisational knowledge transfer,[2] studies that explicitly assess its impact on platform-based innovation processes are non-existent.[3] In the context of double-sided moral hazard, a series of papers have separately addressed its issues in determining contract terms with a number of applications.[4] While there are some recent attempts at devising a tractable model of double moral hazard in agency relationships on contracting (Dai and Toikka, 2022; Carroll and Bolte, 2023), research focusing on its applications is relatively sparser. This is a significant omission, as this problem has been associated with generating market failures within innovative systems, and advancing our understanding requires a functional model of innovation to capture the dynamic and interdependent relationship between the OSS participants.

Thus, our paper fills this gap. We investigate if and under what conditions openness leads to moral hazard behaviours. Our analysis employs a game-theoretic model with double moral hazard to capture the agency relationship in that we make both methodological and substantive contributions by explicitly accounting for (1) specifications of conditions that capture the risk of double moral hazard in the process of OSS developments, (2) an assessment for the governance approaches accounting for the issue in open collaboration and peer production, and (3) the role of institutions in minimising moral hazard. Our findings provide a theoretical basis pertaining to double moral hazard in OSS processes. Practically, the model provides important implications for practitioners and policymakers for considering the sustainable development strategies and regulations that are specific to the evolution of open-source ecosystems.

In addition, another distinctive aspect of our analysis is our assumption regarding rationality in decision-making processes. The defining characteristic of what we refer to as OSS platform participation is to limit the cognitive skills of participants in the model and this is achieved by introducing a simple agent's constraint which can be thought of a behaviour bias that occurs when their cognitive capacity may not be fully rational. We depart from the restrictive assumption faced by the principal in optimal risk sharing and introduce a participation constraint of

[2] See, for example, Caillaud and Jullien (2003); Hagiu (2005); Armstrong (2006); Roger and Vasconcelos (2014); Spulber (2018) and Chen and Li (2021).

[3] Granstrand (2024) provides an up-to-date review on the theoretical approaches to ecosystem analysis and discusses a formal approach to innovation ecosystems (platform-based innovations) based on systems theory and cooperative game theory. Similar to our modelling framework, the analysis in Granstrand (2024) specifies a value creation function involving inputs from small numbers of inter-related innovators. Liu *et al.* (2021) investigate how digital platforms like Uber can mitigate (one-sided) moral hazard based on a buyer-seller model.

[4] An early application where double moral hazard arises is in franchising (Bhattacharyya and Lafontaine, 1995) or warranties (Cooper and Ross, 1985).

low-input-type that needs to be met in the individual's decision problem.

To our knowledge, this paper represents the first attempt at evaluating the implications of coordination failures within an OSS platform by developing a framework of sequential-move structure that involves the costly efforts by each party that are difficult to observe or monitor. Thus, we make contributions to the existing literature by discussing a range of practical and theoretical implications, shedding more light on the underlying mechanism and incentive arrangements that guide policy choices for a successful collaborative project and developing sustainable development strategies for the open-source R&D-intensive sector. Our analysis provides novel insights into policy recommendations, while also being tractable and easily adaptable to designing experiments of enforcement mechanisms and behavioural uncertainties.

Based on the equilibrium solutions (with double moral hazard), further discussions can be conducted from three theoretical and policy perspectives as follows.

First, when the *ex post* participation constraint is binding, the payment of fixed compensation and lack of incentive gradients by the owner motivates them to exert their second-best efforts, while contributing to competing technologies to gain additional income is an incentive for the contributors which undermines the IPR of the OSS project due to the lack of incentive gradients (e.g., fixed compensation). These interaction dynamics of the inputs exerted by both parties explain the reduced effort under moral hazard which could create spillover effects on the project and its community and in many areas that impact productivity at the economy level.

Second, under a limited rationality participation constraint, the model extrapolates that the utilities of owners and contributors are opposite in monotonic nature, but they are able to coexist upon their platform. We find that this is the case that mimics what is observed in reality and represents a middle ground where incentives for private and collective gains coexist. First-order and higher-order free-riding arises for the contributors which makes it more expensive for the owner to exercise monitoring. Cost savings in monitoring leads to increased imbalance of incentives for innovation.

Third, under the incentive-compatibility constraint, a combination of positive and negative incentives exhibits an incentive gradient; however, negative incentives lead to reputational losses from the principal side. The elimination of negative incentives can also fulfill incentive-level differentiation. In typical open-source environments, incomplete information increases the cost of observing and monitoring the contributors' inputs. The owner saves on monitoring costs and compensates the contributor based on the average costs of participation. The contributors turn to other competing OSS firms or proprietary software vendors for additional income, and use technical vulnerabilities and documentation for unfair competition once the costs of participation are not covered.

The rest of the paper is structured as follows. Section 2 reviews related literature to provide context for our study. Section 3 outlines the baseline model and introduces the key assumptions. Based on Section 3, we consider and solve three variants of the model in Section 4 and explore the statistical validation of the model through numerical simulations. Section 5 discusses the results under the three variant formulations and compare the payoff matrices and welfare effects. Section 6 discusses the policy implications of the study in the context of OSS platforms, and finally, Section 7 concludes.

2 Related literature

There are three strands of literature related to our paper, which are discussed in this section. Additionally, our methodology draws from the principal-agent analysis and we discuss how our approach differs from the standard framework to further clarify the contribution of this paper.

2.1 Digital platform and open-source innovation processes

The first strand is a largely digital platform and open innovation literature. This literature focuses on the significant challenges and opportunities arising from the unique characteristics of digital platforms in general and OSS processes in particular. Recent examples of digital platforms include online auction, sharing economy platforms, electronic payment systems, and blockchain-based platforms. The major discussion under this strand of literature is around the challenges arising from the growing distribution of online communities across geographical regions and cultures, and the design theory and stability of digital platforms. A number of recent papers delve into the distinct design features of these platforms while others investigate governance mechanisms. Notably, Carrasco-Farre *et al.* (2022) assess the dynamic characteristics of the stakeholder value proposition design process. Leckel *et al.* (2025) empirically examine optimal platform interaction design. Spagnoletti *et al.* (2015) focus on the platform architecture that supports the social interaction structures of online communities. Empirical studies, including Springer *et al.* (2025), Nguyen and Nguyen (2025) and Uzunca and Kas (2023), examine the mechanisms employed in governing industrial business-to-business platforms, blockchain platforms, and digital labour platforms. Gawer and Harracá (2025) theorise inconsistent platform governance and discuss its implications for compliance behaviour.

This literature is also related to our paper by discussing the unique characteristics of OSS platforms. First, the main purpose of OSS is to share technologies and elicit collaboration, whereas the traditional view of innovation can mainly be described as a closed process in which technology is tradable on the market (i.e., private innovation or proprietary platforms). The studies on the latter usually make an implicit *a priori*: exclusivity is desirable for the innovator (Henkel, 2006), and assuming a market structure with monopolistic competition. Second, OSS participants are both creators and users of technologies who benefit from development support by a user community (Franke and Shah, 2003), but in two-sided markets, firms or individuals are usually distinguishable as either a buyer or a seller of their innovations (Rochet and Tirole, 2006). Third, OSS platforms sometimes rely on voluntary contributions and free revealing of user innovations (Henkel, 2006; von Hippel and von Krogh, 2006) - this is again very different from a traditional market structure that utilises pricing and profit-sharing rules for coordination and contract enforcement (Spulber, 2018). The major discussion under this strand of literature is around the issues of information trading, asymmetry and disclosure, and contractual incompleteness, and the strategic use of R&D investment and development for intellectual or human capital. This motivates us to take account of the concerns that arise in agency relationships.

2.2 Studies on open-source software development

Our research is also motivated by three related studies: von Hippel and von Krogh (2006), Krishnamurthy *et al.* (2014) and Islam *et al.* (2017). Theoretical studies, including von Hippel and von Krogh (2006) and Krishnamurthy *et al.* (2014), employing a benchmark private-collective innovation model, reveal substantial evidence of a mix of private and collective gains in open innovation processes. This literature also expresses concern about the sustainability of the OSS movement and incentives. The empirical study of Islam *et al.* (2017), focusing on the implementation of a decentralised control system in OSS projects, compares the private costs for participation with the average level of contribution. Using firm-level data, Petralia (2025) provides evidence that firms synchronize their internal innovative processes and their contributions to OSS projects. A common theme emerges from the open-source literature. Within the OSS communities, neither the social regimes implied by collective actions which would eliminate free-riding nor the incentives to innovate generated by appropriation for profits from innovation can adequately explain the OSS phenomenon.

Indeed, recent studies such as James *et al.* (2023) show that the inter-personal interactions in an innovative system have the potential to give rise to ethical issues, especially during periods of stress, crisis episodes and uncertainties. The lack of reliable institutions to resolve the disputes in coordination arrangements on open-source platforms has raised serious concerns among practitioners and academic researchers. Within such platforms, there is often no enforceable contract that stipulates the level of effort required by innovators who do not always abide by a set of rules or less formal arrangements codified in social or enforcement norms.

The present paper is then motivated by the observation that the existing literature has not come to an agreement on the dynamic relationship between the OSS owner and contributors. Furthermore, open innovation and real sector innovation are likely jointly determined. OSS user innovations typically exhibit strong heterogeneity of needs (Franke and von Hippel, 2003) and the underlying technologies associated with OSS often exhibit high modularity (MacCormack *et al.*, 2006). Thus, accurately assessing the consequences of OSS interactions, and their impact on open innovation strategies, is not only a formidable task but of significant relevance to policy as the findings are often used in building large-scale models and behavioural experiments for policy guidance.

2.3 The moral hazard problem

Our paper also fits into the broader literature on moral hazard in the context of platforms and R&D management. Indeed, moral hazard and incentive contracts are ubiquitous. There is a large body of research investigating the problem in innovative projects, contracting and market designs. Bergmann and Friedl (2008) focus on the optimal contracts and performance measures to motivate product R&D managers, while Dechenaux *et al.* (2011) study the role of contracts in addressing inventor moral hazard and monitoring effort in the development of licensed inventions. Liu *et al.* (2021) investigate how the pricing schemes derived from technology-enabled market designs and self-enforcement mechanisms mitigate moral hazard. Furthermore, the problem of double moral hazard has recently been studied in the domain of knowledge management in terms of technology licensing (Mooi and Wuyts, 2021), delegated

R&D activities (Poblete and Spulber, 2017), cooperative innovation in a project-based supply chain (Cai and Choi, 2020), and inter-organisational knowledge sharing between knowledge senders and receivers (Chen and Li, 2021).[5]

Yet in many situations, output is determined by the joint efforts of both parties with incentive mechanisms based on having formal incentive contracts or profit-sharing rules. However, OSS platforms account for uncertainties from the environment and often operate under incomplete or relational or nonenforceable contracts and can be fraught with incentive problems. The nature of OSS interactions makes it almost impossible to conditional on all the variables in explicit contracts (Niedermayer, 2013). The challenge of asymmetric information in the provision of project service is evident, as both the OSS project innovators can pursue their private gains to the detriment of the other during their activities. However, the literature has not systematically examined the dynamic interactions among the participants involved within the OSS development-based community, subjecting the relationship to moral hazard from both the owner and contributors, which is the focus of this paper.

2.4 Outline of contributions

Building on the earlier contributions by Arnott and Stiglitz (1988) and recent studies such as Rey and Stiglitz (2024), there are two departures from the standard model that lead to interesting results. First, on the basis of assuming cognitive limitations in decision-making, our model is subject to a *limited rationality* participation constraint, in addition to the participation and incentive-compatibility constraints, that is specific to OSS participants. This case allowing open-source participation coexistence mimics what is observed in reality. We then explore the statistical validation of this modelling assumption through numerical simulations. Second, our player strategy sets encompass variations in the level of free-riding efforts for the contributors, as well as three different compensation schemes provided by the owner to elicit effort. Our formulation of discrete–time principal–agent problem is particularly relevant to the behaviour of participants that is affected by the defining characteristics of open-source environments.

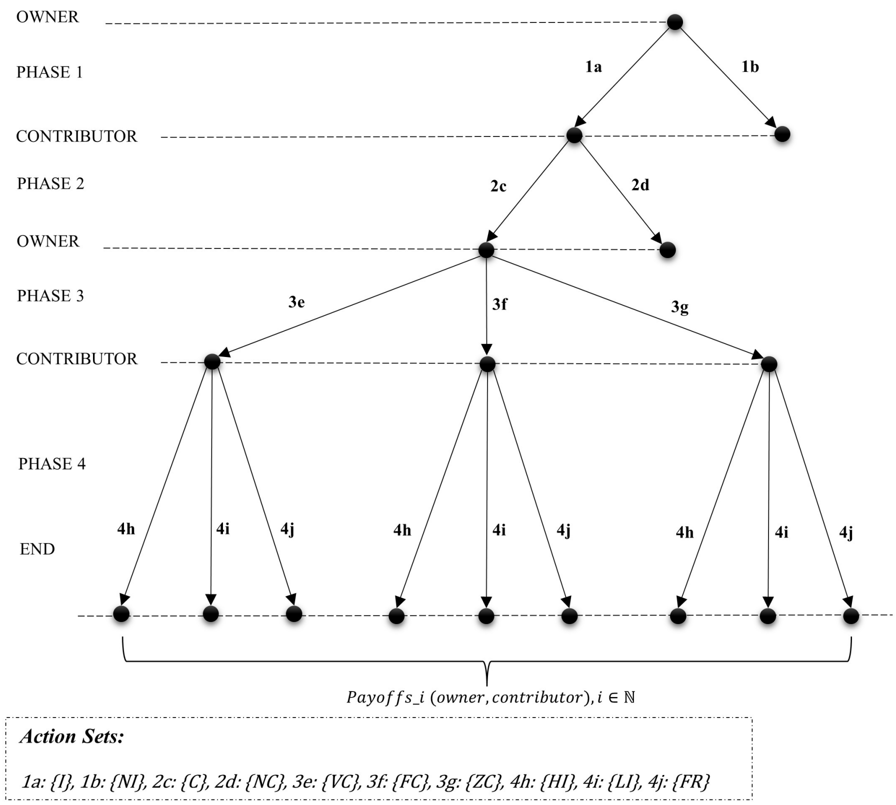
3 A principal-agent model with double moral hazard

We formulate the general framework of a principal-agent relation with double moral hazard, using the terminology in Rasmusen (2006), and describe the key components of our model and order of play. The model is depicted as a static game tree in Figure 1 with participants joining or leaving the platform at any time. Our model also captures the interaction between the decisions of each participant based on the information that they observe and the behaviours of others. Participants' goals and constraints may change over time and be affected by external factors, such as market demand, technology trends and competition, as well as internal factors such as personal goals or resources. In our development-based environment, neither the owner nor the

[5] Other typical applications where double-sided moral hazard tends to arise in contracting and profit/incentive sharing include franchising (Bhattacharyya and Lafontaine, 1995), warranties (Cooper and Ross, 1985), sharecropping (Eswaran and Kotwal, 1985), collaborative business services such as consulting (Roels *et al.*, 2010), and optimal contract schemes for subsidising public transit services (Wen *et al.*, 2023). A review is provided by Carroll and Bolte (2023).

contributors can exactly determine output *ex ante*, hence the terms of inputs and outputs are not fully contractible.

Figure 1: **Extended form of the game with double moral hazard**



3.1 Elements of the game

In what follows, we discuss details of the main modelling elements in the sequential one-shot game.

Participants: The OSS project owner and contributors represent the principal and the agent, respectively, with perfect but incomplete information.

Order of Play: Decisions of the owner and the contributors are made in an alternating manner and each step of decision is affected by the previous decisions and actions of the opponent. The game structure is divided into four stages (phases) with the owner starting the action first and the contributors following.

Actions: Both parties make costly contributions to the project and have their respective strategy sets at each of the following four phases. This is a model of hidden action and both parties can breach the agreement. The contributor has a last-mover advantage.

- Phase One: The owner makes a binary choice $\{\mathcal{I}, \mathcal{NI}\}$ ('Invite', 'Not invite'). $\{\mathcal{I}\}$ means that the project is released through the OSS platform to elicit collaboration. If the owner chooses $\{\mathcal{NI}\}$, then they develop the code in a closed-source, commercialised manner.

- Phase Two: The contributor makes a binary choice $\{\mathcal{C}, \mathcal{NC}\}$ ('Collaborate', 'Not collaborate'). $\{\mathcal{C}\}$ means that both parties form a principal-agent relationship (through an incomplete contract). The contributor can also sell the technologies in the market with the choice of $\{\mathcal{NC}\}$.

- Phase Three: The owner may provide an incentive scheme (compensation for the time invested) to the contributor and has a strategy set $\{\mathcal{VC}, \mathcal{FC}, \mathcal{ZC}\}$ ('Variable compensation', 'Fixed compensation', 'Zero compensation'). Different strategies imply different levels of incentives and the contract encompasses a linear incentive scheme. As the technology of the project is only known to the owner at this stage, they thus can under-declare output, subjecting the contract to moral hazard from the owner's side.

- Phase Four: The contributor has a strategy set $\{\mathcal{HI}, \mathcal{LI}, \mathcal{FR}\}$ ('Hight input', 'Low input', 'Free ride'). After combining strategies in the first three phases, the contributor's strategy choices generate a strategy space consisting of a combination of (3×3) 9 strategies. In contrast, as the owner does not know the contributor's type who now has the chance to review the project, the contributor can equally renege on their inputs once the owner has delivered the compensation.[6]

Payoffs: The strategic combinations each have their own corresponding payoffs and are denoted as $Payoffs_i\,(owner,\ contributor)\,,\ i \in \mathbb{N}$. Specifically, payoffs are associated with costs c, compensation s, utility u and earnings Π. Table 1 below summarises the notation used in the paper.

Costs: The OSS contribution incurs a cost $c_j \in \mathbb{R}^+$, where $j = ex$ or im, denoting an explicit cost and an implicit cost, respectively. c_{ex} are expenses pertaining to the job, where the economic value of exchange can be directly measured and observed. For example, materials or experimental equipment procured by a contributor to overcome a technical challenge are explicit costs, while the opportunity costs - foregoing the investment of that time in other more rewarding work - are implicit costs c_{im}. The cost function is expressed as

$$c_j = c\,(e) = \begin{cases} 0, & \text{if } e = 0 \\ c_j = c_{ex} + c_{im}, & \text{if } e = 1 \end{cases} \tag{1}$$

where c_j is a function of the contributor's action space e.

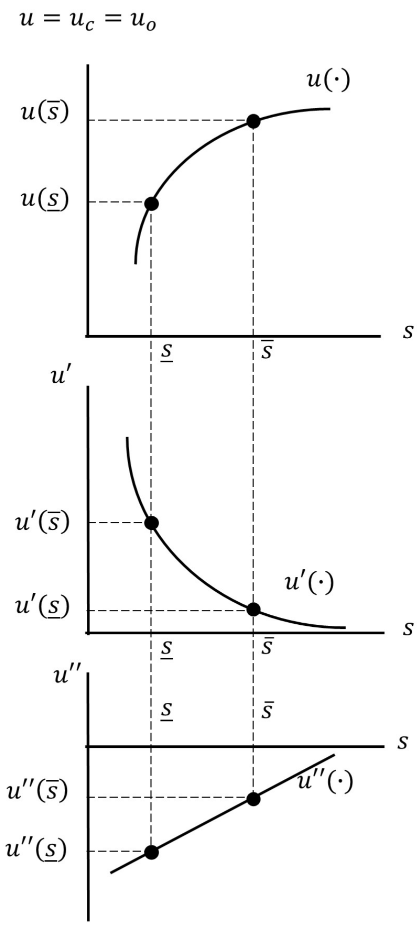
Compensation schemes: The owner evaluates the inputs and outputs of the contributors and then determines the compensation as $s \in \mathbb{R}^+$. Compensation is an expense for the owner, corresponding to the monetary rewards for co-contributing. The former has diminishing marginal utility of compensation $(u_o'\,(s) > 0,\ u_o''\,(s) < 0)$. Compensation is also a transfer payment, an

[6] As the co-contributor, just like any other users, has the right to review the OSS and its on-going development, our assumption holds that openness and the informational asymmetry lead to two-sided moral hazard and the model may not be reduced into a standard single moral hazard case or a signalling game.

incentive for the owner to pay the contributor and redistribute benefits.[7]

Utility function: The net benefit (utility) to the contributor, u_c, is defined based on s paid by the owner and c spent by the contributor. We therefore write the contributor's instantaneous utility as $u_c = E[u(s)|e] - c(e) = s - c$. Marginal utility of the contributor is mainly affected by the utility function $u_o(s)$ of the compensation paid by the owner. That is, $u'_c(s) = u'_o(s) > 0$, which shows a diminishing marginal utility, i.e., $u''_c(s) = u''_o(s) < 0$; $u'_c(s = e) = -c'(e)$. It follows from (1) that the agent is risk averse and dislikes effort. Figure 2 shows the properties of the utility functions.

Figure 2: **Utility and marginal utility functions**



Notes: $\underline{s}$ denotes the low level of compensation offered by the owner and $\bar{s}$ is the high level of compensation.

Earnings: The expected output is bounded with its value domain $q \in \{\underline{q}, \overline{q} \mid \underline{q} \geq 0, \overline{q} - \underline{q} > 0\}$. The OSS projects face market competition and they generate revenue $S(q) \in \mathbb{R}^+$. The con-

[7] Here, double moral hazard arises due to the sequential, contingent timing, as described in Figure 1. As noted above, the nature of the OSS platform implies that, through the one-shot strategic interaction, the equilibrium results in a non-enforced contract. Consequently, the contributors may act in a passive manner and fail to comply with the development obligation. In contrast, the OSS owner shrinks on their commitment, or privatises source codes for private gains, thereby creating moral hazard from both sides.

tributors that have different skill sets and knowledge endowments are subject to chance factors such as inspiration, trial and error, and luck. Thus, there is a nonlinear relationship between the inputs and outputs of contributors.

Information: In each phase, both parties face different decision points, i.e., nodes of the game tree (Figure 1). The information sets contain both the historical information and public knowledge of each node that each side possesses, based on which participants make decisions within their respective strategy spaces and then update the respective information sets.

It follows that there is also a degree of randomness between the OSS inputs and outputs. Let P_0 and P_1 be the probabilities of producing an output associated with two possible effort levels. At low effort level ($\underline{c}$), the probability of a contributor producing a high level of output $\overline{q}$ by supplying a low-cost input $\underline{c}$ is P_0. Thus, the probability of obtaining $\underline{q}$ from $\underline{c}$ is $1{-}P_0$. Similarly, at high level of effort ($\overline{c}$), the probability that a contributor invests a high cost $\overline{c}$ to obtain a high level of output $\overline{q}$ is P_1, while the probability that $\underline{c}$ is invested to produce $\underline{q}$ is $1{-}P_1$, and $1 > P_1 > P_0 \geq 0$, $\sum_{i=0}^{1} P_i = 1$. The owner makes compensation decisions based on the degree of output of contributors from the set of compensation strategies $(\underline{s}, \overline{s})$; Matching the outputs and incentives with positive correlations yields a set of incentive strategies $(\underline{q}, \underline{s}), (\overline{q}, \overline{s})$.

To determine the expected utility of the owner based on different output levels q, we use the following values functions

$$\underline{V_q} = P_0 \left[\Pi\left(\overline{q}\right) - \overline{s}\right] + (1 - P_0)\left[\Pi\left(\underline{q}\right) - \underline{s}\right], \quad c\left(\bullet\right) : e = 0 \rightarrow \underline{c} = 0 \tag{2}$$

$$\overline{V_q} = P_1 \left[\Pi\left(\overline{q}\right) - \overline{s}\right] + (1 - P_1)\left[\Pi\left(\underline{q}\right) - \underline{s}\right], \quad c\left(\bullet\right) : e = 1 \rightarrow \overline{c} = c \tag{3}$$

where $\Pi\left(\bullet\right)$ is the earnings function that quantifies the output q.

3.2 Relevant constraints

Constraints not only balance the interests and behaviours of different participants and ensure that their minimum needs are met but also provide the flexibility needed in order to cope with uncertainties and risks in mechanism design. In our model, we set out three relevant constraints that are specific to the OSS owner/principal who seeks to maximise (2) and (3).

Individual rationality or participation constraint: Each participant can receive at least a minimum level of income or return and their effort level is restricted on the expected utility of the principal. The source of the participation constraint is the need, or expectation, of the participants that pertains to the security of their rights and benefits that are in the contract, ensuring efficient allocations compatible with the agent's individual rationality.

Limited rationality participation constraint: This provides a degree of flexibility, or 'has slack', in the requirements of participants to achieve the minimum return that is required in order to meet the participation constraint. The existence of such constraints may allow participants to accept

Table 1: **Summary of notation used**

Variable	Description
e	Effort input—action space of contributor
$c; c\,(e)$	Total cost of contributor
$\underline{c}$	Low level of total cost
$\overline{c}$	High level of total cost
c_{ex}	Explicit cost
c_{im}	Implicit cost
s	Compensation from OSS project owner to contributor
$\underline{s}$	Low level of compensation
$\overline{s}$	High level of compensation
$u_o\,(\bullet)$	Owner's utility of compensation
$u_o'\,(\bullet)$	Owner's marginal utility of compensation
$u_o''\,(\bullet)$	Second-order derivative of owner's utility of compensation
$u_c\,(\bullet)$	Contributor's utility of compensation
$u_c'\,(\bullet)$	Contributor's marginal utility of compensation
$u_c''\,(\bullet)$	Second-order derivative of contributor's utility of compensation
q	Output of OSS knowledge innovation by contributor
$\underline{q}$	Low level of OSS output
$\overline{q}$	High level of OSS output
$S\,(q)$	Revenues generated by OSS project
P_0	The probability of contributor obtaining $\overline{q}$ by incurring $\underline{c}$
$1{-}P0$	The probability of contributor obtaining $\underline{q}$ by incurring $\underline{c}$
P_1	The probability that a contributor incurs $\overline{c}$ to obtain $\overline{q}$
$1{-}P1$	The probability that a contributor incurs $\underline{c}$ to obtain $\overline{q}$
V_q	Expected utility of owner based on various output levels of q
$\underline{V_q}$	Expected utility of owner based on $\underline{q}$
$\overline{\overline{V_q}}$	Expected utility of owner based on $\overline{q}$
$\Pi\,(\bullet)$	The earnings function that economically quantifies the output q
$\nabla = \left(\frac{\partial}{\partial \underline{s}}, \frac{\partial}{\partial \overline{s}}\right)'$	The surface gradient vector of the objective and constraint functions
u_e	Contributor's additional earnings from investing in competing OSS projects or commercial software projects

lower income or return and, in some cases, to exchange other advantages, or achieve other goals.

Incentive-compatibility constraint: The incentives in the contract should encourage and ensure that participants engage in high level of effort. The source of incentive-compatibility lies in the motivation of participants to maximise their own benefits, while also taking into account the impact of other participants' effort level and optimising behaviour.

3.3 Key assumptions

Before the model is analysed and conjectures are proposed, we make six assumptions attributed to agency theory including the standard assumptions on opportunism and information: Assumption 2 is an underlying assumption for the participation constraint faced by the principal. Assumptions 3-5 are the underlying assumptions for the limited rationality participation constraint, and Assumption 6 is the basic assumption for the incentive-compatibility mechanism in agency theory.

Coordination governance between OSS owners and contributors is characterised by considerable uncertainty which is behavioural and could be influenced by other social and environmental factors. Drawing on behavioural economics and prospect theory, we explain the biases that people use in individual decision problems and the rationale behind our departure from the standard individual rationality participation condition (i.e., Assumptions 3-5).

First, in the context of digital labour, prospect theory suggests that individuals are typically risk-averse and place more weight on perceived gains than perceived losses when deciding between alternatives where the probability of different outcomes is unknown. The loss aversion is closely related to social hedonic editing (Thaler, 1999) which suggests that the fulfilment of individual needs extends well beyond pure utility metrics to encompass broader well-being and avoid self-alienation (Hayry, 2021). In other words, open-source contributors identify a subjective reference point or benchmark, when evaluating digital tasks, and may prefer a smaller but certain reward on one platform over a riskier, high-reward one (see, for example, Leckel *et al.*, 2025; Madanaguli *et al.*, 2023).[8] They project long-term uncertain gains onto the present as non-economic incentives for decision-making (social capital-related) - such as reputational incentives yielding community recognition or skill incentives facilitating better career opportunities.

Second, we refer to relational and social contracting, motivated by the same reasons underlying coordination failures and contractual incompleteness, such as the inability to specify all possible contingencies. These informal agreements, or preferences for social exchange, sustained by the value of future relationships, are normally in the form of unwritten codes of conduct (Baker *et al.*, 2002; Gibbons, 2005; Leckel *et al.*, 2025). The nature of OSS interactions relying on information sharing within a group favours the use of such relational mechanisms, especially if the entire decision problem involves a repeated interaction in the first place and induces long-term relationship enforcement. The intuition is also related to social exchange theory (Molm *et al.*, 2000).

In both cases, the decision maker is not required to have access to full cognitive and compu-

[8] A comprehensive review by Madanaguli *et al.* (2023) provides insights, covering various theories and practice in assessing risks and uncertainties of open innovation.

tational abilities but can think ahead. Such prospects and preferences explain why contributors engage in open-source activities despite short-term losses in expected return. A simple way to evaluate the set of such problems is via a more flexible participation that requires an agent's expected gains from a low-effort input to be at least as great as the reservation utility.

Assumption 1. *Information is incomplete at each phase of the game and each participant is self-interested.*

Assumption 2. *Under the participant constraint (IR), participants are individually rational and have a minimum threshold of participation, typically requiring zero expected utility ($\bar{u} = 0$). This is the lowest level at which a participant does not choose to participate. The participation constraint also emphasises that participants pursue private gains and may choose to withdraw from the OSS platform if they do not receive at least the reservation utility.*

In other words, the reservation utility from a contract needs to meet the following participation constraint

$$\begin{aligned} E[u_c(s,c),c] &\geq \bar{u}_c \quad \textit{or} \\ [P_1 u_c(\overline{s}) + (1-P_1)\, u_c(\underline{s}) - \overline{c}] &\geq 0 \quad \textit{(IR)} \end{aligned} \tag{4}$$

Assumption 3. *Under the limited rationality participation constraint (LR), the idea is that rationality is limited when participants make decisions and participants may be willing to accept a lower immediate return in a particular scenario in exchange for long-term benefits (e.g., a relational contacting device).*

Assumption 4. *Under LR, participants are aware of the risks and uncertainties in the collaboration. In some cases, they may not get the expected rewards or benefits.*

Assumption 5. *Participants are likely to accept the contract under LR when said constraints are weighed against risk and uncertainty; they consider the possible consequences of uncertainty. In a given situation, they may voluntarily accept reduced rewards, or benefits, in order to minimise risks.*

The above constraint can then by summarised by a participation constraint for low-cost input ($\underline{c}$) in OSS contributions

$$\begin{aligned} E[u_c(s,c),c] &\geq \bar{u}_c \quad \textit{or} \\ [P_0 u_c(\overline{s}) + (1-P_0)\, u_c(\underline{s}) - \underline{c}] &\geq 0 \quad \textit{(LR)} \end{aligned} \tag{5}$$

Assumption 6. *The principal anticipates the agent's action and takes the optimising behaviour of the agent as a constraint. A contact that satisfies the incentive-compatibility constraint (IC) should ensure that the expected utility from the high-effort input needs to be at least as large as the expected utility that participants would obtain from the low-effort input.*

For the contract to be incentive compatible, the constraint is given by

$$\begin{aligned} c &\in \arg\max_{c'} E[u_c(s,c'),c'] \quad \textit{or} \\ [P_1 u_c(\overline{s}) + (1-P_1)\, u(\underline{s}) - \overline{c}] &\geq [P_0 u_c(\overline{s}) + (1-P_0)\, u_c(\underline{s}) - \underline{c}] \quad \textit{(IC)} \end{aligned} \tag{6}$$

4 The constrained optimisation problems

In what follows, we present the model variants based on the above assumptions, and using the expressions in the above section and profit functions in (2) and (3), we also derive and discuss the optimality conditions associated with each scenario.

4.1 The problem under participation constraint (Case 1)

The owner looks for incentives to motivate the contributors to engage while maximising the expected utility (profit). The compensation given to a contributor is $\bar{s}$ when the strategy set of the contributor is ensured to be $\{(e, c(e))\} = \{(1, \bar{c})\}$. That is, high compensation $\bar{s}$ motivates the contributors to supply a high level of input at a high level of costs. Our model assumes that compensation is monetary based on a particular open-source contribution, whereby the owner's expenditure and the contributor's income satisfy the zero-sum property but are equal in absolute value. This simplifying assumption ensures that both parties using money as a general equivalent for economic incentives allows us to quantify these preferences as a utility metric. In this case, the corresponding utilities for the owner and contributor are equal and denoted as $u_c(\bar{s}) = u_o(\bar{s})$.

If the contributor in (6) incurs zero costs of innovation, i.e., $\underline{c} = 0$, then, with activities that allow a certain degree of probability P_0 to innovate - such as free-riding - (6) becomes

$$[P_1 u_c(\bar{s}) + (1 - P_1) u(\underline{s}) - \bar{c}] - [P_0 u_c(\bar{s}) + (1 - P_0) u_c(\underline{s})] \geq 0 \tag{7}$$

If we consider the possibility of a policy that enforces the contract, the contributor cannot free-ride if they have incurred zero cost in developing the source code. The compensation in this case is zero and the participation constraint for the contributor is then given by (4).

If there is an employment relationship between the contributor and owner, then the process in the principal–agent relationship is fully contractual. In reality, however, the principal–agent relationship pertaining to the OSS development is an incomplete contract which makes it more difficult for the public sector to enforce compliance with the contract; thus making it possible for both parties to breach the agreement.

If the inputs and outputs of the contributors are completely observed, and the contributors do not free-ride or incur high costs in order to innovate, then the utility of the owner is maximised based on the set of incentive strategies $\{(\underline{q}, \underline{s}), (\bar{q}, \bar{s})\}$ for the objective function set out in (3) subject to (4)

$$\max_{\{(\underline{q},\underline{s}),(\bar{q},\bar{s})\}} \quad P_1 [\Pi(\bar{q}) - \bar{s}] + (1 - P_1) [\Pi(\underline{q}) - \underline{s}] \tag{8}$$

Since (8) is a multivariate extreme value problem subject to inequality constraints, the Lagrange multiplier method is employed for optimisation. This method uses the principle that the surface gradient vectors of the objective function and the constraint function point in the same direction. We next introduce the Lagrange multiplier λ, that scales the gradient of the constraint function, to examine the relationship between the level of compensation s and the cost of developing the source code c incurred by the contributor, by solving the constrainted optimisation of (8). The individual solves the Lagrangian problem which is

$$L = \max_{\{(\underline{q},\underline{s}),(\overline{q},\overline{s})\}} P_1 \left[\Pi\left(\overline{q}\right) - \overline{s}\right] + \left(1 - P_1\right)\left[\Pi\left(\underline{q}\right) - \underline{s}\right] + \lambda\left[P_1 u_c\left(\overline{s}\right) + \left(1 - P_1\right) u_c\left(\underline{s}\right) - \overline{c}\right] \tag{9}$$

First order conditions for a maximum are

$$\nabla = \begin{pmatrix} \frac{\partial L}{\partial \underline{s}} \\ \frac{\partial L}{\partial \overline{s}} \end{pmatrix} = \begin{pmatrix} -\left(1 - P_1\right) + \lambda\left(1 - P_1\right)\frac{\partial u_c(\underline{s})}{\partial \underline{s}} \\ -P_1 + \lambda P_1 \frac{\partial u_c(\overline{s})}{\partial \overline{s}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{10}$$

The following result is obtained from (10)

$$\begin{aligned} \frac{\partial u_c(\underline{s})}{\partial \underline{s}} &= u_c'\left(\underline{s}\right) = \frac{\partial u_c(\overline{s})}{\partial \overline{s}} = u_c'\left(\overline{s}\right) = \frac{1}{\lambda} \\ &\rightarrow \underline{s} = \overline{s} \end{aligned} \tag{11}$$

Since $u_c = u_o$, and there is monotonicity, (11) implies that the contributors receive a fixed level of compensation $s = \underline{s} = \overline{s}$ - regardless of whether they innovate or not - on the proviso that they incur the true cost of the OSS participation.

4.2 The problem under limited rationality participation constraint (Case 2)

Innovators in the form of individuals, or organisations, may participate in the OSS project within the platforms that are led by entities such as firms, non-profit NGOs, or governments. This section takes account of effective behaviour which departs from the assumption of rationality by introducing an additional participation constraint allowing for reduced levels of effort. Variations in the level of effort e by the contributors lead to differences in the high and low costs of inputs $\overline{s}$ or $\underline{s}$, as well as differences in the high- and low-outputs $\overline{q}$ or $\underline{q}$ and the corresponding high and low probabilities P_1 or P_0. We now consider the problem of maximising the objective function in this case subject to an inequality constraint of the form

$$\begin{aligned} \max_{\{(\underline{q},\underline{s}),(\overline{q},\overline{s})\}} \quad & P_1\left[\Pi\left(\overline{q}\right) - \overline{s}\right] + \left(1 - P_1\right)\left[\Pi\left(\underline{q}\right) - \underline{s}\right] + P_0\left[\Pi\left(\overline{q}\right) - \overline{s}\right] + \left(1 - P_0\right)\left[\Pi\left(\underline{q}\right) - \underline{s}\right] \\ \text{s.t.} \quad & \left[P_0 u_c\left(\overline{s}\right) + \left(1 - P_0\right) u_c\left(\underline{s}\right) - \underline{c}\right] \geq 0 \\ & P_1 - P_0 > 0,\ P_0 \geq 0 \end{aligned} \tag{12}$$

(12) can be solved by standard methods. The individual solves the Lagrangian problem which is

$$\begin{aligned} L = \max_{\{(\underline{q},\underline{s}),(\overline{q},\overline{s})\}} & P_1\left[\Pi\left(\overline{q}\right) - \overline{s}\right] + \left(1 - P_1\right)\left[\Pi\left(\underline{q}\right) - \underline{s}\right] + P_0\left[\Pi\left(\overline{q}\right) - \overline{s}\right] + \left(1 - P_0\right)\left[\Pi\left(\underline{q}\right) - \underline{s}\right] \\ & + \lambda\left[P_0 u_c\left(\overline{s}\right) + \left(1 - P_0\right) u_c\left(\underline{s}\right) - \underline{c}\right] \end{aligned} \tag{13}$$

First order conditions for a maximum are

$$\nabla = \begin{pmatrix} \frac{\partial L}{\partial \underline{s}} \\ \frac{\partial L}{\partial \overline{s}} \end{pmatrix} = \begin{pmatrix} -(1-P_1)-(1-P_0)+\lambda(1-P_0)\,u_c'(\underline{s}) \\ -P_1-P_0+\lambda P_0 u_c'(\overline{s}) \end{pmatrix} = \begin{pmatrix} -2+P_1+P_0+\lambda(1-P_0)\,u_c'(\underline{s}) \\ -P_1-P_0+\lambda P_0 u_c'(\overline{s}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{14}$$

Simplification gives the following results

$$u_c'(\underline{s}) = \frac{2-P_1-P_0}{\lambda(1-P_0)} = \frac{1}{\lambda}\left(1+\frac{1-P_1}{1-P_0}\right) \tag{15}$$

$$u_c'(\overline{s}) = \frac{P_0+P_1}{\lambda P_0} = \frac{1}{\lambda}\left(1+\frac{P_1}{P_0}\right) \tag{16}$$

From (15) and (16) we obtain

$$u_c'(\overline{s}) - u_c'(\underline{s}) = \frac{P_1-P_0}{P_0(1-P_0)} > 0 \tag{17}$$

4.3 Simulations of marginal utilities

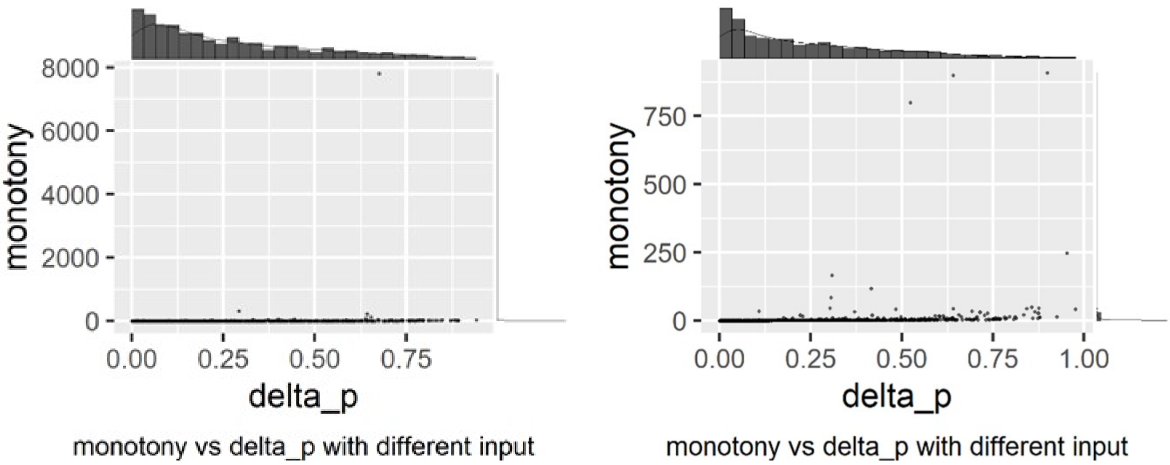
The behavioural bias that occurs under LR implies that the participants of OSS make satisfactory decisions rather than fully optimise. In Figure 3, we carry out numerical simulations for (17), the relationship between the probabilities P_0 and P_1 of the outputs of the contributor, with the probabilities drawn from the uniform and beta distributions, and the corresponding monotonicity of the utility. The monotonicity is found by (15) and (16). The probabilistic distance between a contributor's high- and low-output levels is represented by the x-axis; i.e., $\Delta P = P_1 - P_0$. Similarly, (17) gives the monotonicity difference under the payoffs that correspond to different outputs on the y-axis, which indicates that the difference is related to the contributor's output level.

Since $P_1 - P_0 > 0$, we simulate the model to provide the statistical validation of (17), in addition to characterising the results under both the uniform (Panel a) and normal distributions (Panel b).[9] The inclusion of marginal density histograms in Figure 3 provides a better view of the distribution of probability distances and monotonicity differences between the high and low contributor output levels. The simulations show that the probability distances of the contributor's output are more discrete under a uniform distribution and the monotonicity differences are more discrete under a normal distribution.

It is interesting to note that the utility function of the contributor intensifies in line with increasing compensation. This monotonically-increasing utility function can be explained by the observation that the contributors feel the additional benefit of the increased compensation;

[9] We start with a uniform distribution bounded between 0 and 1, i.e., fractions or probabilities, from which we draw randomly 1000 parameters. We then use (truncated) normal distributions as priors for the bounded parameters when more informative priors seem to be necessary and discard the parameter draws that are outside of $[0, 1]$. Our prior for the probabilities follows a normal distribution with a mean of 0.5 and a standard deviation of 0.1, which implies a prior of 0.5 for P_1 and P_0, respectively. This prior aligns closely with the intuition that the probability that the level of output produced is associated with the two possible effort levels, respectively, is around 50%. Details of the technical content, which contains the algorithm and the R source code for the numerical simulations, are available upon request.

Figure 3: **Numerical simulations of the marginal utility of a contributor**



(a) Uniform Distribution (b) Normal Distribution

Notes: Each panel shows monotonicity of $u_c(\bullet)$ under which the output probabilities P_0 and P_1 follow a uniform or normal distribution.

as compensation increases, the contributors become more sensitive to this increase, and the additional utility that may be derived from each additional unit of compensation is greater.

Nevertheless, the owner's utility $u_o(\bullet)$ monotonically increases while its first-order derivative monotonically decreases and is, therefore, inelastic. As noted, the owner's marginal utility $u_o'(\bullet)$ presents an opposite nature to the contributor's marginal utility $u_c'(\bullet)$. For the owner, their marginal utility decreases relatively quickly as they pay a higher compensation to the contributors. This is because, as compensation increases, the owner may face higher costs and financial pressures, such as increased investment in resources that is needed in order to maintain and expand the project, or it may trigger resentment from other contributors in terms of offering excessive compensation. This reflects the different focus of the two parties in terms of the trade-off between the benefits and risks of the reward.

4.4 Coexistence of open-source participation

The opposite monotonic nature of $u_o'(\bullet)$ and $u_c'(\bullet)$ does not negate (12) and its premises; rather, it demonstrates the coexistence of the utilities of both the owner and contributor under the realistic conditions of LR. The result of the opposite nature of their utility functions and mixed coexistence of actions mimics the reality of OSS platforms where the owner and contributor each have different views and preferences on compensation and cost inputs for innovation, and both parties find a balance of utility to carry out open collaboration. Even taking into account the condition that allows the contributors to free-ride, the open-source platform remains sustainable with varying types of utility functions often coexisting as long as the probability of producing output is higher for those contributors who actually invest in the costs of open-source innovation and also remains higher than the probability for the contributors who free-ride, i.e., $P_1 > P_0$.

In order for $P_1 > P_0$ to be strictly satisfied, the OSS project must have a certain level of complexity and institutional design and arrangement. From the theoretical perspective, under certain conditions, by modifying the assumption and constraint applied to the model, our simulation results represent a realistic middle ground where incentives for private and collective actions can coexist and the contributors obtain private benefits tied to the development of the public good, yielding higher private benefits to innovators than to free-riders. Our simulation results are generally in line with the findings in von Hippel and von Krogh (2006).

4.5 The problem under incentive-compatibility (Case 3)

Turning to the IC mechanism which is a higher-order type that can be achieved through the platform's own evolution, or initial mechanism design. The above section explains the coexistence of utilities under different types of open-source themes on the platform under the LR constraint. Here, *ex post* IC is concerned with the design of incentives by taking into account the optimising behaviour of the agent. In terms of basic incentive principles, the owner always desires to incentivise those contributors who have genuinely invested in the cost of innovating, to encourage original contributions, and to resist free-rider behaviours. The constraint thus presents an incentive gradient for the maximisation problem as follows

$$\begin{aligned} \max_{\{(\underline{q},\underline{s}),(\overline{q},\overline{s})\}} \quad & P_1\left[\Pi\left(\overline{q}\right)-\overline{s}\right]+\left(1-P_1\right)\left[\Pi\left(\underline{q}\right)-\underline{s}\right] \\ \text{s.t.} \quad & \left[P_1 u_c\left(\overline{s}\right)+\left(1-P_1\right)u_c\left(\underline{s}\right)-c\right]-\left[P_0 u_c\left(\overline{s}\right)+\left(1-P_0\right)u_c\left(\underline{s}\right)\right]\geq 0 \\ & P_1 u_c\left(\overline{s}\right)+\left(1-P_1\right)u_c\left(\underline{s}\right)-c\geq 0 \\ & P_1-P_0>0,\ P_0\geq 0 \end{aligned} \tag{18}$$

The individual solves the Lagrangian problem which is

$$\begin{aligned} L=\max_{\{(\underline{q},\underline{s}),(\overline{q},\overline{s})\}} & P_1\left[\Pi\left(\overline{q}\right)-\overline{s}\right]+\left(1-P_1\right)\left[\Pi\left(\underline{q}\right)-\underline{s}\right] \\ & +\lambda_1\left[P_1 u_c\left(\overline{s}\right)+\left(1-P_1\right)u_c\left(\underline{s}\right)-c\right]-\left[P_0 u_c\left(\overline{s}\right)+\left(1-P_0\right)u_c\left(\underline{s}\right)\right] \\ & +\lambda_2\left[P_1 u_c\left(\overline{s}\right)+\left(1-P_1\right)u_c\left(\underline{s}\right)-c\right] \end{aligned} \tag{19}$$

First order conditions for a maximum are

$$\begin{aligned} \nabla=\begin{pmatrix} \frac{\partial L}{\partial \underline{s}} \\ \frac{\partial L}{\partial \overline{s}} \end{pmatrix} &= \begin{pmatrix} -(1-P_1)+\lambda_1(1-P_1)u_c'(\underline{s})-\lambda_1(1-P_0)u_c'(\underline{s})+\lambda_2(1-P_1)u_c'(\underline{s}) \\ -P_1+\lambda_1 P_1 u_c'(\overline{s})-\lambda_1 P_0 u_c'(\overline{s})+\lambda_2 P_1 u_c'(\overline{s}) \end{pmatrix} \\ &= \begin{pmatrix} \left[\lambda_2-(\lambda_1+\lambda_2)P_1+\lambda_1 P_0\right]u_c'(\underline{s})-(1-P_1) \\ \left[(\lambda_1+\lambda_2)P_1-\lambda_1 P_0\right]u_c'(\overline{s})-P_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned} \tag{20}$$

Solving the constrained optimisation problem yields

$$u_c'\left(\underline{s}\right) = u_o'\left(\underline{s}\right)=\frac{1-P_1}{\lambda_2+\lambda_1 P_0-(\lambda_1+\lambda_2)P_1} \tag{21}$$

$$u_c'\left(\overline{s}\right) = u_o'\left(\overline{s}\right)=\frac{P_1}{(\lambda_1+\lambda_2)P_1-\lambda_1 P_0} \tag{22}$$

Given the constraints $P_1 - P_0 > 0$ and $P_0 > 0$, clearly, (22) is strictly positive. In other words, under high compensation $\overline{s}$, the utility of open-source contributions increases incrementally with each additional unit of digital labour (input) contributed to open-source activities, *ceteris paribus*, implying that high levels of incentive promote open-source engagement. (21), however, yields a negative value, implying that low-level incentives decrease overall satisfaction for the contributors, *ceteris paribus*. This may cause the contributors to develop a sense of aversion towards open-source activities. Combined with (22), this scenario forms a gradient where high incentives yield greater contributions, while low incentives result in fewer contributions or elicit lower performance - aligning with the incentive compatibility objectives.

Comparing the properties of $u'_c(\overline{s})$ and $u'_c(\underline{s})$ in (21) and (22), respectively, yields the following result

$$u'_c(\overline{s}) - u'_c(\underline{s}) = \frac{\lambda_1(P_1 - P_0)}{[(\lambda_1+\lambda_2)P_1 - \lambda_1 P_0]\,[\lambda_2 + \lambda_1 P_0 - (\lambda_1+\lambda_2)P_1]} < 0 \quad (23)$$

$$\Rightarrow u'_c(\underline{s}) = u'_o(\underline{s}) \quad (24)$$

Compensation is the expected payoff that the agent receives. (23) indicates that $u'_c(\bullet)$ and $u'_o(\bullet)$ both monotonically decrease. This result leads to (24), which implies that both utility functions remain consistent in their monotonic nature with the incentive-compatibility constraint, proving the existence of an incentive-compatibility mechanism.

The owner chooses the effort level taking into account the optimising behaviour of the agent (IC). Under the condition of an incentive-compatible, incomplete contract, where the monotonic nature of the utility of both parties is the same, the owner tends to pay the lowest possible compensation. The level of compensation that is paid for the high-cost innovations ultimately tends to converge at the level of compensation paid for low-cost innovations. Allowing this limit to exist, a constant transformation of the constraint in (18) yields the following conditions

$$u_c(\underline{s}) = -\frac{cP_0}{(P_1 - P_0)} < 0 \quad (25)$$

$$u_c(\overline{s}) = \frac{c(1-P_0)}{P_1 - P_0} > 0 \quad (26)$$

where, (25), which is the utility of the contributor input at two different innovation costs, reveals that $u_c(\underline{s}) < 0 < u_c(\overline{s})$. The outcome derived from (25) indicates the implementation of negative incentives at low levels of OSS output; negative incentives can result in reputational losses. This demonstrates that different companies set up different incentive systems for OSS platforms.

5 Analysing and comparing the payoff matrices

In this section, we contrast the cases under the IR and LR participation constraints (Case 1 and Case 2, respectively) with the case under IC (Case 3) and compare the payoff matrices from the results obtained in Section 4. In doing so, we provide a further discussion on the issue of moral hazard from both sides based on the model equilibria.

5.1 Case 1

Based on the payoff matrices in Table 2 below, enforcing the participation constraint implies that the contributors must incur the real costs of innovation, regardless of whether the owner is truly exercising supervision, or the contributors have a high degree of self-discipline. Under IR, the contributors produce the OSS output with a higher probability P_1. The owner recognises both the output of the contributors and the innovation cost incurred, even if there is no output with a probability $1–P_1$.

Table 2: **Payoff matrix for moral hazard under IR participation constraint**

		Contributor		
		$\mathcal{HI}$: $c = \overline{c}$	$\mathcal{LI}$: $c = \underline{c}$	$\mathcal{FR}$: $c = 0$
Owner	$\mathcal{VC}$ ($\equiv \mathcal{FC}$):$s = \underline{s} = \overline{s}$	$u_c(s) - \overline{c}$ $P_1 [\Pi(\overline{q}) - s] + (1 - P_1) [\Pi(\underline{q}) - s]$	$u_c(s) - \underline{c}$ $P_0 [\Pi(\overline{q}) - s] + (1 - P_0) [\Pi(\underline{q}) - s]$	$u_c(s) + u_e$ 0
	$\mathcal{FC}$: $s = \underline{s} = \overline{s}$	$u_c(s) - \overline{c}$ $P_1 [\Pi(\overline{q}) - s] + (1 - P_1) [\Pi(\underline{q}) - s]$	$u_c(s) - \underline{c}$ $P_0 [\Pi(\overline{q}) - s] + (1 - P_0) [\Pi(\underline{q}) - s]$	$u_c(s) + u_e$ 0
	$\mathcal{ZC}$: $s \to 0$	$-\overline{c}$ $P_1 \Pi(\overline{q}) + (1 - P_1) \Pi(\underline{q})$	$-\underline{c}$ $P_0 [\Pi(\overline{q}) - t] + (1 - P_0) [\Pi(\underline{q}) - t]$	u_e 0

However, the level of compensation is fixed at s. This leads to a lack of differentiated incentives for the owner with regard to open-source innovation agreement. This compensation is significantly greater than zero at the beginning of the game, i.e., $s \gg 0$, while it tends to veer towards zero during the process, i.e., $s \to 0$, thus implying that a moral hazard behaviour from the owner's side by taking the value of the contributors' input with almost zero compensation.

Suppose IR holds with non-binding IC, the agent's compensation for inducing high effort is unnecessarily high. While the contributors receive a fixed compensation, the efficiency cost of risk-bearing from the agent's side is reduced and there is a singular motivation to increase their own earnings in order to hide true effort, and instead invest the innovation output in competing OSS projects, or proprietary software for additional earnings, u_e, i.e., the moral hazard problem arises from the contributors.

5.2 Case 2

The payoff matrix for moral hazard under LR is summarised in Table 3. This case strongly mimics the reality of platform-based OSS development where moral hazard behind free-riding that is prevalent among the contributor community. In this context, the interests include not only the compensation received by the original contributor in relation to the source code, but also intangible values such as the medium- to long-term industry reputation that the source code gives to the original contributor; even if the original contributor receives zero financial reward, reputation is an important moral and spiritual incentive for the original contributor so that they may continue contributing knowledge to the platforms. These phenomena have generated a free-riding attitude among OSS developers.

Meanwhile, open-source communities within the platform architectures suffer not only from the simple first-order free-riding problem described above, but also from higher-order free-rider behaviour. Higher-order free-riding is a behavioural phenomenon by which simple first-order free-riders choose their targets with a degree of randomness and invisibility. Most contributors

Table 3: **Payoff matrix for moral hazard under LR participation constraint**

		Contributor		
		$\mathcal{HI}$: $c = \overline{c}$	$\mathcal{LI}$: $c = \underline{c}$	$\mathcal{FR}$: $c = 0$
Owner	$\mathcal{VC}$:$s = \{\underline{s}, \overline{s} \mid 0 \leq \underline{s} < \overline{s}\}$	$P_1 u_c(\overline{s}) + (1 - P_1) u_c(\underline{s}) - \overline{c}$ $P_1 [\Pi(\overline{q}) - \overline{s}] + (1 - P_1) [\Pi(\underline{q}) - \underline{s}]$	$P_0 u_c(\overline{s}) + (1 - P_0) u_c(\underline{s}) - \underline{c}$ $P_0 [\Pi(\overline{q}) - \overline{s}] + (1 - P_0) [\Pi(\underline{q}) - \underline{s}]$	u_e 0
	$\mathcal{FC}$: $s = \underline{s} = \overline{s}$	$u_c(s) - \overline{c}$ $P_1 [\Pi(\overline{q}) - s] + (1 - P_1) [\Pi(\underline{q}) - s]$	$u_c(s) - \underline{c}$ $P_0 [\Pi(\overline{q}) - s] + (1 - P_0) [\Pi(\underline{q}) - s]$	u_e 0
	$\mathcal{ZC}$: $s \to 0$	$-\overline{c}$ $P_1 \Pi(\overline{q}) + (1 - P_1) \Pi(\underline{q})$	$-\underline{c}$ $P_0 [\Pi(\overline{q}) - t] + (1 - P_0) [\Pi(\underline{q}) - t]$	u_e 0

choose to ignore such issues, assuming that, sooner or later, a force or mechanism, can emerge to govern the project.

Due to the invisibility of free-riding, it is difficult for the project owner to monitor and verify the true cost of the open-source innovation. Furthermore, monitoring incurs additional costs. The owner can choose to pay a fixed level of compensation, or move towards zero compensation in an attempt to simplify the evaluation of open-source contributions, thus achieving cost savings. Fixed compensation saves the costs of investigation and supervision from the principal's side, but it leads to a lack of incentive gradients and increased imbalance of incentives for innovation.

5.3 Case 3

Ensuring incentive-compatibility is subject to principal-side moral hazard whereby the owner imposes negative incentives on low-input/output contributors (Table 4). Negative incentives mechanically follow a gradient of incentives, ignoring the loss of both reputation and long-term financial gains that result from negative incentives.

Table 4: **Payoff matrix for moral hazard under IC constraint**

		Contributor		
		$\mathcal{HI}$: $c = \overline{c}$	$\mathcal{LI}$: $c = \underline{c}$	$\mathcal{FR}$: $c = 0$
Owner	$\mathcal{VC}$: $s = \{\underline{s}, \overline{s} \mid 0 \leq \underline{s} < \overline{s}\}$	$P_1 u_c(\overline{s}) + (1 - P_1) u_c(\underline{s}) - \overline{c}$ $P_1 [\Pi(\overline{q}) - \overline{s}] + (1 - P_1) [\Pi(\underline{q}) - \underline{s}]$	$P_0 u_c(\overline{s}) + (1 - P_0) u_c(\underline{s}) - \underline{c}$ $P_0 [\Pi(\overline{q}) - \overline{s}] + (1 - P_0) [\Pi(\underline{q}) - \underline{s}]$	$u_c(s) + u_e$ 0
	$\mathcal{FC}$: $s = \underline{s} = \overline{s}$	$u_c(s) - \overline{c}$ $P_1 [\Pi(\overline{q}) - s] + (1 - P_1) [\Pi(\underline{q}) - s]$	$u_c(s) - \underline{c}$ $P_0 [\Pi(\overline{q}) - s] + (1 - P_0) [\Pi(\underline{q}) - s]$	$u_c(s) + u_e$ 0
	$\mathcal{ZC}$: $s \to 0$	$-\overline{c}$ $P_1 \Pi(\overline{q}) + (1 - P_1) \Pi(\underline{q})$	$-\underline{c}$ $P_0 [\Pi(\overline{q}) - t] + (1 - P_0) [\Pi(\underline{q}) - t]$	u_e 0
	Negative incentive:$s < 0$	$-\overline{c}$ $P_1 \Pi(\overline{q}) + (1 - P_1) \Pi(\underline{q})$	$-\underline{c}$ $P_0 [\Pi(\overline{q}) - t] + (1 - P_0) [\Pi(\underline{q}) - t]$	0 0

Another possibility where principal-side moral hazard arises lies in determining payment. Paying only the necessary labour costs and the contributors only recover the economic value that corresponds to the necessary labour time invested in the open-source innovation output; they do not receive additional economic incentives.

Due to having incomplete information, the owner is unable to observe and verify the true costs of the contributors individually. Based on the estimation that the probability distribution of input costs for all contributors will obey a normal distribution (approximately), the owner's expectations of compensation payments for all contributors tend to be at the level of the mean value of input costs for the contributors, i.e., $P_1 u_c(\overline{s}) + (1 - P_1) u_c(\underline{s}) \to c$. Here, the owner saves on the information costs incurred in carrying out monitoring. The owner treats the input costs of contributors as risk neutral, and the deviation of the input costs of any single contributor from the mean of the collective costs is stochastic.

The game between the contributors and the owner is sequential. When the rewards received by the contributors do not cover their costs, the negative financial profit naturally motivates the contributors to withdraw from the OSS project. Therefore, agent-side moral hazard arises from the perspective of the agent when they turn to other competing OSS projects, or proprietary software, for additional revenues.

These moral hazard behaviours manifest in various specific forms and are challenging to detect even when the two parties interact more than once. For example, unobservable actions can arise through retention of novel technical solutions in core modules of the software and strategic concealment of source codes to maintain one's competitive advantage. The contributors often prioritise the cost-benefit analysis of on-going tasks based on short-term personal interests, and limited time and economic resources. Furthermore, moral hazard can be particularly profound in human-machine collaborations given advances in generative Artificial Intelligence (AI), automation, and data analytics, in conjunction with the inadequate governance system. To illustrate, a contributor could exploit AI agents to remotely and anonymously inject vulnerabilities into source codes, then submit solutions under their own name, subjecting the relationship to moral hazard from the contributor's side.

Clearly, the distributed, asynchronous nature of open-source model renders moral hazards among the contributors more discreet. Such interactions provide the contributors with a late-mover advantage. The contributor's moral hazard sends a negative signal to the other party. Indeed, the cumulative impact of contributors' delayed actions will invariably result in a weakening of the owner's control over the open-source project, ultimately leading to the collapse of the original incentive mechanism. The contributors have the opportunity to enhance personal activity levels and accumulate community reputation by reneging on input provision agreement. This approach facilitates the acquisition of higher-value job opportunities without assuming long-term maintenance responsibilities. This form of reputation-based intertemporal and spatial arbitrage has the potential to undermine the incentive compatibility contract.

5.4 Welfare Evaluation

This section further summarises the welfare effects, W, derived from the income and costs of owners and contributors across the aforementioned three cases. We define a simple welfare criterion in the static model

$$W = [\Pi(q) - s] + [s - c(e)] - K \tag{27}$$

where $\Pi(q) - s$ represents the owner's net benefit, while the second item $s - c(e)$, denotes the contributor's net benefit (instantaneous utility). K signifies frictions within the open-source platform, primarily comprising the enforcement costs of incomplete contracts.

Clearly, under IR in Case 1, compensation is fixed at $s = \underline{s} = \overline{s}$, and the welfare effect simplifies to

$$W_1 = \Pi(q) - c(e) - K \tag{28}$$

where W_1 is unrelated to the compensation level, positively correlated with $\Pi(q)$ derived from the quality of open-source contributions, and negatively correlated with the contributor's input costs. Due to the participation constraint, the derived fixed compensation can be independent of the input cost level. At this point, the enforcement cost K of the incomplete contract is approximately zero.

Under LR in Case 2, the level of cost $c(e)$ incurred by the contributor simultaneously influences both the returns $\Pi(q)$ from open-source contributions and the compensation s. In this scenario of flexible participation, a mismatch arises between input and compensation, as well as between input and returns. (27) holds under such conditions of mismatch

$$W_2 = [\Pi(\overline{q}) - \overline{s}] + [\overline{s} - \underline{c}] - K \quad \text{or} \quad W_2 = [\Pi(\overline{q} - \underline{s}] + [\underline{s} - \underline{c}] - K \tag{29}$$

Evidently, attempting to eliminate the mismatches would incur additional expenditure on the enforcement costs K of incomplete contracts.

Therefore, in Case 3, where returns, compensation, and inputs are strictly matched according to the IC constraint rules, internal frictions within the open-source platform are reduced. The enforcement costs of incomplete contracts can be minimised to zero, and the welfare effect simplifies to

$$W_3 = \Pi(\overline{q}) - \overline{s} - \overline{c} \quad \text{or} \quad W_3 = \Pi(\underline{q}) - \underline{s} - \underline{c} \tag{30}$$

The welfare effects observed across the three cases demonstrate that, in the absence of incentive compatibility, owners and contributors on open-source platforms face a dilemma between open-sourcing and non-open-sourcing, and between contributing and non-contributing. Owners worry that the knowledge spillovers from open-source projects provide incentives to free ride, while contributors care that their input costs may not receive fair compensation. When compensation strictly matches input, the contract enforcement cost K approaches zero. At this point, the platform welfare W_3 requires no additional mechanism to sustain itself. High compensation naturally attracts high-cost, high-quality open-source contributions, yielding a welfare outcome exceeding both W_1 and W_2. Furthermore, due to the existence of contract enforcement cost K and the unsustainability of low-cost, low-input open-source contributions yielding high-quality, high-return outcomes, $W_2 > W_1$ only happens by chance.

6 Implications for platform-based OSS development

This simple model delivers some intuitive results. Our model can provide important and practical implications for OSS platforms where moral hazard has been argued to be relevant in determining contract terms or collaborative agreements. Neither the owner nor the contributors can exactly determine output *ex ante*, hence participation interaction is not fully contractible, even if well-functioning institutions are readily accessible. But, for OSS, the relative rudimentary nature of institutions makes it even more difficult to devise incentive and licensing schemes. In the light of the above results, the following governance approaches can be considered in order to monitor and manage the issue and thus strengthening the incentives for both parties to exert effort for proving inputs.

6.1 Enforcement mechanisms

Our findings also suggest that repeated interaction, reputational mechanism, and (repeated) communication in relational contracting devices (e.g., statements of intent on relationship building or a *quid pro quo*) would help provide contract enforcement or the solution mechanisms to coordination failures or on limiting side-selling (to proprietary platforms). This means that both OSS parties are likely to engage in long-term repeated relationships with interim negotiation. Notable studies such as Baker *et al.* (2002) and Gibbons (2005) delve into the effectiveness of these (less formal) enforcement mechanisms on preventing coordination failures or contract non-performance.

In an open-source development-based community, moral hazard is profound in relational contracts, especially when participants are risk-averse, but could be mitigated through a contract that creatively designs incentive strategies, utilises non-financial rewards or observable performance signals (Ho, 2024), or seeks support from external resources. Contributors need to be flexible with regard to innovation investment and actively collaborate with the open-source community network of contributors to reduce the cost of innovation (e.g., through a general public repository). This implies that an agent could be induced to comply with a non-binding agreement if the intended goals and vision of the project are clearly communicated and can inspire the participants' sense of teamwork (joint risk sharing). Contributors are more motivated to contribute to the long-term development of the project.

In addition, our finding reveals the second-best optimum while the *ex post* participation constraints must be satisfied. This implies that another possibility of achieving efficiency with the second-best contract (effort) exists by easing some of the basic assumptions and constraints. Suppose both the OSS principal and agent consider *ex ante* contracting which replaces the *ex post* participation constraints (both the IR- and LR-type), while IC must still be satisfied. Assuming that both parties aim to maximise their expected payoffs, these more efficient contracts allowing the principal to appropriate *ex ante* the informational rents left to the agent *ex post* can reduce or even eliminate the cost of these rents that may arise based on pre-contractual private information (Heumann, 2020; Ho, 2024). The mechanism of smart contracts is also discussed in Onjewu *et al.* (2023).

6.2 Role of institutions and public enforcement

Bergstrom *et al.* (1992) show that, in the provision of public goods, a free-rider problem emerges if private incentives faced by economic agents are not aligned with their shared goals. Indeed, societal cooperation, social capital and collective actions in minimising the costs of uncertainty, knowledge exchange and value transactions, and ultimately enhancing economic performance and social welfare, requires a set of rules or less formal arrangements codified in social or enforcement norms. However, mutually beneficial cooperation that is privately costly is not practically observed.

It is therefore advantageous to establish clear contributor criteria and evaluation mechanisms that determine the level of effort by the contributors for collective action projects in open ecosystems with joint risk sharing and ubiquitous information. Through such mechanisms, contributors in voluntary cooperation can be treated equally, and limited resources can be allocated to those who have truly contributed to the OSS project, in a peer-to-peer democratised environment. For example, incentives in the form of public subsidies may be required to induce adequate contributions and reveal research findings (von Hippel and von Krogh, 2006). Onjewu *et al.* (2023) discuss how encrypted transactions facilitate the congruence of goals in blockchain alliances.

Furthermore, the role of institutions in shaping economic behaviour and developing social capital lies in strengthening social connections, networks, norms, and trust, in order to facilitate societal cooperation and collective action. Informal institutions such as self-enforcing rules, conventions and social norms can offer enforcement and be an important mechanism to overcome free-rider problems, particularly in OSS communities, where long-term repeated interactions between parties are highly valued. The norms of reciprocity and maintaining long-term relationships lead to pro-social action.

7 Conclusions

This paper proposed a principal-agent model to examine the problem prevalent in this environment. We have examined several conditions of the model under which the incomplete and contractual nature of open-source projects gives both parties the opportunity to escape responsibility. Double moral hazard arises as the result of OSS participants acting in their self-interest and shirking on their contractual commitments to maximise their private gains.

Our paper contributes to the theory from three main perspectives. Firstly, the findings address moral hazard that is beyond the pricing structure within the economic theories of platforms. Secondly, our paper focuses on two-sided moral hazard within the OSS platform paradigm, which advances the existing understanding of knowledge transfer, or contract-based knowledge collaboration, that are based on formal IPR systems. Thirdly, we construct a simple model and identify the mechanisms through which double moral hazard is generated by different incentives and primary activities. This method expands a principal-agent model with double moral hazard, and considers the behaviour from both the OSS owner and contributors. The method used in this paper can be generalised and geared towards applications and experiments designed to test the hypothesis within platform-based OSS development.

We recognise that there are possible avenues for future research. In terms of the modelling framework, the single-shot sequential game could be enriched by considering a mechanism that addresses forms of uncertainties such as contracting participants' opportunistic behaviour tailored specifically to our setting of OSS platforms. Additionally, designing self-enforcement mechanisms through dynamic (repeated) interactions that penalise non-performance and address the deficiencies of incomplete contracts represents an important research direction. A methodology that certainly deserves further attention is experimental economics. To test the hypothesis derived from the theoretical model, future work could also examine strategies for enforcement by a laboratory experiment for different contractual treatments and investigate behavioural heterogeneity by an experiment for testing the behavioural factors relating to the characteristics and actions of the OSS participants.

References

Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics*, **37**(3), 668–691.

Arnott, R. and Stiglitz, J. (1988). The basic analytics of moral hazard. *Scandinavian Journal of Economics*, **90**(3), 383–413.

Baker, G., Gibbons, R., and Murphy, K. J. (2002). Relational contracts and the theory of the firm. *The Quarterly Journal of Economics*, **117**(1), 39–84.

Bergmann, R. and Friedl, G. (2008). Controlling innovative projects with moral hazard and asymmetric information. *Research Policy*, **37**(9), 1504–1514.

Bergstrom, T., Blume, L., and Varian, H. (1992). Uniqueness of nash equilibrium in private provision of public goods: An improved proof. *Journal of Public Economics*, **49**(3), 391–392.

Bhattacharyya, S. and Lafontaine, F. (1995). Double-sided moral hazard and the nature of share contracts. *RAND Journal of Economics*, **26**(4), 761–781.

Boudreau, K. J. (2010). Open platform strategies and innovation: granting access vs. devolving control. *Management Science*, **56**(10), 1849–1872.

Cai, Y.-J. and Choi, T.-M. (2020). Platform supported supply chain operations in the blockchain era: Supply contracting and moral hazards. *Decision Sciences*, **52**(4), 866–892.

Caillaud, B. and Jullien, B. (2003). Chicken-and-egg: Competition among intermediation service providers. *RAND Journal of Economics*, **34**, 309–328.

Carrasco-Farre, C., Snihur, Y., Berrone, P., and Ricart, J. E. (2022). The stakeholder value proposition of digital platforms in an urban ecosystem. *Research Policy*, **51**(4), 104488.

Carroll, G. and Bolte, L. (2023). Robust contracting under double moral hazard. *Theoretical Economics*, **18**, 1623–1663.

Chen, G. and Li, B. (2021). A double moral hazard model for inter-organizational knowledge sharing. *Expert Systems*, **8**(8), 12750.

Cooper, R. and Ross, T. (1985). Product warranties and double moral hazard. *RAND Journal of Economics*, **16**(1), 103–113.

Dai, T. and Toikka, J. (2022). Robust incentives for teams. *Econometrica*, **90**(4), 1583–1613.

Dechenaux, E., Thursby, J., and Thursby, M. (2011). Inventor moral hazard in university licensing: The role of contracts. *Research Policy*, **40**(1), 94–104.

Eswaran, M. and Kotwal, A. (1985). A theory of contractual structure in agriculture. *American Economic Review*, **75**(3), 352–367.

Evans, D. S. (2003). Some empirical aspects of multi-sided platform industries. *Review of Network Economics*, **2**, 191–209.

Franke, N. and Shah, S. (2003). How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research Policy*, **32**(1), 157–178.

Franke, N. and von Hippel, E. (2003). Satisfying heterogeneous user needs via innovation toolkits: the case of apache security software. *Research Policy*, **32**(7), 1199–1215.

Gawer, A. (2010). *Platforms, Markets and Innovation.* Northampton, MA, Edward Elgar.

Gawer, A. and Harracá, M. (2025). Inconsistent platform governance and social contagion of misconduct in digital ecosystems: A complementors perspective. *Research Policy*, **54**(8), 105300.

Gibbons, R. (2005). Four formal(izable) theories of the firm? *Journal of Economic Behavior & Organization*, **58**(2), 200–245.

Granstrand, O. (2024). A cooperative game theory systems approach to the value analysis of (innovation) ecosystems. *Technovation*, **130**, 102926.

Hagiu, A. (2005). Pricing and commitment by two-sided platforms. *RAND Journal of Economics*, **37**, 720–737.

Hayry, M. (2021). Just better utilitarianism. *Cambridge Quarterly of Healthcare Ethics*, **30**(2), 343–367.

Henkel, J. (2006). Selective revealing in open innovation processes: The case of embedded linux. *Research Policy*, **35**(7), 953–969.

Heumann, T. (2020). Information design and sequential screening with ex post participation constraint. *Theoretical Economics*, **15**(1), 319–359.

Ho, H. (2024). Loss aversion, moral hazard, and stochastic contracts. *Managerial and Decision Economics.*

Iansiti, M. and Lakhani, K. R. (2020). From disruption to collision: The new competitive dynamics. *MIT Sloan Management Review*, **61**(3), 34–39.

Islam, M., Miller, J., and Park, H. D. (2017). But what will it cost me? how do private costs of participation affect open source software projects? *Research Policy*, **46**(6), 1062–1070.

James, S., Liu, Z., White, G. R. T., and Samuel, A. (2023). Introducing ethical theory to the triple helix model: Supererogatory acts in crisis innovation. *Technovation*, **126**, 102832.

Krishnamurthy, S., Ou, S., and Tripathi, A. K. (2014). Acceptance of monetary rewards in open source software development. *Research Policy*, **43**(4), 632–644.

Leckel, A., Randhawa, K., and Piller, F. T. (2025). Interaction design for open innovation platforms: A social exchange perspective. *Journal of Product Innovation Management*, **42**(4), 641–678.

Liu, M., Brynjolfsson, E., and Dowlatabadi, J. (2021). Do digital platforms reduce moral hazard? the case of uber and taxis. *Management Science*, **67**(8), 4665–4685.

MacCormack, A., Rusnak, J., and Baldwin, C. Y. (2006). Exploring the structure of complex software designs: An empirical study of open source and proprietary code. *Management Science*, **52**(7), 1015–1030.

Madanaguli, A., Dhir, A., Talwar, S., Clauss, T., Kraus, S., and Kaur, P. (2023). Diving into the uncertainties of open innovation: A systematic review of risks to uncover pertinent typologies and unexplored horizons. *Technovation*, **119**, 102582.

Mateos-Garcia, J. and Steinmueller, W. E. (2008). The institutions of open source software: Examining the debian community. *Information Economics and Policy*, **20**(4), 333–344.

Molm, L. D., Takahashi, N., and Peterson, G. (2000). Risk and trust in social exchange: An experimental test of a classical proposition. *American Journal of Sociology*, **105**(5), 1396–1427.

Mooi, E. and Wuyts, S. (2021). Value from technology licensing – the role of monitoring and licensing experience. *International Journal of Research in Marketing*, **38**(4), 1034–1054.

Nguyen, L. T. M. and Nguyen, P. T. (2025). The interplay between governance mechanisms of blockchain platforms. *Economics of Innovation and New Technology*, **34**(3), 367–391.

Niedermayer, A. (2013). On platforms, incomplete contracts, and open source software. *International Journal of Industrial Organization*, **31**(6), 714–722.

Onjewu, A.-K. E., Walton, N., and Koliousis, I. (2023). Blockchain agency theory. *Technological Forecasting and Social Change*, **191**, 122482.

Petralia, S. (2025). Open source software as digital platforms to innovate. *Journal of Economic Behavior & Organization*, **237**, 107109.

Poblete, J. and Spulber, D. (2017). Managing innovation: Optimal incentive contracts for delegated r&d with double moral hazard. *European Economic Review*, **95**, 38–61.

Rasmusen, E. (2006). *Games and Information: An Introduction to Game Theory (4th ed.)*. Blackwell, Oxford.

Rey, P. and Stiglitz, J. E. (2024). Moral hazard and unemployment in competitive equilibrium. NBER Working Papers 32700, National Bureau of Economic Research, Inc.

Rochet, J. C. and Tirole, J. (2006). Two-sided markets: A progress report. *RAND Journal of Economics*, **37**, 645–667.

Roels, G., Karmarkar, U., and Carr, S. (2010). Contracting for collaborative services. *Management Science*, **56**(5), 849–863.

Roger, G. and Vasconcelos, L. (2014). Platform pricing structure and moral hazard. *Journal of Economics and Management Strategy*, **23**(3), 527–547.

Spagnoletti, P., Resca, A., and Lee, G. (2015). A design theory for digital platforms supporting online communities: A multiple case study. *Journal of Information Technology*, **30**(4), 364–380.

Springer, V., Randhawa, K., Jovanović, M., Ritala, P., and Piller, F. T. (2025). Platform design and governance in industrial markets: Charting the meta-organizational logic. *Research Policy*, **54**(6), 105236.

Spulber, D. F. (2018). The economics of markets and platforms. *Journal of Economics and Management Strategy*, **28**, 159–172.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, **12**(3), 183–206.

Uzunca, B. and Kas, J. (2023). Automated governance mechanisms in digital labour platforms: how uber nudges and sludges its drivers. *Industry and Innovation*, **30**(6), 664–693.

von Hippel, E. and von Krogh, G. (2006). Free revealing and the private-collective model for innovation incentives. *R & D Management*, **36**, 295–306.

Wen, X., Chen, X., and Yang, Z. (2023). Subsidization of public transit service under double moral hazard. *Physica A: Statistical Mechanics and its Applications*, **632**, 129304.