



Research article

Stability of scientific big data sharing mechanism based on two-way principal-agent

Wang Zhang^{1,*} and Jingkun Zhang²

¹ School of Economics and Management, Northwest University, Xi'an 710127, China

² School of Accounting, XiJing University, Xi'an 710123, China

* **Correspondence:** Email: zhangwang@nwu.edu.cn; Tel: +8618702958556.

Abstract: In the era of big data, facing the data-intensive scientific paradigm shift and the explosion of scientific big data, there is an urgent need for alliance cooperation between heterogeneous research groups to actively open and share scientific big data to support China's economic development, technological innovation and national security. Therefore, the study of scientific big data sharing mechanism has very important practical significance. We think science big data sharing is an ecosystem that is constantly evolving to higher-order ecological evolution. Based on the dual perspectives of psychological contract and contractual contract, the scientific big data sharing strategy evolution mechanism and sharing strategy incentive mechanism are explored. The research finds that the cooperation of scientific research groups is bound by psychological contract and contractual contract; stochastic evolutionary game has stronger explanatory power for sharing strategy evolution, complementarity is positive indicator, random interference and moral risk are negative indicators; two-way principal agent can describe Alliance members are mutually entrusted, and the shared strategy incentive contract consists of fixed wages and incentive wages, which are proportional to risk.

Keywords: scientific big data; sharing strategy evolution; sharing strategy incentives; stochastic evolutionary game; two-way principal agent

Mathematics Subject Classification: 20C25, 20C35

1. Introduction

In recent years, with the continuous improvement of China's scientific and technological

innovation capacity and innovation investment, the scientific big data collection and management capacity has also been improved step by step, but the scientific big data game strategy selection and open sharing incentive mechanism are relatively weak. In the era of big data, scientific and technological innovation is increasingly dependent on the analysis, mining and comprehensive utilization of scientific big data. At the same time, scientific big data is considered as a new engine of scientific discovery and a new scientific research methodology. The sharing of scientific big data is beneficial to promoting scientific progress, reducing repetitive work, improving scientific research ability and improving scientific research efficiency. It is of great significance. Scientific big data [1] is a compound term that refers to big data related to scientific research, including big data of scientific knowledge and big data of scientific activities (hereinafter referred to as "scientific data"). Scientific data is mainly derived from the physical world and has many characteristics, such as complexity, multi-source, heterogeneity, etc. It is urging the transformation of scientific research paradigm to data-intensive scientific paradigm [2]. Scientific research groups are not only the producers of scientific data, but also the beneficiaries of scientific data. Multi-source heterogeneous data fusion of heterogeneous scientific research groups can effectively release the economic value and scientific research value of scientific data, break down information barriers and information islands among scientific research groups, and achieve the amplification effect of $1+1>2$. Although there are many benefits in the open sharing of scientific data, the scientific data sharing mechanism is still a complex and difficult issue, involving a series of issues such as sharing strategy, incentive mechanism, intellectual property rights, regulations and policies [3].

Since the launch of the scientific data sharing project in 2001, China has implemented strategies, such as the Action Plan for Promoting the Development of Big Data, the 13th Five-Year National Science and Technology Innovation Plan and issued the Scientific Data Management Measures, aiming to improve the scientific data sharing mechanism and improve the scientific data sharing and sharing, so that the open sharing of scientific data is the normal, and the open sharing of scientific data is the exception. The existing research shows that [4]: The deep fusion of multi-source and heterogeneous scientific data is not only to improve the academic and economic value of scientific data, but also to improve the innovation ability and efficiency of scientific research groups. Scientific data sharing includes four parts: basic elements [5], sharing mode [6], sharing mechanism [7] and security mechanism [8]. Under a suitable sharing mechanism, the basic elements realize the circulation of data resources according to the existing sharing mode, fully release the value co-creation function of data, and the security mechanism focuses on policy, technology and financial support. The influencing factors of scientific data sharing strategy mainly include personal factors [9], industry norms [10], platforms and governance [11] and many factors lead to difficulties in obtaining scientific data and low willingness of scientific research groups to share.

With the expansion and in-depth development of science and technology, it has provided support for alliance and cooperation between scientific research groups. Throughout history, few scientific research groups have completed major breakthroughs in the frontier of science, and most of them have been completed under the alliance and cooperation between heterogeneous scientific research groups. In addition, the quality of talents, research costs, scientific research facilities and the marketization of scientific data require exchanges and multilateral cooperation in different regions and fields to form a dynamic cooperation mechanism for sharing data resources. Data sharing by researchers with different disciplines, academic backgrounds and academic ideas can produce data resource fusion and amplification effect. Therefore, the issue of scientific data sharing mechanism for heterogeneous

scientific research groups is crucial. For example, the research and development of the large aircraft C919 has been completed by up to 14 universities in collaboration. A suitable sharing mechanism is the prerequisite for the smooth development of the research and development work.

The core problem of scientific data sharing is the separation of data ownership and use right [12]. Due to the existence of information asymmetry and incomplete rationality, it is difficult to choose the sharing strategy between data owners and data users. Scientific research groups have dual identities as owners and users of data. Heterogeneous scientific research groups have both competition and cooperation game relationship and principal-agent relationship. The game relationship of competition and cooperation is that scientific research groups work together to form strategic alliances for data sharing, and at the same time form competitive concerns when facing the distribution of benefits and risks. The principal-agent relationship refers to that scientific research groups entrust each other with the right to use their own data temporarily. At this time, the heterogeneous groups are in the principal-agent relationship with each other. By building a scientific big data sharing ecosystem, this paper proposes two major constraint mechanisms for the evolution of the sharing ecosystem to a higher level, psychological indenture and contract indenture, which analyzes the path of iterative updating of the sharing ecosystem, clarifies the evolution mechanism of the sharing strategy and the incentive mechanism of the sharing strategy, discusses the influencing factors in an uncertain environment and designs the optimal incentive contract. It is expected to enrich the scientific big data sharing theory and provide policy inspiration for China's data sharing cause.

The second part is the basic theory of the scientific big data sharing ecosystem, the third part is the evolution mechanism of sharing strategies from the perspective of psychological contracts, the fourth part is the incentive mechanism of sharing strategies from the perspective of contract contracts, and the fifth part is the research conclusion and inspiration.

2. Theoretical basis of scientific big data sharing ecosystem

Scientific research group [13] is a relatively loose organization, which generally refers to a group that has common scientific research objectives and is combined in a unique way. There are great differences in the external environment and internal structure of scientific research groups, which leads to heterogeneity among scientific research groups (hereinafter referred to as "heterogeneous scientific research groups"). With the increasing number of scientific research data, open data sharing has become the trend of scientific innovation and even national innovation. In the face of massive multi-source heterogeneous data, scientific research groups voluntarily form strategic alliances for collaborative innovation in order to improve scientific research capacity, improve innovation efficiency, meet diverse innovation and other needs, so as to realize high-speed circulation and sharing of scientific data among different levels, different fields and different standards, effectively reduce data collection costs and improve data utilization efficiency. Scientific research groups and the sharing environment constitute the scientific big data sharing ecosystem (Figure 1). Data is the flowing energy source in the ecosystem. Heterogeneous scientific research group alliance contracts are mainly subject to psychological contracts and contract contracts. Psychological contract is an implicit contract between heterogeneous scientific research groups, which is uncertain and dynamic; The contract is a linear contract between heterogeneous scientific research groups, with legitimacy and equality. Both are referred to as "explicit contract". According to the characteristics of psychological indenture and indenture contract, the stochastic evolutionary game model is used to explore the dynamic equilibrium

and influencing factors of sharing strategy, and the two-way principal-agent model is used to explore the incentive mechanism and optimal contract design of sharing strategy. Stochastic evolutionary game theory originates from the fusion of evolutionary game and stochastic differential equation [14]. It explains how the sharing strategy of scientific research groups achieves equilibrium by simulating the interference caused by uncertain sharing environment. Uncertain environment mainly refers to the social system where scientific research groups are located is full of uncertainty and is affected by external environment (social culture, social interests, macro policies) and internal environment (group emotion, group organization, individual differences) and other factors; Two-way principal-agent [15] is based on asymmetric information game theory, which is used to explain the principal-agent relationship between the participants in the alliance cooperation. The participants share the output value of the alliance and the random risk of the alliance with the goal of maximizing the individual effect function.

The scientific big data sharing ecosystem includes [16]: six links: demand analysis, data collection, data processing, heterogeneous data amplification effect, data configuration and data empowerment. Through data sharing flow, the deep integration and complementarity of multi-source heterogeneous data can be realized, and the scientific research ability and innovation efficiency of scientific research groups can be improved. Demand analysis is the analysis and positioning of innovation demand of scientific research groups. With the implementation of national strategies such as Internet plus and big data, the trend of cross collaboration of heterogeneous scientific research groups and deep integration of multi-source heterogeneous data is becoming increasingly obvious, and the data demand is increasingly diversified; data collection means to collect matching data after accurately grasping the needs of scientific research groups. Data can also be obtained through market-oriented transactions as innovative information; data processing is used to extract high-value digital information from a large number of original data, mainly for data collection, storage, retrieval, processing, transformation and transmission; the amplification effect of heterogeneous data refers to the total effect of data fusion greater than the sum of the two, which mainly reflects the complementarity of multi-source heterogeneous data; data configuration is to reconfigure innovation elements after heterogeneous data fusion to realize new innovation value and optimization of innovation elements; data empowerment refers to the integration of different links in the innovation activities of scientific research groups, mainly including research, development and application.

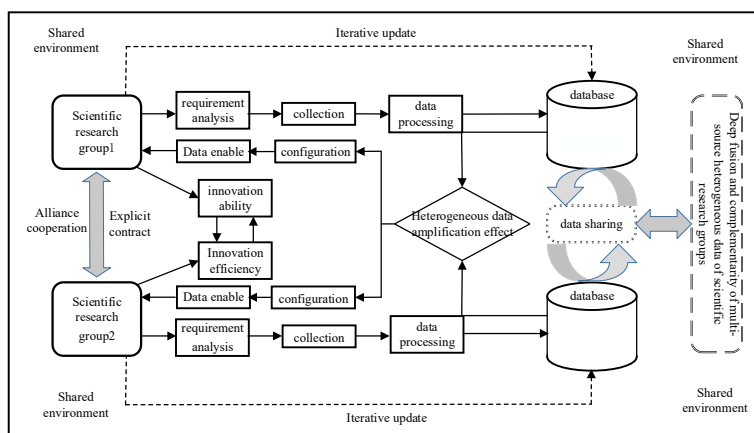


Figure 1. Scientific big data sharing ecosystem.

3. Evolutionary mechanism of sharing strategy from the perspective of psychological contract

3.1. Subheading stochastic evolutionary game analysis of sharing strategy

Heterogeneous scientific research groups can choose to share or not share two strategies under the constraint of psychological contract mechanism in order to achieve deep integration of multi-source heterogeneous data and form alliance cooperation. The choice of strategy is related to the existence of alliance relationship and the distribution of benefits. The internal influencing factors of strategy choice mainly include: Degree of complementarity, willingness to share, random interference and moral hazard. The degree of complementarity is a quantitative indicator of complementarity, and complementarity is a prerequisite for the sharing of scientific data. In an era of rich and diverse data types, the sharing of complementary data or the cross-fusion of multi-source heterogeneous data is the basis for innovation. Therefore, the identification of the degree of complementarity is an important part of the needs analysis of scientific research groups; the willingness to share is an externalized manifestation of the psychological contract and a subjective indicator to measure the degree of sharing of scientific research groups. The low willingness to share is the key problem leading to the difficulty of scientific data sharing in China. The willingness to share is often inseparable from the trust mechanism and incentive contract. The greater the trust between scientific research groups, the higher the willingness to share, the stronger the incentive degree of incentive contract and the higher the willingness to share; Random interference is a local interference caused by the uncertain environment on the sharing strategy of scientific research groups. It is a non-dominant system that dynamically adjusts the energy, only changes the local state and has no obvious propagation law to follow; Moral hazard is caused by information asymmetry, network externalities, "free ride" and other factors of both sides of the game. One side of the game adopts the sharing strategy to make the other party generate moral hazard and increase its own losses. On the basis of previous studies, this section discusses the dynamic equilibrium and influencing factors of sharing strategy evolution in uncertain environment from the perspective of psychological contract. See Table 1 for model symbols and meanings.

Table 1. Symbols and meanings.

Symbol	Meaning	Symbol	Meaning
x	Data scale	b	Cost factor
λ	Complementarity	$\omega(t)$	One-dimensional standard <i>Brown</i> motion, subject to normal distribution $N(0, t)$
δ	Willingness to share	C	Shared cost function
Δx	Moral risk	ρ	Absolute risk aversion measurement
$p(t)$	Share strategy proportion of scientific research group 1	β	Risk
$q(t)$	Sharing strategy proportion of scientific research group 2	α	Output elasticity coefficient
θ	Random interference intensity	w	Fixed salary
h	Simulation step size	\bar{w}	Retained wages

Suppose that the participants are heterogeneous scientific research groups: scientific research group 1 and scientific research group 2. The strategic space of both sides of the game is the same (shared, not shared). Build the benefit matrix of scientific research group game, see Table 2.

Table 2. Benefit matrix of scientific research group game

Scientific research Group 1	Scientific research group 2	
	Share	Don't share
Share	$x_1 + \lambda_1 x_2, x_2 + \lambda_2 x_1$	$x_1 + \lambda_1 \delta_2 x_2 - \Delta x_1, x_2 + \lambda_2 x_1$
Don't share	$x_1 + \lambda_1 x_2, x_2 + \lambda_2 \delta_1 x_1 - \Delta x_2$	x_1, x_2

The benefit matrix is analyzed as follows:

When both sides of the game choose the sharing strategy, the heterogeneous scientific research groups realize complete docking of multi-source and heterogeneous data, and the benefits are $x_1 + \lambda_1 x_2$ and $x_2 + \lambda_2 x_1$; When scientific research group 1 chooses the non-sharing strategy and scientific research group 2 chooses the sharing strategy, the benefit of scientific research group 1 is $x_1 + \lambda_1 x_2$, the benefits of scientific research group 2 are $x_2 + \lambda_2 \delta_1 x_1 - \Delta x_1$. When scientific research group 1 chooses to share and scientific research group 2 chooses not to share, the benefit of scientific research group 1 is $x_1 + \lambda_1 \delta_2 x_2 - \Delta x_1$, the benefits of scientific research group 2 are $x_2 + \lambda_2 x_1$; When both sides of the game choose not to share, the bilateral contractual relationship of the strategic alliance cannot exist, the integration of multi-source heterogeneous data cannot be realized, and the benefit of scientific research group 1 is x_1 , the benefits of scientific research group 2 are x_2 .

The author still follows the traditional approach and uses the replication dynamic equation to describe the dynamic evolution mechanism of the sharing strategy. In comprehensive consideration of the existing literature and space limitations, the details of copying dynamic equations are referred to Hingu [17], Erwin [18] and other achievements. Evolutionary game replication dynamic equation of sharing strategy of heterogeneous scientific research groups:

Scientific research group 1 adopts sharing and non-sharing expected benefits:

$$\begin{cases} \Pi_{yes}^1 = (x_1 + \lambda_1 x_2)q(t) + (x_1 + \lambda_1 \delta_2 x_2 - \Delta x_1)(1 - q(t)) \\ \Pi_{no}^1 = (x_1 + \lambda_1 x_2)q(t) + x_1(1 - q(t)) \end{cases} \quad (1.1)$$

When scientific research groups choose to share expected benefits more than not share expected benefits, the proportion of sharing will increase. On the contrary, it also decreases. Scientific research group 1 evolutionary game replicates dynamic equation:

$$\begin{aligned} dp(t) &= p(t) \cdot (\Pi_{yes}^1 - \Pi_{no}^1) dt \\ &= [-(\lambda_1 \delta_2 x_2 - \Delta x_1)q(t) + (\lambda_1 \delta_2 x_2 - \Delta x_1)] p(t) dt \end{aligned} \quad (1.2)$$

Scientific research group 2 shares and does not share expected benefits:

$$\begin{cases} \Pi_{yes}^2 = (x_2 + \lambda_2 x_1)p(t) + (x_2 + \lambda_2 \delta_1 x_1 - \Delta x_2)(1 - p(t)) \\ \Pi_{no}^2 = (x_2 + \lambda_2 x_1)p(t) + x_2(1 - p(t)) \end{cases} \quad (1.3)$$

Scientific research group 2 evolutionary game replicates dynamic equation:

$$\begin{aligned} dq(t) &= q(t) \cdot (\Pi_{yes}^2 - \Pi_{no}^2) dt \\ &= [-(\lambda_2 \delta_1 x_1 - \Delta x_2)p(t) + (\lambda_2 \delta_1 x_1 - \Delta x_2)] q(t) dt \end{aligned} \quad (1.4)$$

In order to highlight all kinds of random interference in uncertain environment, Gaussian white noise is introduced to describe it. Refer to Xu [19] for details.

$$\begin{cases} dp(t) = [-(\lambda_1 \delta_2 x_2 - \Delta x_1)q(t) + (\lambda_1 \delta_2 x_2 - \Delta x_1)] p(t) dt \\ \quad + \theta_1 \sqrt{(1-p(t))p(t)} d\omega(t) \\ dq(t) = [-(\lambda_2 \delta_1 x_1 - \Delta x_2)p(t) + (\lambda_2 \delta_1 x_1 - \Delta x_2)] q(t) dt \\ \quad + \theta_2 \sqrt{(1-q(t))q(t)} d\omega(t) \end{cases} \quad (1.5)$$

Equilibrium solution and proof of stochastic dynamic systems need to rely on stochastic theory, mainly referring to Huang [20], Hu [21] and other works, without focusing on it. Based on this, the stability conditions of heterogeneous scientific research groups under random interference are obtained:

Lemma 1.1. Gives a stochastic differential equation:

$$dx(t) = f(t, x(t))dt + g(t, x(t))d\omega(t), x(t_0) = x_0. \quad (1.6)$$

Assume existence function $V(t, x)$ and normal numbers c_1, c_2 make

$$c_1 |x|^p \leq V(t, x) \leq c_2 |x|^p, t \geq 0.$$

1) If there is a normal number γ , such that

$$LV(t, x) \leq -\gamma V(t, x), t \geq 0.$$

Then the zero solution of Eq (1.6) is exponentially stable and holds the p -th moment

$$E|x(t, x_0)|^p < (c_2 / c_1) |x|^p e^{-\gamma t}, t \geq 0.$$

2) If there is a normal number γ , such that

$$LV(t, x) \geq \gamma V(t, x), t \geq 0.$$

3) The zero solution of Eq (1.6) is exponentially unstable and holds

$$E|x(t, x_0)|^p \geq (c_2 / c_1) |x|^p e^{-\gamma t}, t \geq 0.$$

Among them, $LV(t, x) = V_x(t, x) + V_x(t, x)f(t, x) + 1/2g^2(t, x)V_{xx}(t, x)$.

Based on Lemma 1.1, obtain the stability judgment criteria for Eq (1.5).

- For scientific research group 1. When $q(t) \geq \frac{\lambda_1 \delta_2 x_2 - \Delta x_1 + 1}{\lambda_1 \delta_2 x_2 - \Delta x_1}$ and $\lambda_1 \delta_2 x_2 - \Delta x_1 < 0$, then the moment of zero solution of shared evolution of scientific research group 1 is exponentially stable; When $q(t) < \frac{\lambda_1 \delta_2 x_2 - \Delta x_1 + 1}{\lambda_1 \delta_2 x_2 - \Delta x_1}$ and $\lambda_1 \delta_2 x_2 - \Delta x_1 > 1$, then the moment index of the shared zero solution of scientific research group 1 is unstable. Scientific research group 1 will eventually give up the sharing strategy and take the non-sharing strategy as a whole after playing with scientific research group 2 for many times under the condition of meeting the index stability; Under the condition of unstable index, scientific research group

1 is more inclined to adopt sharing strategy.

- For scientific research group 2. When $p(t) \geq \frac{\lambda_2 \delta_1 x_1 - \Delta x_2 + 1}{\lambda_2 \delta_1 x_1 - \Delta x_2}$ and $\lambda_2 \delta_1 x_1 - \Delta x_2 < 0$, then the moment of zero solution shared by scientific research group 2 is exponentially stable; When $q(t) < \frac{\lambda_2 \delta_1 x_1 - \Delta x_2 + 1}{\lambda_2 \delta_1 x_1 - \Delta x_2}$ and $\lambda_2 \delta_1 x_1 - \Delta x_2 > 1$, then the moment index of shared zero solution of scientific research group 2 is unstable. Scientific research group 2 will eventually give up the sharing strategy and take the non-sharing strategy as a whole after playing with scientific research group 2 for many times under the condition of meeting the index stability; Under the condition of unstable index, scientific research group 2 is more inclined to adopt sharing strategy.

To sum up, we should meet $q(t) \geq \frac{\lambda_1 \delta_2 x_2 - \Delta x_1 + 1}{\lambda_1 \delta_2 x_2 - \Delta x_1}$ and $\lambda_1 \delta_2 x_2 - \Delta x_1 < 0$,

$p(t) \geq \frac{\lambda_2 \delta_1 x_1 - \Delta x_2 + 1}{\lambda_2 \delta_1 x_1 - \Delta x_2}$ and $\lambda_2 \delta_1 x_1 - \Delta x_2 < 0$, there is a unique stable strategy for shared dynamic

evolution $ESS(0,0)$, that is, scientific research groups will eventually adopt (no sharing, no sharing) strategies. Due to the existence of factors such as opportunism and self-protection awareness of scientific research groups, this non *Nash* equilibrium strategy appears at the end of the game, which is the fundamental reason for the collapse of the scientific data sharing alliance; When

$q(t) < \frac{\lambda_1 \delta_2 x_2 - \Delta x_1 + 1}{\lambda_1 \delta_2 x_2 - \Delta x_1}$ and $\lambda_1 \delta_2 x_2 - \Delta x_1 > 1$, when $q(t) < \frac{\lambda_2 \delta_1 x_1 - \Delta x_2 + 1}{\lambda_2 \delta_1 x_1 - \Delta x_2}$ and $\lambda_2 \delta_1 x_1 - \Delta x_2 > 1$,

At this time, there is a unique stable strategy in the evolution process of the shared strategy $ESS(1,1)$, that is, scientific research groups will eventually adopt (sharing, sharing) strategies.

3.2. Sensitivity analysis of sharing strategy

Based on the above stochastic evolutionary game model, the sharing strategy of scientific research groups ultimately has two stable strategies: (unshared, unshared) and (shared, shared). This is mainly related to the change of some parameters in the benefit matrix. Based on this, sensitivity analysis is carried out on the influencing factors such as the degree of complementarity, willingness to share, random interference and moral hazard. With the help of numerical simulation, the stochastic evolutionary game model is mainly processed by *Milstein* method combined with stochastic *Taylor* expansion and *Itô* stochastic formula. Refer to Huang [20] and Sun [22] for specific algorithms. See Table 3 for initial values of parameters.

Table 3. Initial value of parameters.

Parameters	x_1	x_2	$q(t)$	$p(t)$	θ	λ	δ	Δx_1	Δx_2	h
Initial value	5	4	0.5	0.5	0.5	3	0.1	5	4	0.01

A. Complementarity

In the era of big science, the interdisciplinary integration makes any scientific research group only obtain limited information resources in limited time, space and discipline. Scientific data sharing is to allow massive information resources to flow in scientific research activities and maximize the role of

scientific data resources. However, due to the diversity of scientific data forms, different collection methods and purposes, there are differences in scientific data resources of heterogeneous scientific research groups. Research findings (Figure 2): The degree of complementarity of scientific data can effectively reduce the impact of random interference. The greater the degree of complementarity, the smaller the fluctuation of random interference; The degree of complementarity is in direct proportion to the convergence rate of sharing strategies of scientific research groups. The greater the degree of complementarity, the greater the probability of scientific research groups choosing sharing strategies.

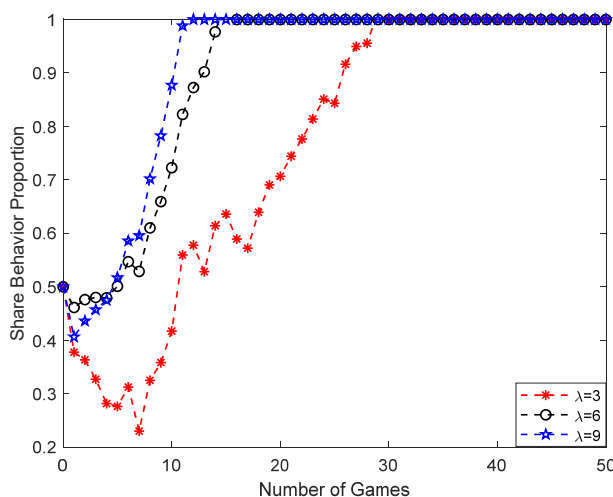


Figure 2. Simulation diagram of complementarity change.

B. Willingness to share

In the face of today's interdisciplinary and integrated development of scientific research environment, scientific data plays an important role in supporting scientific and technological innovation. Scientific researchers want to obtain the original data of others but are not willing to share their own original data, resulting in low willingness to share scientific data and difficulty in sharing scientific data. The belief, reputation and self-worth of scientific researchers are important factors that directly lead to the low willingness to share. The willingness to share is a parameter to describe the psychological contract. The research findings (Figure 3): The sharing intention has no significant impact on group decision-making choices. With the increase of sharing intention, the convergence rate of sharing strategy of scientific research groups has not significantly improved. The main reason is that: the willingness to share is a subjective indicator of individuals, which has no significant impact on group decision-making. Enhancing the willingness to share will not significantly improve the convergence rate of scientific research group sharing strategy.

C. Random interference

In the scientific data sharing ecosystem, the complexity of the relationship between scientific research groups and the uncertainty of individual behavior lead to the randomness of sharing behavior selection. Scientific research groups are composed of many individuals, whose behavior choices are affected by their own and environmental factors, and have the characteristics of random evolution. The choice of scientific research group behavior is not a simple superposition of individual behavior choices, but is determined by the evolution of a large number of individual choices, showing some new characteristics and new laws at the group level. If the process of scientific research group selection is regarded as a process of trial and error, random interference is the threshold in trial and error, which

can correct or fine-tune the behavior selection of scientific research group according to the surrounding historical information. Research findings (Figure 4): Random interference has a significant impact on individual sharing behavior. The random evolution trajectory shows a non-smooth curve fluctuation. The random interference will only change the local stable state of individual sharing, and will not change the stable state of scientific research group sharing. In addition, random interference will slow down the convergence rate of scientific research groups. The stronger the random interference, the slower the stability rate. In reality, when faced with increasing uncertainty risks, participants need enough time to observe the other party's decision changes to develop their own strategies.

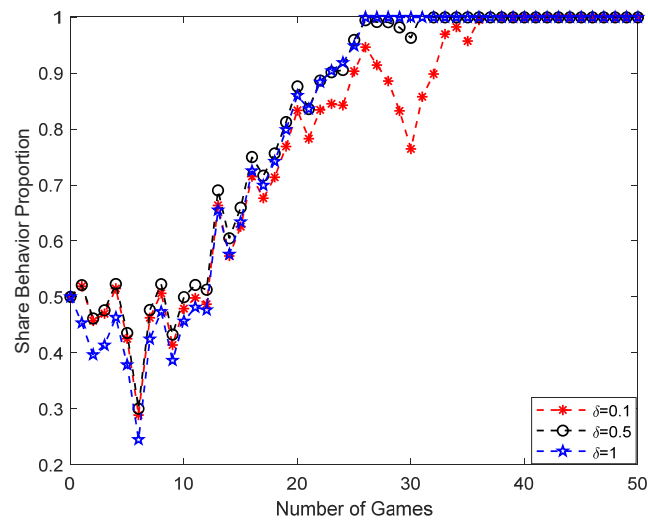


Figure 3. Simulation diagram of sharing intention change.

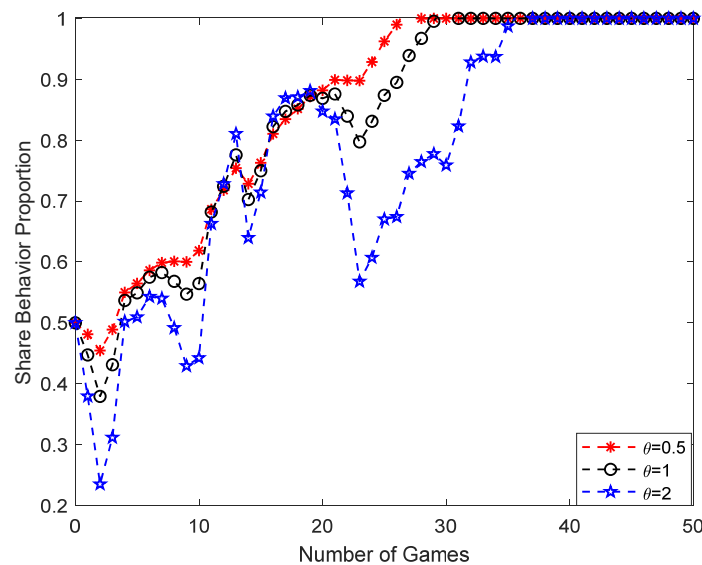


Figure 4. Simulation diagram of random interference intensity change.

D. Moral hazard

Heterogeneous scientific research groups voluntarily form strategic alliances in order to seek data

sharing. Alliance members often take advantage of their own information advantages to adopt strategies that cannot be observed and monitored by the other party, thus causing the objective existence of moral hazard for alliance members. Moral hazard can only reduce the probability of moral hazard by optimizing the shared ecological environment, and cannot completely avoid the occurrence of moral hazard. The study found (Figure 5) that moral hazard was significantly negatively correlated with the rate of group yield. The greater the moral hazard, the slower the convergence rate. The main reason is that moral hazard increases the risk of group selection of sharing strategies. When some individual decisions have a large proportion of non-sharing, group decisions need sufficient time to seek the optimal sharing strategy in the process of trial and error.

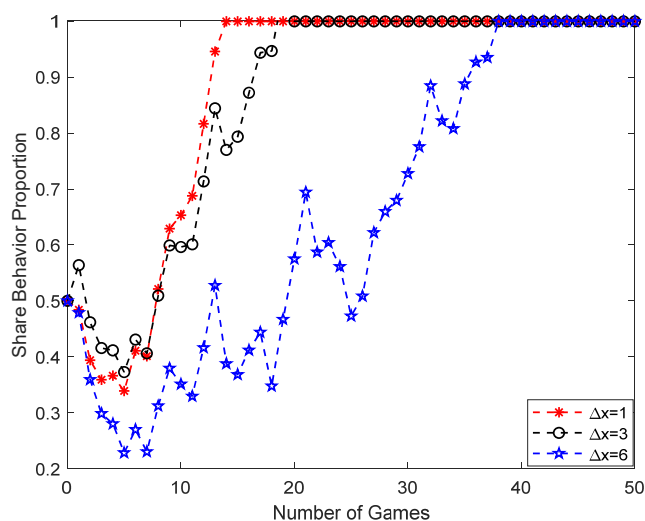


Figure 5. Simulation of moral hazard change.

To sum up, for the stability of scientific research group sharing evolution, the degree of scientific data complementarity is a positive indicator, while external random interference and network moral hazard are negative indicators, and individual sharing willingness has no significant impact on group sharing behavior; The evolution trajectory of shared strategy under uncertain environment is a non-smooth curve, which is basically consistent with the evolution law under certain environment, but there are differences in stability solution and judgment.

4. Incentive mechanism of sharing strategy from the perspective of contract

4.1. A Bidirectional principal-agent analysis of the incentive mechanism of shared strategies

Heterogeneous scientific research groups voluntarily form strategic alliances in order to maximize their individual utility functions, and have their own reserved utility functions to share scientific data with each other, which is a typical two-way principal-agent relationship. The two-way principal-agent relationship of alliance members is constrained by the explicit contract mechanism, and both sides are at a strategic disadvantage in the case of asymmetric information. How to design a sharing strategy incentive mechanism for third parties (governments, platforms, scientific research groups, etc.) to balance benefits and risks? How to maximize the benefits of the alliance?

Following the traditional principal-agent theory paradigm, this paper discusses the optimal

incentive contract design from the perspective of principal and agent. Alliance members (scientific research groups) have the dual identities of principal and agent, construct a virtual member to represent the strategic alliance, and ask the alliance members to turn in the right of entrustment to the virtual member. Virtual members (strategic alliances) are called "virtual principals", and alliance members (scientific research groups) who turn in the power of attorney are called "agents". Build a two-way principal-agent model after the introduction of virtual principals:

$$\begin{aligned} & \max F(x_1, x_2, \alpha) \\ & s.t \begin{cases} F_1(x_1, x_2, \alpha, \beta, w_1, w_2) \geq \bar{w}_1 \\ F_2(x_1, x_2, \alpha, \beta, w_1, w_2) \geq \bar{w}_2 \\ x_1, x_2 \in N \end{cases} \end{aligned} \quad (4.1)$$

See Table 1 for the meaning of variables in model (4.1). F , F_1 and F_2 represent the utility functions of virtual client, scientific research group 1 and scientific research group 2, and the utility functions are concave functions. They represent the utility functions of virtual client, scientific research group 1 and scientific research group 2, and the utility functions are concave functions. For the convenience of discussion, the costs shared by scientific research groups are as follows: $C_1(x_1) = \frac{1}{2}b_1x_1^2$ and $C_2(x_2) = \frac{1}{2}b_2x_2^2$, use *Cobb-Douglas* production function to describe the overall income of strategic alliance, then virtual principal production function $Y = x_1^\alpha x_2^{1-\alpha} + \varepsilon$.

Among them, ε is a normal distribution random variable with zero mean and variance equal to σ^2 , it represents exogenous uncertainties. Then the expected output of the virtual client is $E(Y) = x_1^\alpha x_2^{1-\alpha}$. When designing incentive contracts, use Holmstrom et al. [23] proposed a linear function as the distribution function of the alliance members' income: $y_1 = w_1 + \frac{1}{2}\beta(Y - w_1 - w_2)$; $y_2 = w_2 + \frac{1}{2}\beta(Y - w_1 - w_2)$. Among them, the utility function of the alliance members has an invariable absolute risk aversion, and the risk cost $\frac{1}{2}\rho\beta^2\sigma^2$ proposed by Zhang [24] is used to represent the random risk. Then:

Virtual client utility function:

$$F(x_1, x_2, \alpha) = E(Y) - C_1(x_1) - C_2(x_2) - \frac{1}{2}\rho\beta^2\sigma^2.$$

Utility function of scientific research group 1 and 2:

$$\begin{aligned} F_1(x_1, x_2, \alpha, \beta, w_1, w_2) &= y_1 - C_1(x_1) - \frac{1}{4}\rho\beta^2\sigma^2 \\ F_2(x_1, x_2, \alpha, \beta, w_1, w_2) &= y_2 - C_2(x_2) - \frac{1}{4}\rho\beta^2\sigma^2 \end{aligned}$$

The two-way principal-agent model (4.1) can be converted into model (4.2):

$$\begin{aligned} & \max \left\{ x_1^\alpha x_2^{1-\alpha} - \frac{1}{2} b_1 x_1^2 - \frac{1}{2} b_2 x_2^2 - \frac{1}{2} \rho \beta^2 \sigma^2 \right\} \\ & \text{s.t.} \begin{cases} w_1 + \frac{1}{2} \beta (x_1^\alpha x_2^{1-\alpha} - w_1 - w_2) - \frac{1}{2} b_1 x_1^2 - \frac{1}{4} \rho \beta^2 \sigma^2 \geq \bar{w}_1 \\ w_2 + \frac{1}{2} \beta (x_1^\alpha x_2^{1-\alpha} - w_1 - w_2) - \frac{1}{2} b_2 x_2^2 - \frac{1}{4} \rho \beta^2 \sigma^2 \geq \bar{w}_2 \\ x_1, x_2 \in N \end{cases} \end{aligned} \quad (4.2)$$

For solving the optimization problem of model (4.2), the optimal solution is obtained by constructing the Lagrange function. Optimize two first-order conditions to obtain:

$$x_1^* = x_2^* \cdot \sqrt{\frac{b_2}{b_1} \cdot \frac{\alpha}{1-\alpha}}. \quad (4.3)$$

Formula (4.3) shows that under the optimal solution of model (4.2), the optimal sharing amount of scientific data between scientific research group 1 and scientific research group 2 is linear, inversely proportional to the sharing cost coefficient and directly proportional to the output elasticity coefficient.

4.2. Optimal contract design of incentive mechanism for sharing strategy

The incentive of scientific researchers is the core issue of scientific data sharing and sharing, and the main explicit constraint mechanism is incentive contract. For the third party, how to design the optimal incentive contract is the key issue of whether the data sharing ecosystem can be healthy and sustainable, and is also an effective means to achieve open data sharing. It is impossible for one member of the alliance to use the mandatory contract to force the other party to adopt the desired strategy, but can only use the incentive contract to induce the other party to choose the desired strategy. For simplicity, suppose $x_1 = x_2 = x$. Model (4.2) can be simplified into model (4.4):

$$\begin{aligned} & \max_{x, \beta} F = \left\{ x - \frac{1}{2} (b_1 + b_2) x^2 - \frac{1}{2} \rho \beta^2 \sigma^2 \right\} \\ & \text{s.t.} \begin{cases} w_1 + \frac{1}{2} \beta (x - w_1 - w_2) - \frac{1}{2} b_1 x^2 - \frac{1}{4} \rho \beta^2 \sigma^2 \geq \bar{w}_1 \\ w_2 + \frac{1}{2} \beta (x - w_1 - w_2) - \frac{1}{2} b_2 x^2 - \frac{1}{4} \rho \beta^2 \sigma^2 \geq \bar{w}_2 \\ x \in N \end{cases} \end{aligned} \quad (4.4)$$

The first-order condition of model (4.4) optimization means:

$$x^* = \frac{1}{b_1 + b_2}; \beta^* = 0. \quad (4.5)$$

Bring formula (4.5) into the constraint of scientific research group participation:

$$w_1 = \bar{w}_1 + \frac{1}{2}b_1 \frac{1}{(b_1 + b_2)^2}; \quad w_2 = \bar{w}_2 + \frac{1}{2}b_2 \frac{1}{(b_1 + b_2)^2}. \quad (4.5)$$

Formula (4.5) indicates that the Pareto optimal contract under the condition of information symmetry requires the members of the alliance not to bear any risk ($\beta^* = 0$), and the members' income is composed of two parts: the retained wage and the shared cost, which is independent of the amount of scientific data sharing. In the cooperative work of actual alliance members, Pareto optimality under the condition of information symmetry is obviously not achievable. Because: it is impossible for one member of the alliance to completely and truly observe the amount of scientific data shared by the other, the member will choose to maximize their utility function. Given $\beta = 0$, the income of the alliance member is independent of the amount of data shared, then $x = \beta / (b_1 + b_2) = 0$. This shows that under the condition that the alliance members do not bear any risks, they will choose to adopt the non-sharing strategy to minimize the sharing costs.

Next, we discuss the optimal incentive contract under the condition of asymmetric information. Given (x, β) , the incentive compatibility constraint of the alliance members means $x = \beta / (b_1 + b_2)$ and the problem of the virtual client is to choose the optimal (x, y) solution model (4.6) optimization problem.

$$\begin{aligned} & \max \left\{ x - \frac{1}{2}(b_1 + b_2)x^2 - \frac{1}{2}\rho\beta^2\sigma^2 \right\} \\ & \text{s.t.} \left\{ \begin{array}{l} w_1 + \frac{1}{2}\beta(x - w_1 - w_2) - \frac{1}{2}b_1x^2 - \frac{1}{4}\rho\beta^2\sigma^2 \geq \bar{w}_1 \\ w_2 + \frac{1}{2}\beta(x - w_1 - w_2) - \frac{1}{2}b_2x^2 - \frac{1}{4}\rho\beta^2\sigma^2 \geq \bar{w}_2 \\ x = \frac{\beta}{b_1 + b_2} \\ x \in N \end{array} \right. \end{aligned} \quad (4.6)$$

The optimal first-order condition of model (4.6) means:

$$\begin{aligned} \beta &= \frac{1}{1 + (b_1 + b_2)\rho\sigma^2} \\ x &= \frac{1}{(b_1 + b_2) + (b_1 + b_2)^2\rho\sigma^2} \end{aligned} \quad (4.7)$$

Formula (4.7) indicates that the members of the alliance must bear certain risks, and the size of the risks is inversely proportional to the cost coefficient, risk aversion and output variance. In short, the more members of the alliance are afraid of scientific data sharing, the less risk they will take. Because under a certain cost coefficient, in order to encourage the members of the alliance to choose the same amount of data sharing, the greater the risk, the virtual client would prefer to exchange less data for saving the risk cost.

5. Conclusions and enlightenment

5.1. Conclusions

By building a scientific big data sharing ecosystem, this paper proposes the psychological contract and the explicit and implicit constraint mechanism of the contract contract. On this basis, the evolution mechanism of the sharing strategy and the incentive mechanism of the sharing strategy are discussed respectively. The main conclusions are as follows:

A. The alliance of heterogeneous scientific research groups cooperates to carry out innovative research, which is bound by psychological contract and contract contract. The scientific big data sharing and sharing ecosystem includes six links: demand analysis, data collection, data processing, heterogeneous data amplification effect, data configuration and data empowerment. Data is the flowing energy source in the ecosystem.

B. The solution of stochastic evolutionary game equilibrium conditions is more complex than the solution of evolutionary game equilibrium conditions. The former has stronger explanatory power for the evolution of scientific research group sharing behavior. The evolution trajectory of stochastic evolutionary game is a non-smooth curve, which is closer to the actual problem. The degree of complementarity of scientific big data is a positive indicator, while external random interference and network moral hazard are negative indicators. Individual sharing willingness does not affect the evolution and convergence rate of group sharing strategy.

C. The two-way principal-agent theory can explain the mutual principal-agent relationship of alliance cooperation, providing a theoretical basis and research paradigm for the follow-up research. The optimal incentive contract consists of fixed wage and incentive wage, which is an increasing function of risk. Risks must exist and can only be reduced by optimizing the sharing environment. The more researchers are afraid of sharing, the smaller the risk cost they bear. The core problem is to find the optimal solution between incentive cost and risk cost.

5.2. Enlightenment

At present, there is still a certain gap between China's science big data sharing business and that of developed countries such as Europe and the United States. Although China's science big data collection capacity has been continuously improved, the international policy level is still relatively weak. In view of the weak links in China's science big data sharing, combined with the contents of this paper, the policy implications are as follows:

A. Create a healthy scientific big data sharing ecosystem and improve the conversion efficiency of mobile energy sources. The scientific research community and the shared environment are an inseparable whole, which is called the ecosystem. It can be seen from Figure 1 that the amplification effect of heterogeneous data is the core link of ecosystem health, efficiency and evolution. We should learn from the successful experience of developed countries in Europe and the United States to support the formation of a number of national-level scientific big data sharing centers, and strengthen the standardization of scientific big data collection and production, processing and processing, open sharing and other links.

B. Establish and improve the scientific big data sharing legal system and strengthen the protection of intellectual property rights. This paper believes that it can be carried out from three aspects: first, improve the trading mechanism of scientific big data and make scientific big data realize formal trading. Big data as a resource must have typical trading characteristics; Second, improve the punishment

mechanism for scientific big data abuse and protect the rights and interests of the original data owners. Big data as a resource will inevitably have theft, abuse, counterfeiting and other behaviors; Third, improve the scientific big data sharing mechanism, strictly follow the principles of openness, fairness and equity to carry out data sharing and sharing, and minimize the costs caused by random interference and moral hazard.

C. Establish and optimize the incentive mechanism for scientific big data sharing, and incorporate the sharing behavior into the performance evaluation. This paper believes that suitable psychological contract and incentive contract are the two main mechanisms to constrain the sharing behavior of scientific big data. For the special relationship between cooperative work of heterogeneous scientific research groups, new theories should be adopted to explain. The choice of sharing strategies for scientific research groups is ultimately a matter of interest distribution. The government should fully consider the proportion of fixed wages and incentive wages when formulating policies, and should design incentive contracts that change with risk.

5.3. Future research directions

Application of blockchain technology. Integrate blockchain technology into the system to enhance security, transparency and traceability. Trusted collaboration, where data is available but not visible. The encryption algorithm of blockchain technology enables permission management for different entities involved in data sharing, effectively reducing security risks during the data sharing process, providing users with strong privacy protection and ensuring the security and accuracy of data during transmission. At the same time, the data that needs to be shared is stored on the chain. If there is a data ownership dispute in the future, records can be searched on the chain [25]. Authorization on the data chain and sharing off the chain. By building a data sharing blockchain alliance chain, data providers import data information according to settings. Assuming that one party wants to use the data inside, they can view who has the data they want through the blockchain network and then apply on the chain. This operation will also be provided to the data provider synchronously. After the other party sees the application information, they will place the authorization information on the chain to complete data authorization. At this time, the data user can obtain the data from the data provider through offline means. This is the on chain authorization and off chain sharing, which has a wide range of applications, especially in the supply chain field [26].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Social Science Foundation Program of China (No. 22CJY040).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. R. Kitchin, Big data, new epistemologies and paradigm shift, *Big Data Soc.*, **1** (2014), 1–12.
2. H. Guo, L. Wang, C. Fang, D. Liang, Scientific big data and digital earth, *Chinese Sci. Bull.*, **59** (2014), 5066–5073. <https://doi.org/10.1007/s11434-014-0645-3>
3. C. L. Borgman, The conundrum of sharing research data, *New Technol. Library Inf. Ser.*, **5** (2013), 1–20. <https://doi.org/10.11925/infotech.1003-3513.2013.05.01>
4. Z. Yang, Y. Shi, Y. Li, Analysis of intellectual property cooperation behavior and its simulation under two types of scenarios using evolutionary game theory, *Comput. Ind. Eng.*, **125** (2018), 739–750. <https://doi.org/10.1016/j.cie.2018.02.040>
5. J. Zuo, Y. Chen, The analysis on sharing mode of scientific data in the environment of big data, *J. Intell.*, **32** (2013), 151–154. <https://doi.org/10.1007/978-3-319-50627-2>
6. D. S. Sayogo, T. A. Pardo, Exploring the determinants of scientific data sharing: understanding the motivation to publish research data, *Gov. Inform. Q.*, **30** (2013), 519–531. <https://doi.org/10.1016/j.giq.2012.06.011>
7. C. Li, L. Zhang, Scientific big data opening and sharing: models and mechanisms, *Inf. Stud.: Theory Pract.*, **40** (2017), 45–51.
8. D. L. Nazareth, J. Choi, A system dynamics model for information security management, *Inf. Manage.*, **52** (2015), 123–134. <https://doi.org/10.1016/j.im.2014.10.009>
9. Y. Du, R. Zhang, M. Li, Research on a security mechanism for cloud computing based on virtualization, *Telecommun. Syst.*, **53** (2013), 19–24. <https://doi.org/10.1007/s11235-013-9672-7>
10. Y. Kim, M. Adler, Social scientists' data sharing behaviors: investigating the roles of individual motivations, institutional pressures, and data repositories, *Int. J. Inf. Manag.*, **35** (2015), 408–418. <https://doi.org/10.1016/j.ijinfomgt.2015.04.007>
11. M. Janssen, Y. Charalabidis, A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, *Inf. Syst. Manag.*, **29** (2012), 258–268. <https://doi.org/10.1080/10580530.2012.716740>
12. B. V. Loenen, H. J. Onsrud, Geographic data for academic research: assessing access policies, *Cartogr. Geogr. Inf. Sci.*, **31** (2004), 3–17. <https://doi.org/10.1559/152304004773112730>
13. C. Peng, X. Song, H. Jiang, Q. Zhu, H. Chen, J. M. Chen, et al., Towards a paradigm for open and free sharing of scientific data on global change science in China, *Ecosyst. Health Sustain.*, **2** (2016), 23–45. <https://doi.org/10.1002/ehs2.1225>
14. X. Gao, W. Zhong, S. Mei, Stochastic evolutionary game dynamics and their selection mechanisms, *Comput. Econ.*, **41** (2013), 233–247. <https://doi.org/10.1007/s10614-012-9320-4>
15. S. Dai, G. Nie, N. Xiao, The study of the two-way principal-agent model based on asymmetric information, *Wireless Pers. Commun.*, **102** (2018), 629–639. <https://doi.org/10.1007/s11277-017-5071-9>
16. Y. Wei, G. Liu, Analysis of research data management and sharing policy in foreign universities based on data lifecycle, *J. Intell.*, **5** (2017), 153–158.
17. D. Hingu, K. S. M. Rao, A. J. Shaiju, Evolutionary stability of polymorphic population states in continuous games, *Dyn. Games Appl.*, **8** (2016), 141–156. <https://doi.org/10.1007/s13235-016-0207-1>
18. A. Erwin, P. Alex, On the stability of evolutionary dynamics in games with incomplete information, *Math. Soc. Sci.*, **58** (2009), 310–321.
19. Y. Xu, B. Hu, R. Qian, Analysis on stability of strategic alliances based on stochastic evolutionary game including simulations, *Syst. Eng.-Theory Pract.*, **31** (2011), 920–926.

20. J. M. Huang, H. W. Zhang, A method for selecting defense strategies based on stochastic evolutionary game model, *Acta Electron. Sinica*, **46** (2018), 2222–2228. <https://doi.org/10.3969/j.issn.0372-2112.2018.09.025>
21. S. Hu, C. Huang, *Stochastic differential equation*, Beijing: Science Press, **5** (2008), 66–70.
22. H. Sun, X. Wang, Y. Xue, Stochastic evolutionary game model for unexpected incidents involving mass participation based on different scenarios, *Operat. Res. Manage. Sci.*, **25** (2016), 23–29.
23. B. Holmstrom, P. Milgrom, Aggregation and linearity in the provision of intertemporal incentives, *Econometrica*, **55** (1987), 303–328.
24. W. Zhang, *Game theory and information economics*, Shanghai: People's Publishing House, **5** (2012), 242–262.
25. Y. Ren, F. Zhu, P. K. Sharma, T. Wang, J. Wang, O. Alifarraaj, et al., Data query mechanism based on hash computing power of blockchain in internet of things, *Sensors*, **20** (2020), 207–229. <https://doi.org/10.3390/s20010207>
26. Y. Ren, F. Zhu, J. Wang, P. K. Sharma, U. Ghosh, Novel vote scheme for decision-making feedback based on blockchain in internet of vehicles, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 1639–1648. <https://doi.org/10.1109/TITS.2021.3100103>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)