ソシオネットワーク戦略ディスカッションペーパーシリーズ ISSN 2434-9445
第 122 号　2025 年 4 月
RISS Discussion Paper Series
No.122　April, 2025

Collaborative knowledge creation and size of institutions

Kentaro Hatsumi

文部科学大臣認定　共同利用・共同研究拠点

関西大学ソシオネットワーク戦略研究機構

Research Institute for Socionetwork Strategies,
Kansai University
Joint Usage / Research Center, MEXT, Japan
Suita, Osaka, 564-8680, Japan
URL: https://www.kansai-u.ac.jp/riss/index.html
e-mail: riss@ml.kandai.jp
tel. 06-6368-1228
fax. 06-6330-3304

Collaborative knowledge creation and size of institutions

Author. Kentaro Hatsumi
Address. Faculty of Policy Studies, Kansai University, Suita, Osaka 564-8680, Japan
Email. hatsumi@kansai-u.ac.jp
This Version. February 21, 2025
First Presented. March 1, 2024

Keywords. computational simulation; innovation and invention; organization study

Abstract. We investigate the collaborative knowledge-creation process within an institution using the computational simulation method. The key assumption is that both common and different knowledge previously possessed by agents who conduct collaborative knowledge creation affects productivity sensitively, which is borrowed from Berliant and Fujita. The main question is how institution size affects productivity. If matching by agents for collaboration occurs randomly and there is transmission of ideas between agents, or if similar ability agents match frequently, institution size and productivity have an inverted-U shaped relationship. These findings may be helpful in understanding the appropriate size of institutions for knowledge creation.

1 Introduction

How institution size affects knowledge creation activities, such as industrial innovation and university research, has been a long-standing problem and has been studied from various viewpoints in economics and management science (Acs and Audretsch, 1987; Damanpour, 1992; Cohen and Klepper, 1996; Teirlinck, 2017). However, this topic remains complicated and is not yet fully understood. Empirical studies present conflicting views. Camisón-Zornoza et al (2004) confirm the existence of a positive relationship between institution size and productivity, whereas Shefer and Frenkel (2005) and Mote et al (2016) report a negative relationship[1]. This paper reexamines this classical problem from a micro-focused viewpoint on innovative organizations. We construct a computational simulation model with a single institution that can be interpreted as a firm's R&D division or a university department. We do not consider any factors outside the institution[2]. Agents (knowledge workers) in this institution repeatedly engage in joint trials of knowledge creation, and the success rate of each trial depends on the numbers of both common and differentiated ideas previously possessed by the agents. This assumption originates from Berliant and Fujita (2008), and we convert their deterministic formulation into a probabilistic framework. We propose several plausible settings for agents' matching procedures and the role

[1]The optimal project team size for research activities is a topic relevant to this study. Wu et al (2019) report that small teams tend to disrupt science and technology with new ideas, whereas large teams tend to develop existing ones. Hu et al (2021) find an inverted-U shaped relationship between team size and high-quality innovation in the pharmaceutical patent data.

[2]Some studies on this topic emphasize the importance of an institution's financial accessibility for innovation projects (Czarnitzki and Hottenrott, 2011) and collaboration networks with external members (Andries and Thorwarth, 2014), whereas we do not consider these factors.

of knowledge and demonstrate that, depending on the setting, there is a different appropriate institution size for maximizing productivity of knowledge creation. Moreover, under some of these settings, a medium-sized institution appears to be the most effective. We derive several insightful results solely from agents' internal interactions within an institution.

The structure of the *basic* simulation model is as follows. First, we fix the number of agents in the institution (institution size), which is even. There are several discrete *periods* that constitute a *run*, which is the unit of simulation. In each period of a run, every agent matches with another agent and forms a pair. A pair of agents conducts a joint trial of the creation of an idea, which is a unit of knowledge, with a certain success rate. The number of ideas that both agents possess in *common* and the number of *differentiated* ideas that one of the two agents solely possesses positively and sensitively affect the success rate. If the joint creation is a success, both agents receive the created idea. Therefore, if the period proceeds, the success rate increases because the number of each agent's ideas increases. At the end of a run, we can measure each agent's *individual productivity*, which is the total number of ideas that she contributes and the main objective of this paper. The total number of periods in each run is fixed at 20 throughout the paper[3]. We allow the number of agents to range from 4 to 40. For each fixed number of agents, we simulate enough runs to generate data for analysis.

This paper considers two types of matching procedures and three types of model themes. Hence, we have $2 \cdot 3 = 6$ models. We use *random* and *ability-ordered* matching procedures. Even though random matching among agents is not often observed in actual situations, this is one of the simplest forms of matching and is worth analyzing as a *benchmark*. In addition, random matching is enforceable if the administrator of an institution with sufficient authority believes that the randomization of members has merit. In the ability-ordered matching procedure, high-ability workers are more likely to match with each other, as are low-ability workers. This is sometimes observed in actual institutions as a spontaneous procedure of members. This matching is indeed incentive compatible if agents are quite myopic and have no memory of past matching partners. For each of the two matching procedures, in addition to the basic model explained in the previous paragraph, we consider models with *education* and the *transmission* of ideas. The model with education adds to the basic model the preliminary education that provides a common idea for each member with a certain success rate. The model with the transmission of ideas adds the transmission of a differentiated idea from the agent who possesses it to the other, who does not possess it, during the matching for joint idea creation. In addition, in each of the $2 \cdot 3 = 6$ models, we consider three parameter settings for the success rate of joint creation: (i) the number of differentiated ideas highly affects the success rate, (ii) the numbers of both common and differentiated ideas have a balanced impact on it, and (iii) the number of common ideas highly affects it.

The results for the main question are summarized in Table 1. For the **inverted-U** shaped relationship, (average) *productivity* increases with the number of agents until a certain point, which we sometimes call the *peak*, and then decreases. For the **plateau** shaped relationship, productivity increases with the number of agents until a certain point similar to the inverted-U

[3]We briefly review an alternative parameter setting for the total number of periods in Section 5.

	(i) differentiation	(ii) medium	(iii) common
RndBM	decreasing	decreasing	decreasing
RndEdM	plateau	decreasing	decreasing
RndTmM	inverted-U	inverted-U	inverted-U
AbiBM	inverted-U	inverted-U	inverted-U
AbiEdM	increasing	plateau	inverted-U
AbiTmM	inverted-U	inverted-U	inverted-U

Table 1: The overall results
Note. The prefixes "Rnd" and "Abi" mean random and ability-ordered, respectively. The suffixes "BM", "EdM", and "TmM" mean the basic model, the model with education, and the model with transmission of ideas, respectively. The cases (i), (ii) and (iii) coinside with different parameters for the success rate calculation in the main part of this paper, which are $q = 0.1, 0.5$ and 0.9, respectively.

shaped relationship and then remains at roughly the same value. Here, we note the three main findings. First, the inverted-U shaped relationship is observed in many models, especially when the transmission of ideas among agents occurs or ability-ordered matching is employed. Second, if the number of agents is large, a lack of common ideas among agents is a severe bottleneck for the improvement of productivity except in the case where education works well. Third, where there is a transmission of ideas, peak productivity is higher in the random model than in the ability-ordered model. Hence, in this case, introducing randomization of matching partners in the institution could be helpful for improving productivity.

The structure of this paper is as follows. Section 2 provides a brief review of the related literature. Section 3 describes the models with random matching and the results. Section 4 describes the models with ability-ordered matching and the results. Section 5 discusses alternative parameter settings of the simulation and provides concluding remarks. The electronic supplementary material contains notes for the alternative parameter settings, tables that contain more detailed numerical results omitted from this paper, and the program codes for the simulation.

2 Related Literature

This paper builds on the assumption introduced by Berliant and Fujita (2008) that the productivity of collaborative knowledge creation depends on both differentiated and common ideas among agents[4]. Berliant and Fujita (2008) calculate the optimal group size formed by *rational* knowledge workers[5]. Although Berliant and Fujita (2008) and this paper share a similar

[4]Empirical studies, such as Horwitz and Horwitz (2007), Østergaard et al (2011), and Huo et al (2019), suggest that diversity in acquired knowledge, such as educational background and areas of expertise among members, positively affects innovative outputs. By contrast, several studies, including Nonaka and Takeuchi (1995), Wang and Chen (2010), and Hsiao and Hsu (2018), emphasize the importance of common knowledge among members, such as firm-specific knowledge in innovative activities. These two views, which may seem opposed at first glance, may support the assumption that both common and differentiated ideas among agents are crucial. This assumption may also be considered an extension of the "cognitive distance" hypothesis by Nooteboom (2000) that there is an inverted-U shaped effect of cognitive distance between collaborators on innovation performance. Nooteboom et al (2007) confirm this hypothesis empirically for an inter-firm collaboration case.

[5]Even though Berliant and Fujita (2008) refer to their group formation concept as the "myopic core," the agents in their model are rational enough to form a coalition.

objective of determining the optimal group (or institution) size for knowledge creation, there are two fundamental differences. First, we employ a computational and probabilistic approach, whereas they use a mathematical and deterministic approach. Our approach is so simple that it is easy to add extra settings, such as education and knowledge transmission, into the basic framework. Second, we consider agents' matching for collaboration that is quite different from rational agents' stable one. In an actual institution, there are not only agents' myopic decisions but also some enforced matching by the administrator. This paper takes such situations into account.

This study can also be understood from the perspective of computational simulation studies in economics and management science. Although the current study does not adopt a network science approach, research on knowledge diffusion and creation over networks is a major trend within this perspective. Cowan and Jonard (2004), Delre et al (2007), Qiao et al (2019), and Ozman and Parker (2023) all examine knowledge diffusion on networks. König et al (2011) consider how collaborative networks affect firms' knowledge productivity. Tur and Azagra-Caro (2018) develop a model that incorporates the interaction between agents' knowledge creation and collaborative network formation. In their model, both positive and negative feedback loops emerge between knowledge creation and collaboration activities. Müller et al (2021) represent knowledge as a complex network and examine firms' collaboration in discovering new knowledge. Caminati (2016) assumes that the degree of dissimilarity between agents' knowledge portfolios affects collaborative productivity and examines how this assumption influences the formation of collaborative networks. van den Bergh (2008) constructs an innovation model that is different from the network science approach and derives the long-term importance of maintaining technological diversity. Gräbner (2016), Wall (2016), and Axtell and Farmer (2025) provide comprehensive surveys of computational simulation studies in economics and management science.

3 Random matching models

We consider a **model** to have a formal and foundational structure of the simulation procedure that is used to generate data. In this section, we consider three types of random matching models; basic model (**RndBM**), model with education (**RndEdM**), and model with transmission of ideas (**RndTmM**). The details of each model are described in the following subsections. In the random matching model, agents are extremely myopic, and one may consider this model unrealistic. Here, we provide two reasons why random matching models are worth considering. First, random matching is one of the simplest forms of matching, so this is our *bench mark* when comparing with a more realistic model. Second, it is often observed in some institutions that the administrator enforces the randomization of junior members among several groups. Such random matching may roughly approximate such an enforceable situation.

3.1 Basic model (RndBM)

Some terms and notations are borrowed from Berliant and Fujita (2008) and more recently Mori and Sakaguchi (2019), although our model is substantially different from theirs[6]. A **run** is a unit of the simulation process that generates data for analysis. An arbitrary period is denoted by r. In a run, there is a discrete and finite number of **periods** $0, \dots, T-1$[7]. An arbitrary period is denoted by t. T is the total number of periods in a run.

There are **agents** $0, \dots, n-1$. Arbitrary agents are denoted by i, j, and so on. Then, the number of agents is n. We assume that n is an even number. **Ideas** $0, 1, 2, \dots$ are basic pieces of knowledge. An arbitrary idea is denoted by x. The number of potential ideas is infinite. Any idea is treated symmetrically. In other words, we do not consider the quality of each idea.

In each period t, an agent *randomly* matches with another agent and conducts a joint creation of new ideas. Every agent can match with someone else once and only once in each period. As the number of agents is n and this is even, the number of matches in each period is $n/2$. Joint creations by pairs of agents in each period are sequentially executed and the order is randomly assigned[8].

If the joint creation of an idea by agents i and j is a success, then both receive the same new idea. If there are ideas $0, \dots, x-1$ created by any pairs in the previous and current periods before the joint creation, then the created new idea is named x. If the joint creation is a failure, they receive nothing. The success rate, the calculation of which is explained in the following paragraphs, is always positive. An agent is assumed to be better off creating new ideas as much as possible. An interpretation of this preference is that she can earn income through patent license fees for created ideas that she contributes to. Under this interpretation, it is natural that she has an incentive to conduct a joint trial of idea creation in each period[9].

Let y_i^t denote the number of total ideas that agent i has contributed to create and possesses at the beginning of period t. In the basic model, we assume that no agent has an idea at the beginning of period 0, *i.e.*, for all agents i, $y_i^0 = 0$. Let c_{ij}^t denote the number of agents i and j's **common ideas**, which are the total number of ideas that both i and j possess in common, at the beginning of period t. Note that $c_{ij}^t = c_{ji}^t$. Let d_{ij}^t denote the number of agent i's **differentiated ideas** from agent j's, which are the total number of ideas that i possesses, but j does not have, at the beginning of period t. Note that $y_i^t = c_{ij}^t + d_{ij}^t$.

The success rate of joint creation of a new idea p_{ij}^t by agents i and j at period t follows the composition function:

$$p_{ij}^t = p(f_{ij}^t) = \frac{\ell}{1 + \exp(-k(f_{ij}^t - m))},$$

[6]Mori and Sakaguchi (2019) is an empirical study of collaborative knowledge creation based on Berliant and Fujita (2008)'s theory using patent data in Japan.

[7]Although the numbering of objects such as agents and periods starting with 0 is currently not common in economics, this numbering style is used in discrete-time simulation studies and programming especially using Python and thus we adopt this style.

[8]It is essentially the same whether joint creations in each period occur simultaneously or sequentially. The assumption of sequential joint trials in each period is adopted for the convenience of model description.

[9]In Berliant and Fujita (2008), an agent has the option of a solitary trial to create an idea. We do not consider this option in the model for simplicity.

where

$$f_{ij}^t = f(c_{ij}^t, d_{ij}^t, d_{ji}^t) = (c_{ij}^t)^q \cdot (d_{ij}^t \cdot d_{ji}^t)^{\frac{1-q}{2}}.$$

Because $f : \mathbb{Z}_+^3 \longrightarrow \mathbb{R}_+$ is borrowed from Berliant and Fujita (2008), we call this **the BF (Berliant–Fujita) function**. The BF function states that the number of i and j's common ideas, i's differentiated ideas from j's, and j's differentiated ideas from i's, all positively increase the outcome f_{ij}^t, which determines the success rate of idea creation via function p. $q \in (0,1)$ is the parameter of the BF function. If q is close to 1, the number of common ideas affects the success rate substantially. If q is close to 0, the numbers of differentiated ideas of both agents affect the success rate substantially. Note that if either c_{ij}^t, d_{ij}^t, or d_{ji}^t equals 0, then f_{ij}^t equals 0.

$p : \mathbb{R}_+ \longrightarrow (0,1)$ is the standard logistic function with f_{ij}^t as the variable. p monotonically converts f_{ij}^t into probability p_{ij}^t. ℓ, k, and m are the coefficients of the logistic function. $\ell \in (0,1]$ is the ceiling of p_{ij}^t, *i.e.*, $p_{ij}^t \in [0,\ell)$ holds and if $f_{ij}^t \to +\infty$, then $p_{ij}^t \to \ell$. $k \in (0,+\infty)$ is the steepness of the logistic curve. $m \in (0,+\infty)$ is the parameter such that if $f_{ij}^t = m$, then $p_{ij}^t = \ell/2$. This means that the success rate is at the midpoint when the value of the logistic function reaches m. We call this composition function $p \circ f : \mathbb{Z}_+^3 \longrightarrow (0,1)$ the **logistic BF function**.

Note that if either c_{ij}^t, d_{ij}^t, or d_{ji}^t equals zero, then $f_{ij}^t = 0$ and p_{ij}^t is the minimum. Thus, to have the success rate p_{ij}^t beyond the minimum, we need that all of c_{ij}^t, d_{ij}^t, and d_{ji}^t are larger or equal to 1. Table 2 provides some numerical examples of the logistic BF function. Figure 1 also presents illustrations of this function.

			$q = 0.1$		$q = 0.5$		$q = 0.9$	
c_{ij}^t	d_{ij}^t	d_{ji}^t	f_{ij}^t	p_{ij}^t	f_{ij}^t	p_{ij}^t	f_{ij}^t	p_{ij}^t
0	0	0	0.000	0.215	0.000	0.215	0.000	0.215
1	0	0	0.000	0.215	0.000	0.215	0.000	0.215
1	1	1	1.000	0.400	1.000	0.400	1.000	0.400
2	1	1	1.072	0.414	1.414	0.482	1.867	0.563
1	2	1	1.366	0.472	1.190	0.438	1.035	0.407
1	2	2	1.867	0.563	1.414	0.482	1.072	0.414
2	2	2	2.000	0.585	2.000	0.585	2.000	0.585

TABLE 2: Neumerical examples of the logistic BF function
Note. In Table 2, parameters of the logistic BF function are fixed as $\ell = 0.8$, $k = 1$, and $m = 1$.

If agents i and j's joint creation of a new idea is a success, then they both obtain the created idea. Assume that the previously created ideas before this joint creation in the run are $0, 1, \ldots, x-1$, then the newly created idea by agents i and j is named x. Then, at the beginning of the next period $t+1$, $y_i^{t+1} = y_i^t + 1$ and $y_j^{t+1} = y_j^t + 1$. If the joint creation is a failure, they receive nothing in this period. Then, at the beginning of the next period $t+1$, $y_i^{t+1} = y_i^t$ and $y_j^{t+1} = y_j^t$.

By abuse of notation, T represents not only the total number of periods but also the end of

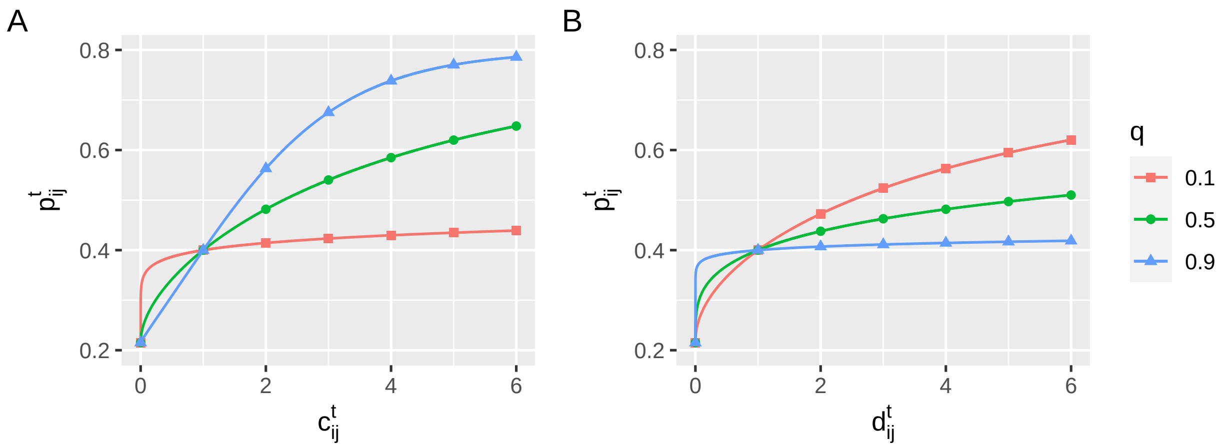


FIGURE 1: Illustrations of the logistic BF function
Note. The parameters of the logistic BF function are fixed as $\ell = 0.8$, $k = 1$, and $m = 1$. Panel A illustrates the relationship between c_{ij}^t and p_{ij}^t for each q when $d_{ij}^t = d_{ji}^t = 1$. Panel B illustrates the relationship between d_{ij}^t and p_{ij}^t for each q when $c_{ij}^t = d_{ji}^t = 1$.

the run immediately after the final period. Let $y_i = y_i^T$, which is the total number of ideas that agent i contributes to create and possesses at the end of the final period $T - 1$. We call this agent i's **individual productivity**. Let X denote the total number of created ideas in each run. At that time, the created ideas are $0, 1, \ldots, X - 1$. Because each created idea is owned by two agents, $\sum_{i=0}^{n-1} y_i = 2X$.

Parameter setup

The total number of periods in each run is fixed as $T = 20$.[10] The number of agents considered here is $n = 4, 6, \ldots, 38, 40$. Throughout this paper, the (numerical) parameters of the logistic BF functions are fixed as $\ell = 0.8$, $k = 1$, and $m = 1$. We select three values of the parameter of the BF function: $q = 0.1, 0.5$, and 0.9. $q = 0.1$ is the case in which the number of differentiated ideas is more important than the number of common ideas for creating a new idea. $q = 0.9$ is the opposite case to the previous one. $q = 0.5$ is the mid case. More precisely, when $q = 0.5$, agents i and j's common idea and a pair of their two differentiated ideas are equally weighted.

A **collection** is a set of runs with a fixed q that generate an adequate amount of data. In a collection, for each number of agents n, $\lfloor 100{,}000/n \rfloor$ runs are conducted, where $\lfloor 100{,}000/n \rfloor$ is the maximal integer that is smaller than or equal to $100{,}000/n$, *i.e.*, $\max\{z \in \mathbb{Z} : z \leq 100{,}000/n\}$. For example, 25,000 runs for $n = 4$, 16,666 runs for $n = 6$, and so on. Hence, for each q, we have $25{,}000 + 16{,}666 + \cdots + 2{,}500 = 129{,}881$ runs, which constitutes a collection.

The main objective of the paper is productivity. In each run, the number of productivity observations equals the number of agents, n. Therefore, for each n in a collection, the number of productivity observations is $\lfloor 100{,}000/n \rfloor n$, which is close to 100,000. Let y denote the average productivity for each n of almost 100,000 observations in a collection. We simply refer to y as

[10]Note that the parameter setting of ℓ, k, and m for generating meaningful results is highly relevant to that of T. The combination of parameters of ℓ, k, m, and T chosen in this paper is tractable and an interesting one for analysis.

productivity because this is the exact measure of the main objective[11]. The total number of observations in a collection is $\sum_{n=4,\dots,40} \lfloor 100{,}000/n \rfloor n = 1{,}899{,}862$. For reference purposes, we also calculate the (average) **overall productivity** of each collection with 1,899,862 observations.

Here, we note below the reason why we drop the case of $n = 2$ from the main analysis.

Fact 1. When $n = 2$, productivity y is at its lowest regardless of q.

Fact 1 is derived directly from the property of the BF function. When $n = 2$, the pair of agents 0 and 1 always conduct a joint project and possess the same ideas. Hence, one of them always has no differentiated idea, *i.e.*, $d^t_{10} = d^t_{01} = 0$ for any t. This ensures that $f^t_{01} = 0$ and p^t_{01} are minima for any t regardless of q. Simulating $n = 2$ with 50,000 runs, productivity is 4.311 with a standard deviation of 1.829[12]. This is far smaller than that in any case of $n = 4$ to 40. Fact 1 holds throughout this paper. By Fact 1, $n = 2$ is a particular case so we remove it from the main analysis.

Summary statistics and the results

A simulation using these three collections is conducted and the summary statistics are presented in Table 3 (a). In the table, by abuse of terminology, even though the relationship is not inverted-U shaped but increasing, decreasing, or plateau shaped, the point at which productivity is at a maximum is called the **peak**. Figure 2 A shows the relationship between the number of agents n and productivity y in each q. The 99%-confidence interval for each n of each q is calculated and given in the supplementary material, whereas it is omitted in Figure 2 for simplicity. The upper and lower limits of the interval are roughly $y \pm 0.03$.

The main finding in this subsection is summarized as follows.

Result 1. *In the RndBM, regardless of q, the number of agents n in the institution (i.e., institution size) has a negative effect on productivity y.*

The reason behind this result is simple. In this model, the success rate of joint creation does not increase *without the existence of at least one common knowledge* between the two agents. When the number of agents is small, it is easy for each pair to obtain a common idea. However, when the number of agents is large, it becomes quite difficult to obtain a common idea. For example, in each case of $n = 4$ and 40 at the collection with $q = 0.1$, there are 1,000,000 matches for idea creation by pairs of agents. The numbers of matches by agents i and j with more than one common idea (*i.e.*, $c^t_{ij} \geq 1$) are 460,594 when $n = 4$ and 50,613 when $n = 40$.

The RndBM is so simple that this result is not often observed in actual institutions. Result 1 can be considered as the benchmark. Many actual institutions possess instruments to provide members basic and common knowledge for collaboration such as education and on-the-job knowledge transmission systems, which are investigated in the following subsections.

[11] Productivity y is precisely a function $y(n)$ from the set of agents to $\mathbb{R}_+$.

[12] Note that in the case of $n = 2$, the parameter q does not affect the outcome.

	overall prod (std div)	peak prod (std div)	n
(a) RndBM			
$q = 0.1$	5.142 (2.500)	6.748 (3.434)	4
$q = 0.5$	4.989 (2.389)	6.535 (3.332)	4
$q = 0.9$	4.821 (2.255)	6.300 (3.193)	4
(b) RndEdM			
$q = 0.1$	9.547 (3.384)	9.623 (3.358)	30
$q = 0.5$	8.489 (3.204)	8.836 (3.660)	4
$q = 0.9$	7.158 (2.840)	8.472 (3.597)	4
(c-i) RndTmM			
$q = 0.1$	8.585 (2.834)	9.621 (2.776)	10
$q = 0.5$	8.137 (2.817)	9.256 (2.906)	8
$q = 0.9$	7.344 (2.763)	8.867 (2.886)	8
(c-ii) RndTmM *tm*			
$q = 0.1$	4.410 (1.830)	4.333 (1.767)	10
$q = 0.5$	4.407 (1.826)	4.199 (1.744)	8
$q = 0.9$	4.407 (1.828)	4.188 (1.751)	8

TABLE 3: Summary statistics for the random matching models
Note. "overall prod" and "peak prod" mean overall productivity and productivity at peak, respectively. "std div" and "n" mean the standard deviation and the number of agents that give the peak, respectively. (c-ii) shows the statistics for the number of transmitted ideas for each q in the RndTmM. The numbers of observations for *overall prod* and *peak prod* in each row are 1,899,862 and $\lfloor 100{,}000/n \rfloor n$ which is close to 100,000, respectively.

3.2 Model with education (RndEdM)

To avoid the shortage of common ideas in the RndBM, education is often employed in many institutions to provide common knowledge among agents. In this subsection, we add a simple preliminary education element to the basic model.

At the beginning of each run before the matching and joint creation process in the basic model (*i.e.*, before $t = 0$ in each run), all agents in the institution receive the same education. If an agent i succeeds in education, she can possess idea x^{ed} learned in education. If an agent i fails in education, she receives nothing. Let $p^{ed} \in [0, 1]$ denote the success rate of education, which is the same for all agents.

If agent i possesses x^{ed}, this works as a common or differentiated idea in the idea creation process, similarly to any ideas created by pairs of agents. However, x^{ed} is not counted for y_i^t, which is the number of ideas that agent i has contributed to creating at period t. If i possesses x^{ed}, $y_i^t + 1 = c_{ij}^t + d_{ij}^t$ holds for the joint creation with agent j. Otherwise, $y_i^t = c_{ij}^t + d_{ij}^t$ holds. All other structures of the RndEdM are the same as the RndBM.

Parameter setup and results

If p^{ed} is very low, then the result may be similar to that in the basic model. In this section, we use a high success rate and set $p^{ed} = 0.9$ as the parameter. All other parameters of the simulation are the same as in the basic model. The summary statistics of the simulation are given in Table 3 (b). Figure 2 B shows the relationship between the number of agents and

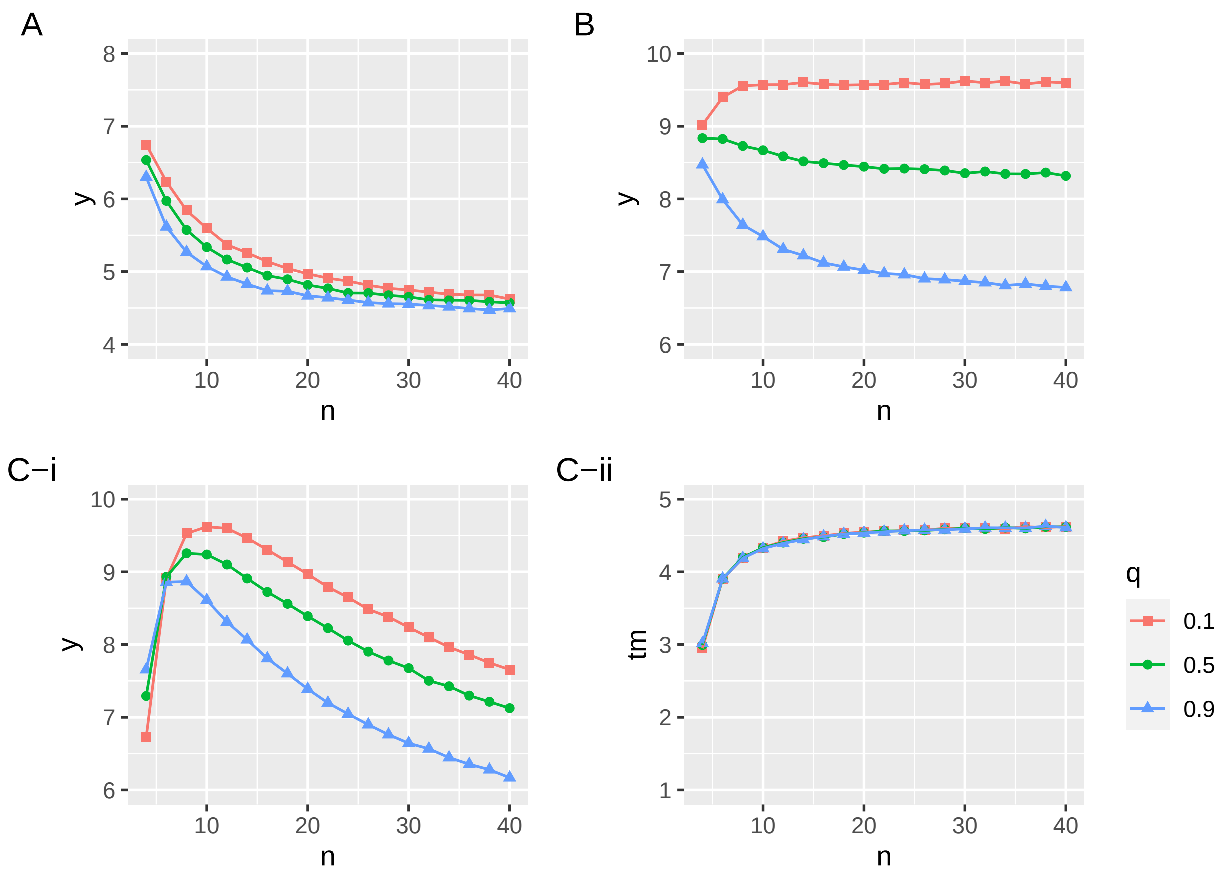


FIGURE 2: Results of the random matching models
Note: Panels A, B, and C-i depict the relationship between the number of agents n and productivity y in the RndBM, the RndEdM, and the RndTmM, respectively. Panel C-ii depicts the relationship between the number of agents n and the average number of transmitted ideas tm in the RndTmM.

productivity for each q of this model. The main result is summarized as follows.

Result 2. *In the RndEdM with high success rate (i.e., $p^{ed} = 0.9$), (i) if $q = 0.1$, the relationship is plateau shaped. As the number of agents n increases, productivity y increases until $n = 8$, and then becomes almost constant. (ii) If $q = 0.5$ or 0.9, n has a negative effect on productivity y, similarly to the RndBM.*

Unlike the RndBM, there is a common idea, which is learned in education by agents in most joint creations. Meanwhile, even in this model, if the number of agents increases, (1) it becomes easy to obtain differentiated ideas and (2) it becomes difficult to obtain an additional common idea that is similar to the basic model. In the case of $q = 0.1$, where the number of differentiated ideas affects the success rate substantially, property (1) is sufficiently effective to balance with property (2) when n is larger than 8. In the case of $q = 0.5$ or 0.9, where the number of additional common ideas is needed to increase the success rate, property (2) is more effective than property (1) and leads to Result 2 (ii).

An interpretation of Result 2 is as follows. Preliminary education works well in this model as expected. However, if education time is short and the number of ideas agents can obtain is restricted to one, it is not enough for a larger institution to be efficient especially when the common ideas are important for productivity. In such a case, additional preliminary education or on-the-job training may be effective.

3.3 Model with transmission of ideas (RndTmM)

A joint research project is sometimes a learning process for participants of each other's knowledge. In the words of current research, the transmission of a differentiated idea from one agent to another and vice versa happens frequently at each joint idea creation. In this subsection, we add this type of idea transmission process into the basic model. To examine the effect of idea transmission, we do not employ the preliminary education in the previous subsection here.

Consider a joint idea creation by a pair of agents i and j in each matching of period t, which is not the final period $T-1$. If there are j's differentiated ideas from i (*i.e.*, $d^t_{ji} \geq 1$), then a differentiated idea x is randomly chosen from them and agent i learns x from j in parallel with new idea creation. Let $p^{tm} \in [0, 1]$ be the success rate of this idea transmission, which is fixed throughout a collection. If the transmission is a success, then i receives x. At the same idea creation, if there are i's differentiated ideas from j (*i.e.*, $d^t_{ij} \geq 1$), then agent j learns an idea randomly chosen from them. Similarly to the opposite case, this idea transmission is a success with probability p^{tm}. If the transmission is a success, agent j receives this idea from agent i. Even though the number of agents who contribute to creating an idea is always two, there can be more than two agents who possess that idea through the creation and transmission.

Let tm^t_i denote the number of total ideas that agent i has obtained through the transmission processes and possesses at the beginning of period t. Obviously, for any agent i, $tm^0_i = 0$. At period t, if the transmission of i's differentiated idea from the other is a success, then $tm^{t+1}_i = tm^t_i + 1$. Otherwise, $tm^{t+1}_i = tm^t_i$. Then, the number of total ideas that agent i possesses at the beginning of period t is $y^t_i + tm^t_i$. We assume that both their created and learned ideas affect the success rate of joint creation equally. Hence, in agent i's joint creation with agent j at period t, $y^t_i + tm^t_i = c^t_{ij} + d^t_{ij}$ holds.

In joint creations at the final period $T-1$, we assume that there is no transmission of ideas between two agents. Remember that the objective of this study is productivity. Even if the transmission of ideas in the final period is possible, it does not affect productivity. Therefore, we drop off the transmission process from the final period. Let tm denote the average number of transmitted ideas for each n in a collection, which is an analog of productivity y. We also calculate the (average) overall number of transmitted ideas in each collection for reference. All other components of each run are the same as those in the RndBM.

Parameter setup and results

As the parameter, we select $p^{tm} = 0.3$ and all other parameters are the same as in the basic model. The summary statistics of the simulation are given in Table 3 (iii). Figure 2 C-i shows the relationship between the number of agents and productivity. Figure 2 C-ii shows the relationship between the number of agents and the average number of ideas obtained through the transmission process. Of course, Panel C-i is the graph of the main objective of this subsection, and Panel C-ii is for reference purposes. The main result is summarized as follows.

Result 3. *In the RndTmM with the transmission success rate* $p^{tm} = 0.3$*, regardless of* q*, the number of agents* n *and productivity* y *has an inverted-U shaped relationship. For* $q = 0.1$*,* 0.5*, and* 0.9*, productivity is at the peak when* n *is* 10*,* 8*, and* 8*, respectively.*

In this model, ideas that affect the success rate of joint creation are (1) those created by the conducted agents themselves and (2) those transmitted in the learning process. The effect of (1) on the result is the same as for the RndBM, in which the number of agents affects productivity negatively in any of q. The effect of (2) on the result is the opposite. Figure 2 C-ii suggests that the number of agents affects the number of transmitted ideas positively regardless of q, even though the slope becomes gentle when n is larger than 14. The transmitted ideas always affect productivity positively, regardless of whether they work as common or differentiated ideas. At the point of maximal productivity, the effects of (1) and (2) are balanced. Even without the preparatory education, common ideas among agents emerge at that time.

Result 3 suggests that if the transmission of ideas works well, there may be an appropriate institution size. The optimal size may be helpful in maintaining the important common ideas among the group of agents.

4 Ability-ordered matching models

In the previous section, we considered the benchmark random matching models. Here, we consider more realistic models and compare them with the previous benchmark models. In some actual institutions, there often arises a self-organized members' hierarchy based on their ability. In such a case, a member with "high ability" is tacitly authorized to choose her joint project partner. Then, she may prefer to match with another high-ability member to improve the success rate of a joint project. At the same time, a member with "low ability" cannot match with a high-ability agent and is forced to match with another low-ability one. In this section, we introduce such *ability-ordered* matching models. Similarly to the previous section, we consider the basic model (**AbiBM**), the model with education (**AbiEdM**), and the model with transmission of ideas (**AbiTmM**).

4.1 Basic model (AbiBM)

The modified part of the ability-ordered model from the random matching model relates to the matching procedure. An agent i's **ability** at each period t is defined simply as the number of ideas y_i^t that she possesses. At each period, matching happens in order of ability as follows. The agent with the highest ability in each period matches with the agent with the second-highest ability, the agent with the third-highest ability matches with the agent with the fourth-highest ability, and so on. If there are several agents with the same ability, then the order within them is randomized. Apart from the matching process of each period, both the other procedures of each run and the parameters of the variables are the same as for the RndBM.

	overall prod (std div)	peak prod (std div)	n
(a) AbiBM			
$q = 0.1$	6.028 (3.380)	6.463 (3.403)	10
$q = 0.5$	6.078 (3.560)	6.685 (3.671)	8
$q = 0.9$	6.062 (3.690)	6.901 (3.953)	8
(b) AbiEdM			
$q = 0.1$	9.678 (4.222)	10.312 (4.308)	40
$q = 0.5$	9.284 (4.103)	9.559 (4.088)	16
$q = 0.9$	8.530 (3.969)	9.046 (4.106)	12
(c-i) AbiTmM			
$q = 0.1$	8.066 (3.753)	8.683 (3.789)	22
$q = 0.5$	7.882 (3.699)	8.451 (3.756)	18
$q = 0.9$	7.417 (3.538)	8.033 (3.652)	16
(c-ii) AbiTmM *tm*			
$q = 0.1$	3.571 (2.242)	3.915 (2.187)	22
$q = 0.5$	3.571 (2.237)	3.774 (2.171)	18
$q = 0.9$	3.570 (2.233)	3.672 (2.168)	16

TABLE 4: Summary statistics for ability-ordered matching models
Note. The meanings of abbreviations are the same as those in Table 3. (c-ii) shows the statistics for the number of transmitted ideas for each q in the AbiTmM. The numbers of observations for *overall prod* and *prod peak* in each row are 1,899,862 and $\lfloor 100{,}000/n \rfloor n$ which is close to 100,000, respectively.

Let us remark on incentives in this model. Suppose that each agent is extremely myopic in the sense that she can neither observe the details of ideas that other agents possess nor remember her matched partners in the previous periods, and she can only glance at the numbers of the ideas that other agents possess. Then, she cannot distinguish other agents' characteristics except their abilities that affect the success rate of joint creation. If she is authorized to choose her partner from a set of agents, she may choose the one with the highest ability. In this sense, the ability-ordered matching is *incentive compatible*. This holds throughout Section 4. Note that because of her myopic view, an agent often cannot choose an agent with whom the success rate of collaboration is the highest. This well resembles actual situations in which people are not fully rational and quite often make mistakes.

Results

The summary statistics of the simulation are given in Table 4 (a). Figure 3 A shows the relationship between firm size and productivity. It is interesting that along with Result 3, an inverted-U shaped relationship once again exists.

Result 4. *In the basic model with ability-ordered matching, for any q, the number of agents n and productivity y have an inverted-U shaped relationship. For $q = 0.1$, 0.5, and 0.9, productivity y is at the peak when n is 10, 8, and 8, respectively.*

The mechanism behind Result 4 is as follows. If the number of agents is small, there is a situation in which two high-ability agents always match with each other, and therefore cannot

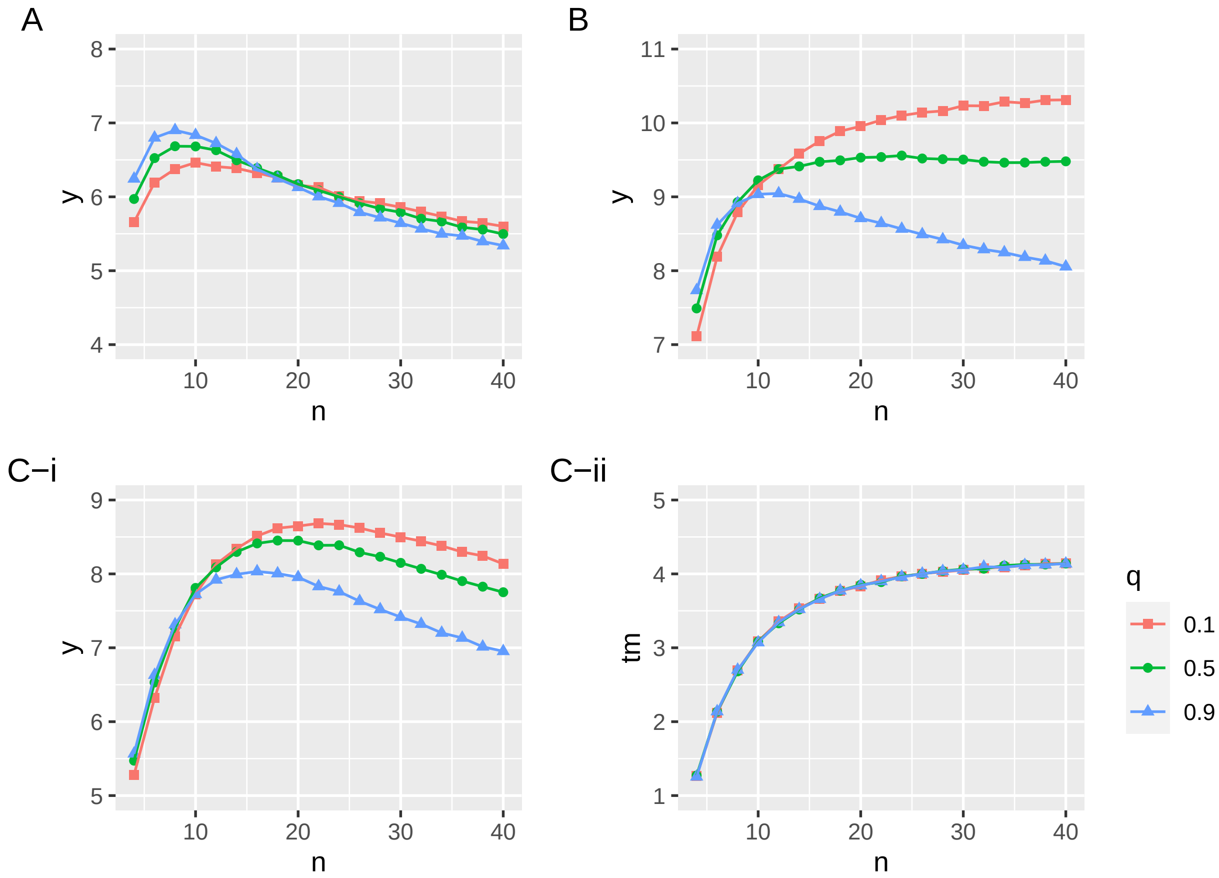


FIGURE 3: Results of the ability-ordered matching models
Note: Panels A, B, and C-i depict the relationship between the number of agents n and productivity y in the AbiBM, the AbiEdM, and the AbiTmM, respectively. Panel C-ii depicts the relationship between the number of agents n and the average number of transmitted ideas tm in the AbiTmM.

increase the number of their differentiated ideas or the success rate, and two low-ability agents always match with each other and continue to create an idea with a low success rate. If the number of agents is around 8, there are enough high-ability agents and the number of common and differentiated ideas are appropriately balanced to increase the success rate of joint creation. If the number of agents increases from 8, then the lack of common ideas, even among high-ability agents, becomes severe and productivity decreases, similarly to the RndBM.

There are several other interesting findings. When n is smaller than 16, productivity for $q = 0.9$ is larger than that for $q = 0.1$. This is because when n is small, matching among high-ability agents occurs frequently and the common ideas among them are generated. Furthermore, note that in any q, the overall productivity in this model is larger than that in the RndBM, even though in cases of $q = 0.1$, productivity at the peak is lower in this model than that in the RndBM. Institutions often introduce a device to force randomized matching of researchers to improve not only low-ability researchers' performance but also the average productivity of institutions. However, this finding suggests that the randomization is not meaningful for the latter purpose for many institution sizes. The variance of average productivity is larger in the

AbiBM than that in the RndBM. This finding may coincide with intuition. In the AbiBM, high-ability agents match with each other and increase their productivity on the one hand, however, low-ability agents remain at low productivity on the other. Hence, the disparity between high- and low-ability agents becomes large in the AbiBM.

4.2 Model with education (AbiEdM)

The AbiEdM discussed here is the RndEdM discussed in Subsection 3.2 with the replacement of random matching with ability-ordered matching. Apart from the matching process of each period, all the other procedures of each run and the parameters are the same as the RndEdM.

Results

The summary statistics of this model are given in Table 4 (b). Figure 3 B shows the relationship between firm size and productivity.

Result 5. *In the AbiEdM with a high success rate (i.e.,* $p^{ed} = 0.9$*), (i) if* $q = 0.1$*, the number of agents* n *in the institution has a positive effect on productivity* y*. (ii) If* $q = 0.5$*, the relationship is plateau shaped. As* n *increases,* y *goes up until* $n = 16$*, and then it becomes almost constant. (iii) If* $q = 0.9$*, the relationship is inverted-U shaped. Productivity is at its peak when* $n = 12$*.*

In the AbiBM, if n becomes large, the lack of common ideas among collaborating agents becomes a bottleneck to improving productivity. In the cases of $q = 0.1$ and 0.5, the common idea via education in the AbiEdM works well to avoid this bottleneck. By contrast, in the case of $q = 0.9$, because of the lack of additional common ideas, productivity is not improved when n becomes large. In any q, both overall and peak productivity at the peak are larger in the AbiEdM than those in the RndEdM, whereas both variances are also larger in the AbiEdM than those in the RndEdM.

4.3 Model with transmission of ideas (AbiTmM)

Here, we consider the AbiTmM, which is the RndTmM discussed in Subsection 3.3 with the replacement of the random matching by the ability-ordered matching. All the parameters in the AbiTmM are the same as those in the RndTmM.

Results

The summary statistics of the simulation are presented in Table 4 (c-i) and (c-ii). Figure 3 C shows the relationships between (i) the number of agents and productivity, and (ii) the number of agents and the average number of ideas that an agent obtained through the transmission process. The main result is summarized as follows.

Result 6. *In the AbiTmM with the transmission success rate* $p^{tm} = 0.3$*, for any* q*, the number of agents* n *and productivity* y *have an inverted-U shaped relationship. For* $q = 0.1$*, 0.5, and 0.9, productivity is at the peak when* n *is 22, 18, and 16, respectively.*

Remember that the AbiTmM is the AbiBM with the addition of idea transmission among agents. Result 4 suggests that even in the original AbiBM, institution size and productivity have an inverted-U shaped relationship. Figure 3 C-ii suggests that regardless of q, the number of agents affects the number of transmitted ideas positively, similarly to the RndTmM. Hence, by this additional effect of transmitted ideas, the number of agents that result in peak productivity is shifted to the right-hand side in the AbiTmM from that in the AbiBM. Contrary to the AbiBM and AbiEdM, productivity in the AbiTmM is smaller than that in the RndTmM. This may indicate that the randomization of the members is helpful for the improvement of productivity if on-the-job knowledge transmissions are observed frequently. The variance of productivity in the AbiTmM is larger than that in the RndTmM, which is similar to the AbiBM and AbiEdM.

5 Discussion and Concluding Remarks

This paper has studied how institution size affects productivity in the collaborative knowledge-creation process. Throughout the main analysis, the total number of periods is fixed as $T = 20$. Here, we briefly note the results of the simulation with an alternative parameter setting of $T = 10$[13]. We conducted a simulation with all six models, with the same model structures and parameter values as in the main analysis, but with a different total number of periods. The results are almost the same although the overdispersion of productivity is stronger in the case of $T = 10$. The only slight difference is that in the RndEdM with $q = 0.5$, the negative effect of institution size on productivity in Result 2 is not observed when $T = 10$. The summary statistics and the figures for $T = 10$ are given in the electronic supplementary material.

Finally, we remark on two possible extensions of this study. This paper is simple, particularly in two aspects. First, we consider the case of a single institution. Extending the model to include multiple institutions with their interactions may be important and interesting[14]. Second, this paper assumes that all ideas are symmetric. However, in real-world knowledge creation processes, some ideas are more influential than others. In other words, there is asymmetry in the quality of ideas. Developing models that incorporate such asymmetries of ideas could be an interesting direction for future research.

Acknowledgement. The author thanks Yukihisa Fujita, Naoki Watanabe, Daisuke Yoshikawa, and several seminar participants for helpful suggestions. Any remaining errors are the author's responsibility.
Funding. This study received no specific grant from any funding agency.
Conflict of Interest. The author declares that there is no conflict of interests.
Data Availability. Data for this study were generated using computational methods. The program codes for generating data are available in the electronic supplementary material.

[13] It may not be meaningful to consider the case in which T is much larger than 20 for the following two reasons. First, in an actual institution, repeated collaborative knowledge creation for many periods by the same members is rarely observed. Second, in the latter part of long periods, there are many joint creations with high success rates close to the ceiling $\ell = 0.8$ without differences in both models and the numbers of agents, which makes comparisons unclear.

[14] Berliant and Fujita (2012) extend their theory to a large-scale model with two regions, and show that under a certain condition, regional segregation with weak interaction helps to improve productivity.

References

Acs ZJ, Audretsch DB (1987) Innovation, market structure, and firm size. Review of Economics and Statistics 69 (4): 567–574 https://doi.org/10.2307/1935950

Andries P, Thorwarth S (2014) Should firms outsource their basic research? The impact of firm size on in-house versus outsourced R&D productivity. Creativity and Innovation Management 23 (3): 303-317

Axtell RL, Farmer JD (2025) Agent-based modeling in economics and finance: Past, present, and future. Journal of Economic Literature, forthcoming

Berliant M, Fujita M (2008) Knowledge creation as a square dance on the Hilbert cube. International Economic Review 49 (4): 1251–1295 https://doi.org/10.1111/j.1468-2354.2008.00512.x

Berliant M, Fujita M (2012) Culture and diversity in knowledge creation. Regional Science and Urban Economics 42 (4): 648–662 https://doi.org/10.1016/j.regsciurbeco.2012.02.008

Camisón-Zornoza C, Lapiedra-Alcamí R, Segarra-Ciprés M, Boronat-Navarro M (2004) A meta-analysis of innovation and organizational size. Organization Studies 25 (3): 331–361 https://doi.org/10.1177/0170840604040039

Caminati M (2016) Knowledge specialization and R&D collaboration. Journal of Evolutionary Economics 26 (2): 247–270 https://doi.org/10.1007/s00191-016-0449-5

Cohen WM, Klepper S (1996) A reprise of size and R&D. Economic Journal 106 (437): 925–951 https://doi.org/10.2307/2235365

Cowan R, Jonard N (2004) Network structure and the diffusion of knowledge. Journal of Economic Dynamics and Control 28 (8): 1557–1575 https://doi.org/10.1016/j.jedc.2003.04.002

Czarnitzki D, Hottenrott H (2011) R&D investment and financing constraints of small and medium-sized firms. Small Business Economics 36 (1): 65–83 https://doi.org/10.1007/s11187-009-9189-3

Damanpour F (1992) Organizational size and innovation. Organization Studies 13 (3): 375–402 https://doi.org/10.1177/017084069201300304

Delre SA, Jager W, Janssen MA (2007) Diffusion dynamics in small-world networks with heterogeneous consumers. Computational and Mathematical Organization Theory 13 (2): 185–202 https://doi.org/10.1007/s10588-006-9007-2

Gräbner C (2016) Agent-based computational models: A formal heuristic for institutionalist pattern modelling? Journal of Institutional Economics, 12 (1): 241–261 https://doi.org/10.1017/S1744137415000193

Horwitz SK, Horwitz IB (2007) The effects of team diversity on team outcomes: A meta-analytic review of team demography. Journal of Management 33 (6): 987–1015 https://doi.org/10.1177/0149206307308587

Hsiao Y-C, Hsu Z-X (2018) Firm-specific advantages-product innovation capability complementarities and innovation success: A core competency approach. Technology in Society 55: 78–84 https://doi.org/10.1016/j.techsoc.2018.06.009

Hu D, She M, Ye L, Wang Z (2021) The more the merrier? Inventor teamsize, diversity, and innovation quality. Science and Public Policy 48 (4): 508–520 https://doi.org/10.1093/scipol/scab033

Huo D, Motohashi K, Gong H (2019) Team diversity as dissimilarity and variety in organizational innovation. Research Policy 48 (6): 1564–1572 https://doi.org/10.1016/j.respol.2019.03.020

König MD, Battiston S, Napoletano M, Schweitzer F (2011) Recombinant knowledge and the evolution of innovation networks. Journal of Economic Behavior and Organization 79 (3): 145–164 https://doi.org/10.1016/j.jebo.2011.01.007

Mori T, Sakaguchi S (2019) Creation of knowledge through exchanges of knowledge: Evidence from Japanese patent data. arXiv 1908.01256 https://doi.org/10.48550/arXiv.1908.01256

Mote J, Jordan G, Hage J, Hadden W, Clark A (2016) Too big to innovate? Exploring organizational size and innovation processes in scientific research. Science and Public Policy 43 (3): 332–337 https://doi.org/10.1093/scipol/scv045

Müller M, Kudic M, Vermeulen B (2021) The influence of the structure of technological knowledge on inter-firm R&D collaboration and knowledge discovery: An agent-based simulation approach. Journal of Business Research 129: 570–579 https://doi.org/10.1016/j.jbusres.2020.01.067

Nonaka I, Takeuchi H (1995) *The Knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press

Nooteboom B (2000) *Learning and innovation in organizations and economies*. Oxford University Press

Nooteboom B, Vanhaverbeke W, Duysters G, Gilsing V, van den Oord A (2007) Optimal cognitive distance and absorptive capacity. Research Policy 36 (7): 1016–1034 https://doi.org/10.1016/j.respol.2007.04.003

Østergaard CR, Timmermans B, Kristinsson K (2011) Does a different view create something new? The effect of employee diversity on innovation. Research Policy 40 (3): 500–509 https://doi.org/10.1016/j.respol.2010.11.004

Ozman M, Parker A (2023) The effect of social networks, organizational coordination structures, and knowledge heterogeneity on knowledge transfer and aggregation. Journal of Evolutionary Economics 33 (2): 249–278 https://doi.org/10.1007/s00191-023-00811-z

Qiao T, Shan W, Zhang M, Liu C (2019) How to facilitate knowledge diffusion in complex networks: The roles of network structure, knowledge role distribution and selection rule. International Journal of Information Management 47: 152–167 https://doi.org/10.1016/j.ijinfomgt.2019.01.016

Shefer D, Frenkel A (2005) R&D, firm size and innovation: an empirical analysis. Technovation 25 (1): 25–32 https://doi.org/10.1016/S0166-4972(03)00152-4

Teirlinck P (2017) Configurations of strategic R&D decisions and financial performance in small-sized and medium-sized firms. Journal of Business Research 74: 55–65 https://doi.org/10.1016/j.jbusres.2017.01.008

Tur EM, Azagra-Caro JM (2018) The coevolution of endogenous knowledge networks and knowledge creation. Journal of Economic Behavior and Organization 145: 424–434 https://doi.org/10.1016/j.jebo.2017.11.023

van den Bergh J (2008) Optimal diversity: Increasing returns versus recombinant innovation. Journal of Economic Behavior and Organization 68 (3-4), 565–580 https://doi.org/10.1016/j.jebo.2008.09.003

Wall F (2016) Agent-based modeling in managerial science: An illustrative survey and study. Review of Managerial Science: 10 (1): 135–193 https://doi.org/10.1007/s11846-014-0139-3

Wang H, Chen W-R (2010) Is firm-specific innovation associated with greater value appropriation? The roles of environmental dynamism and technological diversity. Research Policy 39 (1): 141–154 http://dx.doi.org/10.1016/j.respol.2009.09.015

Wu L., Wang D, Evans JA (2019) Large teams develop and small teams disrupt science and technology. Nature 566: 378–382 https://doi.org/10.1038/s41586-019-0941-9

Electronic supplementary material for "Collaborative knowledge creation and size of institutions"

Author. Kentaro Hatsumi
Address. Faculty of Policy Studies, Kansai University, Suita, Osaka 564-8680, Japan
Email. hatsumi@kansai-u.ac.jp
This Version. February 21, 2025

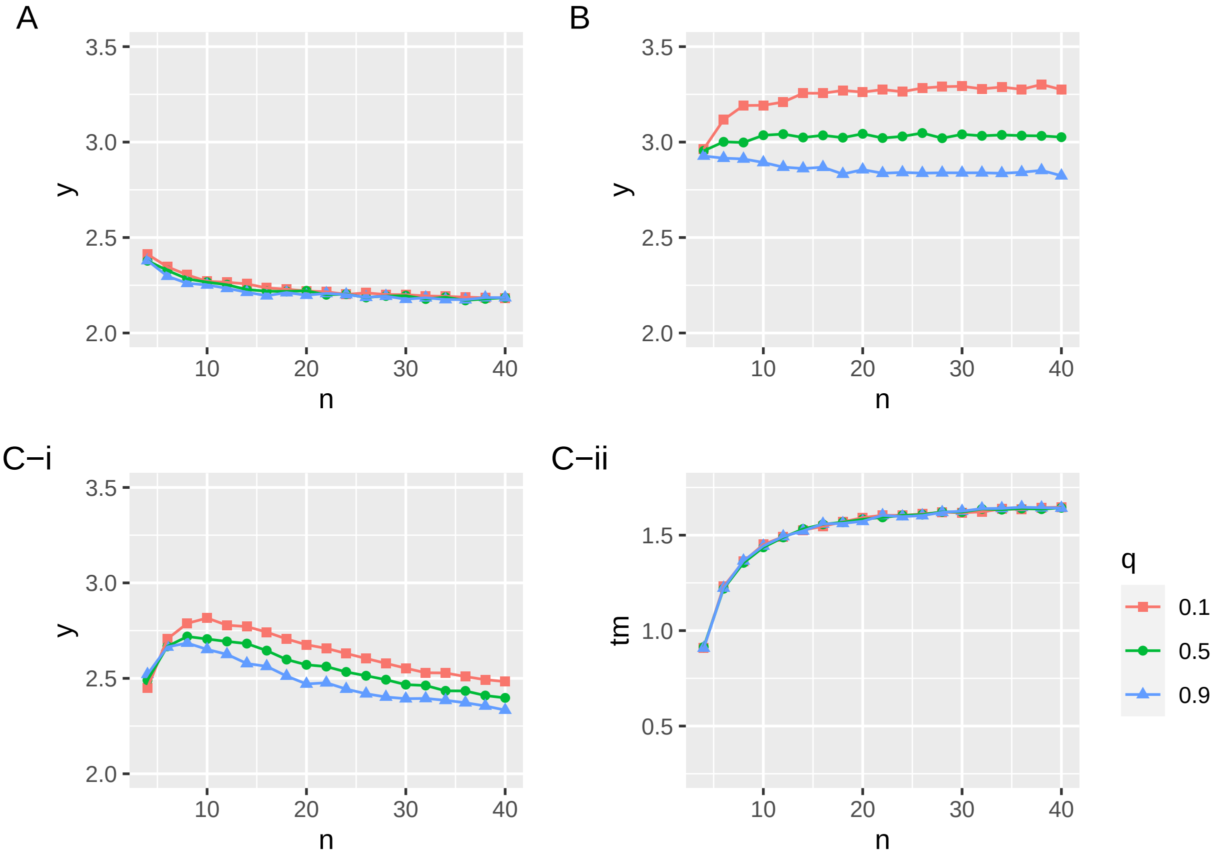
The present electronic supplementary material provides (1) summary tables and figures for an alternative parameter setting $T = 10$ in pages 2–5, (2) tables that contain detailed statistics omitted in the main paper in pages 6–13, and (3) the program codes for simulation in pages 14–27. The codes for the RndBM, the RndEdM, the RndTmM, and the AbiBM are available here. The codes for the AbiEdM and the AbiTmM are omitted because they are combinations of available codes. The author confirms that the codes can run at Google Colaboratory using Python 3.10.12 and NumPy 1.26.4 as of November 2024.

Table S1. Summary statistics for the random matching models ($T = 10$ case)

	overall prod (std div)	peak prod (std div)	n
(a) RndBM			
$q = 0.1$	2.238 (1.405)	2.412 (1.600)	4
$q = 0.5$	2.225 (1.389)	2.379 (1.583)	4
$q = 0.9$	2.216 (1.376)	2.379 (1.572)	4
(b) RndEdM			
$q = 0.1$	3.239 (1.949)	3.302 (1.949)	38
$q = 0.5$	3.025 (1.826)	3.047 (1.803)	26
$q = 0.9$	2.859 (1.714)	2.927 (1.912)	4
(c-i) RndTmM			
$q = 0.1$	2.632 (1.590)	2.816 (1.702)	10
$q = 0.5$	2.552 (1.548)	2.719 (1.676)	8
$q = 0.9$	2.487 (1.510)	2.685 (1.654)	8
(c-ii) RndTmM *tm*			
$q = 0.1$	1.522 (1.089)	1.451 (1.051)	10
$q = 0.5$	1.521 (1.088)	1.616 (1.099)	8
$q = 0.9$	1.522 (1.089)	1.623 (1.102)	8

Note. "overall prod" and "peak prod" mean overall productivity and productivity at peak, respectively. "std div" and "n" mean the standard deviation and the number of agents that give the peak, respectively. (c-ii) shows the statistics for the number of transmitted ideas for each q in the RndTmM. The numbers of observations for *overall prod* and *peak prod* in each row are 1,899,862 and $\lfloor 100{,}000/n \rfloor n$ which is close to 100,000, respectively.

Figure S1. Results of the random matching models ($T = 10$ case)



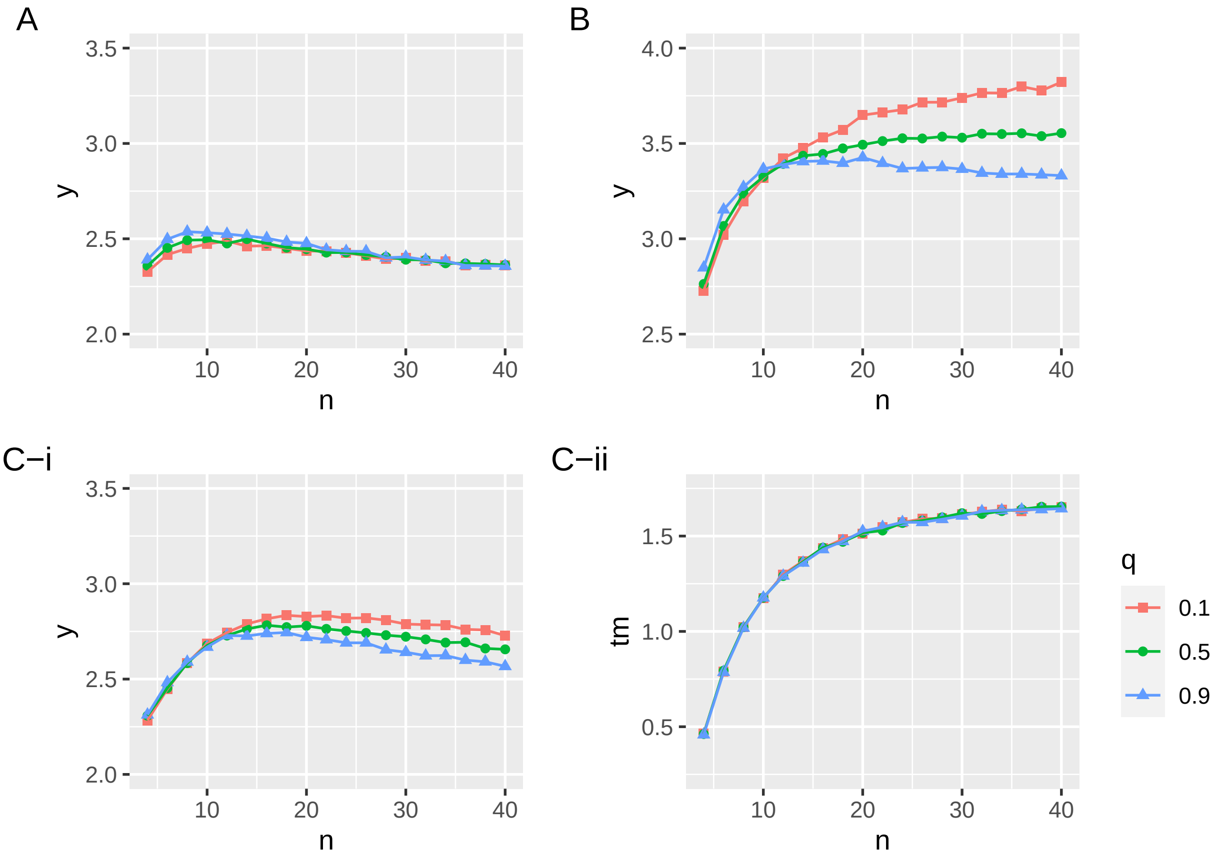
Note: Panels A, B, and C-i depict the relationship between the number of agent n and the mean productivity y in the RndBM, the RndEdM, and the RndTmM, respectively. Panel C-ii depicts the relationship between the number of agent n and the mean number of transmitted ideas tm in the RndTmM.

Table S2. Summary statistics for the ability-ordered matching models ($T = 10$ case)

	overall prod (std div)	peak prod (std div)	n
(a) AbiBM			
$q = 0.1$	2.415 (1.673)	2.488 (1.709)	12
$q = 0.5$	2.425 (1.704)	2.498 (1.742)	14
$q = 0.9$	2.443 (1.756)	2.537 (1.823)	8
(b) AbiEdM			
$q = 0.1$	3.545 (2.285)	3.822 (2.438)	40
$q = 0.5$	3.422 (2.212)	3.554 (2.253)	40
$q = 0.9$	3.328 (2.181)	3.426 (2.219)	20
(c-i) AbiTmM			
$q = 0.1$	2.731 (1.833)	2.834 (1.894)	18
$q = 0.5$	2.682 (1.803)	2.782 (1.860)	16
$q = 0.9$	2.637 (1.762)	2.744 (1.833)	18
(c-ii) AbiTmM *tm*			
$q = 0.1$	1.177 (1.192)	1.242 (1.188)	18
$q = 0.5$	1.176 (1.191)	1.203 (1.167)	16
$q = 0.9$	1.176 (1.189)	1.241 (1.185)	18

Note. The meanings of abbreviations are the same as those in Table S1. (c-ii) shows the statistics for the number of transmitted ideas for each q in the AbiTmM. The numbers of observations for *overall prod* and *prod peak* in each row are 1,899,862 and $\lfloor 100{,}000/n \rfloor n$ which is close to 100,000, respectively.

Figure S2. Results of the ability-ordered matching models ($T = 10$ case)



Note: Panels A, B, and C-i depict the relationship between the number of agent n and the mean productivity y in the AbiBM, the AbiEdM, and the AbiTmM, respectively. Panel C-ii depicts the relationship between the number of agent n and the mean number of transmitted ideas tm in the AbiTmM.

Table S3 A. Detailed results of the RndBM that correspond to Result 1 and Fig. 2 A

q=0.1

n	y	min	max	sd	lb	ub	obs
4	6.748	0	20	3.434	6.720	6.776	100000
6	6.238	0	19	3.079	6.212	6.263	99996
8	5.842	0	18	2.840	5.819	5.865	100000
10	5.597	0	18	2.695	5.575	5.619	100000
12	5.370	0	18	2.558	5.349	5.391	99996
14	5.259	0	16	2.481	5.239	5.279	99988
16	5.138	0	17	2.403	5.119	5.158	100000
18	5.043	0	16	2.347	5.024	5.062	99990
20	4.971	0	16	2.304	4.952	4.989	100000
22	4.910	0	17	2.272	4.891	4.928	99990
24	4.871	0	16	2.236	4.853	4.889	99984
26	4.813	0	15	2.202	4.795	4.831	99996
28	4.771	0	15	2.180	4.753	4.788	99988
30	4.749	0	15	2.157	4.731	4.766	99990
32	4.718	0	15	2.132	4.701	4.736	100000
34	4.690	0	15	2.125	4.673	4.707	99994
36	4.680	0	16	2.115	4.662	4.697	99972
38	4.679	0	15	2.095	4.662	4.696	99978
40	4.622	0	15	2.078	4.605	4.639	100000

q=0.5

n	y	min	max	sd	lb	ub	obs
4	6.535	0	20	3.332	6.508	6.562	100000
6	5.974	0	18	2.926	5.950	5.998	99996
8	5.573	0	17	2.690	5.551	5.595	100000
10	5.337	0	18	2.542	5.316	5.357	100000
12	5.167	0	16	2.436	5.147	5.187	99996
14	5.056	0	18	2.346	5.037	5.075	99988
16	4.945	0	15	2.302	4.927	4.964	100000
18	4.895	0	16	2.242	4.876	4.913	99990
20	4.816	0	15	2.195	4.798	4.834	100000
22	4.769	0	16	2.170	4.752	4.787	99990
24	4.706	0	15	2.135	4.689	4.724	99984
26	4.705	0	15	2.122	4.688	4.722	99996
28	4.675	0	14	2.097	4.657	4.692	99988
30	4.653	0	15	2.083	4.636	4.670	99990
32	4.612	0	15	2.066	4.595	4.629	100000
34	4.608	0	15	2.056	4.591	4.625	99994
36	4.605	0	14	2.040	4.589	4.622	99972
38	4.585	0	16	2.030	4.569	4.602	99978
40	4.570	0	15	2.022	4.554	4.587	100000

q=0.9

n	y	min	max	sd	lb	ub	obs
4	6.300	0	19	3.193	6.274	6.326	100000
6	5.617	0	17	2.718	5.595	5.639	99996
8	5.268	0	16	2.481	5.248	5.288	100000
10	5.070	0	16	2.352	5.051	5.090	100000
12	4.929	0	15	2.268	4.910	4.947	99996
14	4.829	0	15	2.190	4.811	4.847	99988
16	4.738	0	15	2.141	4.721	4.756	100000
18	4.730	0	15	2.125	4.713	4.747	99990
20	4.668	0	14	2.081	4.651	4.685	100000
22	4.642	0	15	2.060	4.625	4.658	99990
24	4.609	0	15	2.044	4.592	4.626	99984
26	4.575	0	15	2.027	4.558	4.591	99996
28	4.559	0	14	2.012	4.543	4.575	99988
30	4.555	0	14	2.004	4.538	4.571	99990
32	4.534	0	15	1.994	4.518	4.550	100000
34	4.517	0	14	1.988	4.501	4.534	99994
36	4.493	0	16	1.971	4.477	4.509	99972
38	4.474	0	14	1.964	4.458	4.490	99978
40	4.493	0	15	1.955	4.477	4.509	100000

Note. **n** is the number of agents. **y** is the (mean) productivity for each *n*. **min** and **max** are the minimum and maximum of individual productivity for each *n*, respectively. **sd** is the standard deviation of productivity. **lb** and **ub** are the lower and upper bounds of the 99%-confidence interval for *y*, respectively. **obs** is the number of observations.

Table S3 B. Detailed results of the RndEdM that correspond to Result 2 and Fig. 2 B

q=0.1

n	y	min	max	sd	lb	ub	obs
4	9.019	0	20	3.633	8.990	9.049	100000
6	9.398	0	20	3.453	9.370	9.426	99996
8	9.556	0	20	3.425	9.528	9.584	100000
10	9.571	0	20	3.385	9.543	9.599	100000
12	9.572	0	20	3.387	9.544	9.600	99996
14	9.603	0	20	3.361	9.575	9.630	99988
16	9.578	0	19	3.374	9.550	9.605	100000
18	9.566	0	20	3.356	9.538	9.593	99990
20	9.571	0	20	3.356	9.544	9.598	100000
22	9.574	0	19	3.359	9.546	9.601	99990
24	9.602	0	20	3.355	9.574	9.629	99984
26	9.575	0	20	3.352	9.548	9.602	99996
28	9.588	0	20	3.344	9.561	9.615	99988
30	9.623	0	19	3.358	9.596	9.651	99990
32	9.598	0	20	3.370	9.571	9.625	100000
34	9.620	0	20	3.343	9.593	9.647	99994
36	9.581	0	20	3.354	9.554	9.609	99972
38	9.611	0	19	3.336	9.583	9.638	99978
40	9.596	0	20	3.334	9.569	9.623	100000

q=0.5

n	y	min	max	sd	lb	ub	obs
4	8.836	0	20	3.660	8.806	8.866	100000
6	8.825	0	20	3.401	8.797	8.853	99996
8	8.729	0	20	3.323	8.702	8.756	100000
10	8.669	0	20	3.258	8.642	8.695	100000
12	8.586	0	19	3.222	8.560	8.612	99996
14	8.516	0	19	3.207	8.490	8.543	99988
16	8.492	0	19	3.176	8.466	8.517	100000
18	8.467	0	19	3.173	8.441	8.493	99990
20	8.444	0	20	3.167	8.418	8.470	100000
22	8.414	0	19	3.148	8.389	8.440	99990
24	8.418	0	19	3.140	8.392	8.443	99984
26	8.409	0	19	3.114	8.384	8.435	99996
28	8.391	0	19	3.120	8.366	8.417	99988
30	8.354	0	20	3.113	8.328	8.379	99990
32	8.377	0	20	3.110	8.352	8.403	100000
34	8.345	0	19	3.106	8.319	8.370	99994
36	8.344	0	19	3.103	8.318	8.369	99972
38	8.363	0	19	3.104	8.338	8.388	99978
40	8.317	0	18	3.104	8.292	8.342	100000

q=0.9

n	y	min	max	sd	lb	ub	obs
4	8.472	0	20	3.597	8.442	8.501	100000
6	7.992	0	19	3.223	7.966	8.019	99996
8	7.643	0	18	3.055	7.618	7.668	100000
10	7.483	0	18	2.945	7.459	7.507	100000
12	7.308	0	19	2.859	7.284	7.331	99996
14	7.221	0	19	2.812	7.198	7.244	99988
16	7.121	0	17	2.768	7.098	7.144	100000
18	7.066	0	17	2.740	7.044	7.089	99990
20	7.020	0	19	2.713	6.998	7.042	100000
22	6.975	0	18	2.697	6.953	6.997	99990
24	6.959	0	18	2.661	6.938	6.981	99984
26	6.903	0	17	2.663	6.882	6.925	99996
28	6.891	0	18	2.660	6.869	6.913	99988
30	6.869	0	18	2.635	6.848	6.890	99990
32	6.849	0	17	2.626	6.828	6.871	100000
34	6.810	0	17	2.623	6.789	6.831	99994
36	6.831	0	18	2.606	6.809	6.852	99972
38	6.801	0	17	2.603	6.780	6.822	99978
40	6.782	0	17	2.601	6.761	6.804	100000

Note. **n** is the number of agents. **y** is the (mean) productivity for each *n*. **min** and **max** are the minimum and maximum of individual productivity for each *n*, respectively. **sd** is the standard deviation of productivity. **lb** and **ub** are the lower and upper bounds of the 99%-confidence interval for *y*, respectively. **obs** is the number of observations.

Table S3 C-i. Detailed results of the RndTmM that correspond to Result 3 and Fig. 2 C-i

q=0.1

n	y	min	max	sd	lb	ub	obs
4	6.724	0	20	3.389	6.697	6.752	100000
6	8.909	0	20	3.228	8.883	8.935	99996
8	9.528	0	20	2.914	9.504	9.552	100000
10	9.621	0	19	2.776	9.598	9.643	100000
12	9.598	0	20	2.713	9.576	9.620	99996
14	9.465	0	19	2.689	9.443	9.487	99988
16	9.303	0	18	2.678	9.282	9.325	100000
18	9.137	0	19	2.671	9.115	9.159	99990
20	8.968	0	18	2.651	8.947	8.990	100000
22	8.791	0	19	2.641	8.770	8.813	99990
24	8.648	0	18	2.650	8.627	8.670	99984
26	8.484	0	19	2.628	8.463	8.506	99996
28	8.381	0	19	2.607	8.360	8.402	99988
30	8.236	0	18	2.609	8.215	8.258	99990
32	8.099	0	19	2.588	8.078	8.120	100000
34	7.964	0	18	2.576	7.943	7.985	99994
36	7.860	0	18	2.560	7.839	7.881	99972
38	7.749	0	17	2.551	7.728	7.770	99978
40	7.655	0	19	2.537	7.635	7.676	100000

q=0.5

n	y	min	max	sd	lb	ub	obs
4	7.293	0	20	3.566	7.264	7.322	100000
6	8.932	0	19	3.143	8.907	8.958	99996
8	9.256	0	20	2.906	9.232	9.279	100000
10	9.238	0	20	2.786	9.216	9.261	100000
12	9.101	0	18	2.750	9.079	9.123	99996
14	8.909	0	19	2.705	8.887	8.931	99988
16	8.724	0	19	2.692	8.702	8.746	100000
18	8.561	0	19	2.676	8.539	8.582	99990
20	8.390	0	18	2.657	8.368	8.412	100000
22	8.226	0	19	2.635	8.205	8.247	99990
24	8.055	0	19	2.623	8.033	8.076	99984
26	7.903	0	19	2.600	7.882	7.924	99996
28	7.780	0	18	2.592	7.759	7.801	99988
30	7.675	0	18	2.573	7.654	7.696	99990
32	7.502	0	18	2.554	7.482	7.523	100000
34	7.426	0	18	2.537	7.405	7.447	99994
36	7.298	0	17	2.521	7.277	7.318	99972
38	7.213	0	18	2.517	7.193	7.234	99978
40	7.125	0	19	2.497	7.104	7.145	100000

q=0.9

n	y	min	max	sd	lb	ub	obs
4	7.656	0	20	3.640	7.626	7.685	100000
6	8.858	0	19	3.096	8.833	8.883	99996
8	8.867	0	19	2.886	8.843	8.890	100000
10	8.611	0	19	2.773	8.588	8.634	100000
12	8.312	0	18	2.709	8.290	8.334	99996
14	8.067	0	18	2.650	8.045	8.089	99988
16	7.809	0	18	2.629	7.787	7.830	100000
18	7.601	0	17	2.584	7.580	7.622	99990
20	7.389	0	17	2.542	7.368	7.410	100000
22	7.198	0	17	2.506	7.178	7.219	99990
24	7.045	0	18	2.486	7.025	7.066	99984
26	6.898	0	18	2.454	6.878	6.918	99996
28	6.762	0	17	2.433	6.742	6.781	99988
30	6.643	0	17	2.407	6.623	6.663	99990
32	6.566	0	16	2.385	6.547	6.585	100000
34	6.445	0	16	2.359	6.426	6.465	99994
36	6.354	0	16	2.346	6.335	6.373	99972
38	6.279	0	16	2.317	6.260	6.298	99978
40	6.169	0	17	2.306	6.150	6.188	100000

Note. **n** is the number of agents. **y** is the (mean) productivity for each *n*. **min** and **max** are the minimum and maximum of individual productivity for each *n*, respectively. **sd** is the standard deviation of productivity. **lb** and **ub** are the lower and upper bounds of the 99%-confidence interval for *y*, respectively. **obs** is the number of observations.

Table S3 C-ii. Detailed supporting results of the RndTmM that correspond to Fig. 2 C-ii

q=0.1

n	tm	min	max	sd	lb	ub	obs
4	2.949	0	11	1.587	2.936	2.962	100000
6	3.904	0	13	1.715	3.890	3.918	99996
8	4.190	0	12	1.742	4.175	4.204	100000
10	4.333	0	13	1.767	4.319	4.347	100000
12	4.420	0	14	1.779	4.406	4.435	99996
14	4.469	0	12	1.789	4.454	4.484	99988
16	4.496	0	13	1.799	4.481	4.511	100000
18	4.528	0	13	1.801	4.513	4.543	99990
20	4.549	0	13	1.798	4.534	4.564	100000
22	4.556	0	14	1.802	4.541	4.570	99990
24	4.569	0	13	1.808	4.554	4.584	99984
26	4.574	0	13	1.814	4.560	4.589	99996
28	4.599	0	13	1.817	4.584	4.614	99988
30	4.603	0	13	1.830	4.588	4.617	99990
32	4.601	0	13	1.823	4.586	4.616	100000
34	4.594	0	14	1.821	4.579	4.609	99994
36	4.621	0	13	1.826	4.606	4.635	99972
38	4.611	0	14	1.830	4.596	4.626	99978
40	4.624	0	14	1.829	4.609	4.639	100000

q=0.5

n	tm	min	max	sd	lb	ub	obs
4	2.994	0	11	1.618	2.981	3.007	100000
6	3.906	0	12	1.705	3.893	3.920	99996
8	4.199	0	12	1.744	4.184	4.213	100000
10	4.333	0	13	1.756	4.319	4.347	100000
12	4.403	0	13	1.773	4.388	4.417	99996
14	4.455	0	13	1.788	4.441	4.470	99988
16	4.478	0	12	1.788	4.464	4.493	100000
18	4.521	0	14	1.801	4.506	4.536	99990
20	4.539	0	14	1.809	4.525	4.554	100000
22	4.563	0	13	1.805	4.548	4.578	99990
24	4.562	0	12	1.809	4.548	4.577	99984
26	4.570	0	13	1.810	4.555	4.585	99996
28	4.584	0	13	1.812	4.570	4.599	99988
30	4.600	0	13	1.818	4.585	4.615	99990
32	4.590	0	14	1.810	4.575	4.605	100000
34	4.609	0	13	1.828	4.594	4.624	99994
36	4.598	0	13	1.822	4.583	4.612	99972
38	4.619	0	14	1.822	4.604	4.634	99978
40	4.618	0	13	1.818	4.603	4.632	100000

q=0.9

n	tm	min	max	sd	lb	ub	obs
4	3.016	0	11	1.640	3.003	3.030	100000
6	3.906	0	12	1.710	3.892	3.920	99996
8	4.188	0	13	1.751	4.174	4.203	100000
10	4.321	0	12	1.756	4.307	4.335	100000
12	4.393	0	13	1.785	4.378	4.407	99996
14	4.446	0	13	1.786	4.431	4.460	99988
16	4.486	0	12	1.794	4.471	4.500	100000
18	4.520	0	13	1.796	4.506	4.535	99990
20	4.536	0	13	1.801	4.522	4.551	100000
22	4.551	0	14	1.798	4.536	4.566	99990
24	4.568	0	13	1.809	4.553	4.583	99984
26	4.576	0	13	1.819	4.561	4.591	99996
28	4.579	0	14	1.816	4.564	4.593	99988
30	4.592	0	13	1.816	4.577	4.607	99990
32	4.606	0	13	1.821	4.592	4.621	100000
34	4.605	0	12	1.821	4.590	4.620	99994
36	4.603	0	14	1.824	4.588	4.618	99972
38	4.629	0	15	1.828	4.615	4.644	99978
40	4.609	0	13	1.822	4.595	4.624	100000

Note. **n** is the number of agents. **tm** is the mean number of transmitted ideas for each *n*. **min** and **max** are the minimum and maximum numbers of transmitted ideas for each *n*, respectively. **sd** is the standard deviation of productivity. **lb** and **ub** are the lower and upper bounds of the 99% confidence interval for *y*, respectively. **obs** is the number of observations.

Table S4 A. Detailed results of the AbiBM that correspond to Result 4 and Fig. 3 A

q=0.1

n	y	min	max	sd	lb	ub	obs
4	5.657	0	18	2.980	5.633	5.681	100000
6	6.194	0	18	3.225	6.168	6.221	99996
8	6.376	0	19	3.357	6.349	6.403	100000
10	6.463	0	19	3.403	6.435	6.491	100000
12	6.410	0	19	3.439	6.381	6.438	99996
14	6.387	0	19	3.460	6.359	6.415	99988
16	6.323	0	19	3.457	6.295	6.351	100000
18	6.259	0	19	3.466	6.230	6.287	99990
20	6.158	0	18	3.465	6.130	6.186	100000
22	6.130	0	20	3.466	6.101	6.158	99990
24	6.013	0	20	3.424	5.985	6.041	99984
26	5.943	0	18	3.407	5.916	5.971	99996
28	5.912	0	19	3.398	5.884	5.940	99988
30	5.860	0	20	3.385	5.832	5.887	99990
32	5.798	0	19	3.379	5.770	5.826	100000
34	5.736	0	19	3.328	5.708	5.763	99994
36	5.669	0	19	3.314	5.642	5.696	99972
38	5.647	0	20	3.304	5.620	5.674	99978
40	5.600	0	19	3.291	5.573	5.627	100000

q=0.5

n	y	min	max	sd	lb	ub	obs
4	5.971	0	19	3.341	5.944	5.998	100000
6	6.524	0	19	3.583	6.495	6.553	99996
8	6.685	0	19	3.671	6.655	6.715	100000
10	6.682	0	20	3.702	6.652	6.712	100000
12	6.631	0	20	3.716	6.601	6.661	99996
14	6.495	0	19	3.687	6.465	6.525	99988
16	6.390	0	20	3.684	6.360	6.420	100000
18	6.290	0	19	3.643	6.260	6.320	99990
20	6.172	0	20	3.613	6.143	6.202	100000
22	6.088	0	20	3.604	6.059	6.117	99990
24	5.999	0	19	3.547	5.970	6.027	99984
26	5.913	0	20	3.518	5.885	5.942	99996
28	5.839	0	20	3.506	5.811	5.868	99988
30	5.791	0	20	3.462	5.763	5.819	99990
32	5.704	0	20	3.431	5.676	5.732	100000
34	5.667	0	20	3.421	5.639	5.695	99994
36	5.588	0	20	3.361	5.561	5.615	99972
38	5.557	0	19	3.370	5.529	5.584	99978
40	5.497	0	20	3.320	5.470	5.524	100000

q=0.9

n	y	min	max	sd	lb	ub	obs
4	6.244	0	20	3.630	6.214	6.273	100000
6	6.800	0	20	3.897	6.768	6.831	99996
8	6.901	0	20	3.953	6.869	6.933	100000
10	6.835	0	20	3.944	6.803	6.868	100000
12	6.725	0	20	3.920	6.693	6.757	99996
14	6.576	0	20	3.890	6.544	6.608	99988
16	6.374	0	20	3.829	6.342	6.405	100000
18	6.247	0	20	3.789	6.217	6.278	99990
20	6.130	0	20	3.725	6.099	6.160	100000
22	6.004	0	20	3.704	5.974	6.034	99990
24	5.915	0	20	3.622	5.886	5.945	99984
26	5.792	0	20	3.572	5.762	5.821	99996
28	5.719	0	20	3.529	5.690	5.747	99988
30	5.644	0	20	3.502	5.615	5.672	99990
32	5.568	0	20	3.443	5.540	5.596	100000
34	5.499	0	20	3.384	5.472	5.527	99994
36	5.472	0	20	3.382	5.444	5.500	99972
38	5.396	0	20	3.313	5.369	5.423	99978
40	5.340	0	20	3.276	5.314	5.367	100000

Note. **n** is the number of agents. **y** is the (mean) productivity for each *n*. **min** and **max** are the minimum and maximum of individual productivity for each *n*, respectively. **sd** is the standard deviation of productivity. **lb** and **ub** are the lower and upper bounds of the 99%-confidence interval for *y*, respectively. **obs** is the number of observations.

Table S4 B. Detailed results of the AbiEdM that correspond to Result 5 and Fig. 3 B

q=0.1

n	y	min	max	sd	lb	ub	obs
4	7.111	0	20	3.551	7.083	7.140	100000
6	8.188	0	20	3.805	8.157	8.219	99996
8	8.791	0	20	3.965	8.759	8.824	100000
10	9.156	0	19	4.027	9.123	9.188	100000
12	9.373	0	20	4.075	9.340	9.406	99996
14	9.584	0	20	4.107	9.550	9.617	99988
16	9.754	0	20	4.135	9.720	9.788	100000
18	9.888	0	20	4.145	9.854	9.922	99990
20	9.955	0	20	4.189	9.921	9.989	100000
22	10.038	0	20	4.197	10.004	10.073	99990
24	10.100	0	20	4.228	10.066	10.135	99984
26	10.142	0	20	4.214	10.108	10.176	99996
28	10.162	0	20	4.251	10.127	10.197	99988
30	10.233	0	20	4.247	10.199	10.268	99990
32	10.231	0	20	4.274	10.196	10.266	100000
34	10.287	0	20	4.270	10.252	10.321	99994
36	10.269	0	20	4.294	10.234	10.304	99972
38	10.309	0	20	4.300	10.274	10.344	99978
40	10.312	0	20	4.308	10.277	10.347	100000

q=0.5

n	y	min	max	sd	lb	ub	obs
4	7.490	0	20	3.864	7.459	7.522	100000
6	8.479	0	20	4.018	8.446	8.512	99996
8	8.929	0	20	4.063	8.896	8.962	100000
10	9.223	0	20	4.066	9.190	9.256	100000
12	9.376	0	20	4.056	9.343	9.409	99996
14	9.412	0	20	4.088	9.379	9.445	99988
16	9.474	0	20	4.083	9.441	9.507	100000
18	9.494	0	20	4.072	9.461	9.527	99990
20	9.531	0	20	4.098	9.498	9.565	100000
22	9.538	0	20	4.069	9.505	9.572	99990
24	9.559	0	20	4.088	9.525	9.592	99984
26	9.519	0	20	4.102	9.485	9.552	99996
28	9.510	0	20	4.103	9.477	9.543	99988
30	9.505	0	20	4.092	9.471	9.538	99990
32	9.475	0	20	4.101	9.441	9.508	100000
34	9.463	0	20	4.106	9.429	9.496	99994
36	9.464	0	20	4.108	9.430	9.497	99972
38	9.475	0	20	4.099	9.441	9.508	99978
40	9.480	0	20	4.116	9.447	9.514	100000

q=0.9

n	y	min	max	sd	lb	ub	obs
4	7.735	0	20	4.110	7.701	7.768	100000
6	8.620	0	20	4.218	8.585	8.654	99996
8	8.913	0	20	4.190	8.879	8.947	100000
10	9.035	0	20	4.136	9.001	9.069	100000
12	9.046	0	20	4.106	9.012	9.079	99996
14	8.971	0	20	4.054	8.938	9.004	99988
16	8.873	0	20	4.037	8.840	8.906	100000
18	8.797	0	20	3.997	8.764	8.829	99990
20	8.710	0	20	3.955	8.677	8.742	100000
22	8.643	0	20	3.945	8.610	8.675	99990
24	8.566	0	20	3.916	8.534	8.598	99984
26	8.492	0	20	3.884	8.460	8.524	99996
28	8.423	0	20	3.857	8.392	8.455	99988
30	8.346	0	20	3.824	8.315	8.377	99990
32	8.288	0	20	3.811	8.257	8.319	100000
34	8.247	0	20	3.770	8.216	8.278	99994
36	8.184	0	20	3.749	8.154	8.215	99972
38	8.135	0	20	3.757	8.104	8.165	99978
40	8.055	0	20	3.729	8.025	8.086	100000

Note. **n** is the number of agents. **y** is the (mean) productivity for each *n*. **min** and **max** are the minimum and maximum of individual productivity for each *n*, respectively. **sd** is the standard deviation of productivity. **lb** and **ub** are the lower and upper bounds of the 99%-confidence interval for *y*, respectively. **obs** is the number of observations.

Table S4 C-i. Detailed results of the AbiTmM that correspond to Result 6 and Fig. 3 C-i

q=0.1								*q=0.5*								*q=0.9*							
n	**y**	**min**	**max**	**sd**	**lb**	**ub**	**obs**	**n**	**y**	**min**	**max**	**sd**	**lb**	**ub**	**obs**	**n**	**y**	**min**	**max**	**sd**	**lb**	**ub**	**obs**
4	5.281	0	18	2.707	5.259	5.303	100000	4	5.472	0	19	2.872	5.449	5.496	100000	4	5.566	0	19	2.962	5.542	5.590	100000
6	6.319	0	19	3.304	6.292	6.346	99996	6	6.530	0	20	3.436	6.502	6.558	99996	6	6.631	0	19	3.472	6.602	6.659	99996
8	7.157	0	19	3.616	7.127	7.186	100000	8	7.269	0	19	3.695	7.239	7.299	100000	8	7.315	0	19	3.673	7.285	7.345	100000
10	7.725	0	19	3.758	7.695	7.756	100000	10	7.810	0	19	3.820	7.779	7.842	100000	10	7.725	0	20	3.741	7.694	7.755	100000
12	8.129	0	19	3.844	8.098	8.161	99996	12	8.088	0	19	3.833	8.057	8.119	99996	12	7.920	0	19	3.713	7.890	7.950	99996
14	8.343	0	19	3.850	8.311	8.374	99988	14	8.297	0	19	3.821	8.266	8.328	99988	14	7.994	0	20	3.689	7.964	8.024	99988
16	8.513	0	20	3.843	8.481	8.544	100000	16	8.413	0	19	3.783	8.382	8.443	100000	16	8.033	0	19	3.652	8.003	8.063	100000
18	8.619	0	19	3.844	8.587	8.650	99990	18	8.451	0	20	3.756	8.421	8.482	99990	18	8.005	0	20	3.605	7.976	8.034	99990
20	8.647	0	19	3.795	8.616	8.678	100000	20	8.450	0	20	3.728	8.420	8.481	100000	20	7.955	0	20	3.580	7.926	7.984	100000
22	8.683	0	19	3.789	8.652	8.714	99990	22	8.387	0	19	3.711	8.356	8.417	99990	22	7.831	0	20	3.542	7.802	7.859	99990
24	8.667	0	20	3.734	8.637	8.698	99984	24	8.388	0	19	3.675	8.358	8.417	99984	24	7.760	0	20	3.502	7.731	7.788	99984
26	8.621	0	19	3.697	8.591	8.651	99996	26	8.292	0	20	3.661	8.262	8.322	99996	26	7.631	0	19	3.473	7.602	7.659	99996
28	8.554	0	20	3.687	8.524	8.584	99988	28	8.232	0	20	3.611	8.203	8.261	99988	28	7.519	0	20	3.452	7.491	7.548	99988
30	8.498	0	19	3.660	8.468	8.528	99990	30	8.149	0	19	3.630	8.120	8.179	99990	30	7.416	0	19	3.426	7.388	7.443	99990
32	8.443	0	20	3.641	8.413	8.472	100000	32	8.068	0	20	3.606	8.039	8.098	100000	32	7.321	0	19	3.378	7.294	7.349	100000
34	8.378	0	20	3.643	8.349	8.408	99994	34	7.988	0	20	3.563	7.959	8.017	99994	34	7.201	0	20	3.378	7.174	7.229	99994
36	8.297	0	20	3.602	8.267	8.326	99972	36	7.903	0	19	3.547	7.874	7.932	99972	36	7.133	0	20	3.335	7.106	7.161	99972
38	8.244	0	19	3.585	8.214	8.273	99978	38	7.827	0	19	3.540	7.798	7.856	99978	38	7.014	0	19	3.320	6.987	7.041	99978
40	8.136	0	19	3.578	8.107	8.165	100000	40	7.752	0	19	3.507	7.723	7.780	100000	40	6.953	0	20	3.297	6.926	6.980	100000

Note. **n** is the number of agents. **y** is the (mean) productivity for each *n*. **min** and **max** are the minimum and maximum of individual productivity for each *n*, respectively. **sd** is the standard deviation of productivity. **lb** and **ub** are the lower and upper bounds of the 99%-confidence interval for *y*, respectively. **obs** is the number of observations.

Table S4 C-ii. Detailed supporting results of the AbiTmM that correspond to Fig. 3 C-ii

q=0.1

n	tm	min	max	sd	lb	ub	obs
4	1.263	0	10	1.460	1.251	1.275	100000
6	2.120	0	11	1.834	2.105	2.135	99996
8	2.696	0	11	2.014	2.680	2.713	100000
10	3.088	0	13	2.102	3.071	3.105	100000
12	3.355	0	12	2.156	3.337	3.372	99996
14	3.537	0	12	2.173	3.519	3.555	99988
16	3.663	0	14	2.184	3.645	3.681	100000
18	3.768	0	13	2.190	3.750	3.786	99990
20	3.835	0	13	2.183	3.817	3.853	100000
22	3.915	0	13	2.187	3.898	3.933	99990
24	3.965	0	13	2.181	3.947	3.983	99984
26	4.004	0	13	2.176	3.986	4.021	99996
28	4.026	0	13	2.170	4.009	4.044	99988
30	4.052	0	13	2.159	4.035	4.070	99990
32	4.079	0	12	2.163	4.061	4.096	100000
34	4.092	0	13	2.160	4.074	4.109	99994
36	4.113	0	14	2.163	4.096	4.131	99972
38	4.138	0	12	2.153	4.121	4.156	99978
40	4.141	0	14	2.147	4.124	4.159	100000

q=0.5

n	tm	min	max	sd	lb	ub	obs
4	1.272	0	9	1.466	1.260	1.284	100000
6	2.135	0	11	1.847	2.120	2.150	99996
8	2.681	0	12	2.002	2.665	2.698	100000
10	3.085	0	13	2.115	3.068	3.102	100000
12	3.331	0	12	2.148	3.314	3.349	99996
14	3.518	0	12	2.158	3.500	3.535	99988
16	3.672	0	13	2.168	3.654	3.690	100000
18	3.774	0	13	2.171	3.756	3.792	99990
20	3.853	0	13	2.174	3.835	3.870	100000
22	3.889	0	13	2.177	3.871	3.907	99990
24	3.967	0	14	2.177	3.949	3.984	99984
26	4.000	0	14	2.169	3.983	4.018	99996
28	4.039	0	15	2.153	4.022	4.057	99988
30	4.064	0	14	2.162	4.046	4.082	99990
32	4.070	0	13	2.160	4.053	4.088	100000
34	4.112	0	13	2.154	4.095	4.130	99994
36	4.128	0	14	2.152	4.110	4.145	99972
38	4.127	0	14	2.158	4.110	4.145	99978
40	4.140	0	12	2.147	4.123	4.158	100000

q=0.9

n	tm	min	max	sd	lb	ub	obs
4	1.255	0	11	1.475	1.243	1.267	100000
6	2.138	0	11	1.851	2.123	2.153	99996
8	2.701	0	12	2.015	2.685	2.718	100000
10	3.071	0	12	2.084	3.054	3.088	100000
12	3.345	0	13	2.133	3.328	3.363	99996
14	3.523	0	12	2.154	3.505	3.540	99988
16	3.659	0	13	2.160	3.641	3.676	100000
18	3.773	0	12	2.167	3.755	3.790	99990
20	3.843	0	13	2.171	3.826	3.861	100000
22	3.899	0	13	2.167	3.882	3.917	99990
24	3.958	0	13	2.160	3.941	3.976	99984
26	4.001	0	12	2.160	3.984	4.019	99996
28	4.036	0	14	2.166	4.019	4.054	99988
30	4.055	0	13	2.165	4.037	4.073	99990
32	4.096	0	12	2.156	4.078	4.113	100000
34	4.091	0	14	2.155	4.073	4.108	99994
36	4.118	0	13	2.153	4.100	4.135	99972
38	4.126	0	13	2.147	4.108	4.143	99978
40	4.138	0	12	2.144	4.120	4.155	100000

Note. **n** is the number of agents. **tm** is the mean number of transmitted ideas for each *n*. **min** and **max** are the minimum and maximum numbers of transmitted ideas for each *n*, respectively. **lb** and **ub** are the lower and upper bounds of the 99% confidence interval for *y*, respectively. **obs** is the number of observations.

```
# Code for the RndBM
import numpy as np
import random
import csv
import datetime
# Define parameters
# n_min: the minimum number of agents
# n_max: the maximum number of agents
# q, l, k, m: parameters in the logistic Berliant-Fujita function
# te: the total number of periods
# tbc: the basic parameter for calculating the number of runs
n_min = 2
n_max = 40
q = 0.9
l = 0.8
k = 1
m = 1
te = 20
tbc = 100000

# Define the Berliant-Fujita function
def s(c0, d10, d20, q0):
  f = (c0**q0)*((d10*d20)**((1-q0)/2))
  return f
# Define the logistic function
def logistic(x, l0, k0, m0):
    f = l0 / (1 + np.exp(-k0 * (x - m0)))
    return f
# Define the logistic BF function
def p(c0, d10, d20, q0, l0, k0, m0):
  f = logistic(s(c0, d10, d20, q0), l0, k0, m0)
  return f

# Define files for data export and column labels
fl1 = open('/content/drive/MyDrive/ind-basic-full-q'+str(int(q*10))+'-
'+str(datetime.date.today())+'.csv', mode='a', newline="")
wf1 = csv.writer(fl1)
label_fu = ['r', 'n', 't', 'i', 'j', 'c', 'dij', 'dji', 'pr', 'suc']
```

```
wf1.writerow(label_fu)

fl2 = open('/content/drive/MyDrive/ind-basic-te'+str(int(te))+'-
q'+str(int(q*10))+'-'+str(datetime.date.today())+'.csv', mode='a',
newline="")
wf2 = csv.writer(fl2)
label_reg = ['y', 'n', 'r', 't', 'i', 'q']
wf2.writerow(label_reg)

# r: the run number in the collection
r = 0

for n in range(n_min, n_max+1,2):
# tr: the total number of runs for each n
   tr = int(tbc/n)

# run: the run number for each n
   for run in range(tr):

# Define variables for data storage
# a_ra: list of agents with randomized orders for each period
# set_id: list of sets of agents' ideas
# idea: name of each idea
     a_ra = list(range(n))
     set_id = [set() for i in range(n)]
     idea = 0
     exdata1 = []
     exdata2 = []

# t: each period of run
     for t in range(te):
# Within a randomized order of agents, the 1st and 2nd agents conduct a joint
creation, 3rd and 4th agents conduct the same, and so on
# c: the number of ideas that the two agents have in common
# d1, d2: the number of ideas that 1st agent knows but 2nd agent does not
know, and vice versa
# With probability calculated via the logistic BF function, the knowledge
creation is a success and add a new idea to the two agents' sets of ideas
```

```
      random.shuffle(a_ra)
      rdm = np.random.rand(int(n/2))
      for i in range(0,n-1,2):
        c = len(set_id[a_ra[i]].intersection(set_id[a_ra[i+1]]))
        d1 =  len(set_id[a_ra[i]].difference(set_id[a_ra[i+1]]))
        d2 =  len(set_id[a_ra[i+1]].difference(set_id[a_ra[i]]))
        pr = p(c,d1,d2,q,l,k,m)
        if pr > rdm[int(i/2)]:
          set_id[a_ra[i]].add(idea)
          set_id[a_ra[i+1]].add(idea)
          idea += 1
          ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 1]
          wf1.writerow(ex_da1)
        else:
          ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 0]
          wf1.writerow(ex_da1)

      if t == te - 1:
        for i in range(n):
          ex_da2 = [len(set_id[i]), n, r, t, i, q]
          wf2.writerow(ex_da2)

    r += 1

fl1.close()
fl2.close()
```

```
# Code for the RndEdM
import numpy as np
import random
import csv
import datetime
# Define parameters
# n_min: the minimum number of agents
# n_max: the maximum number of agents
# q, l, k, m: parameters in the logistic Berliant-Fujita function
# te: the total number of periods
# tbc: the basic parameter for calculating the number of runs
n_min = 4
n_max = 40
q = 0.9
l = 0.8
k = 1
m = 1
te = 20
tbc = 100000

# Additional code for education 1 starts
# p_ed: rate of education success
p_ed = 0.9
# Additional code for education 1 ends

# Define the Berliant-Fujita function
def s(c0, d10, d20, q0):
  f = (c0**q0)*((d10*d20)**((1-q0)/2))
  return f
# Define the logistic function
def logistic(x, l0, k0, m0):
    f = l0 / (1 + np.exp(-k0 * (x - m0)))
    return f
# Define the logistic BF function
def p(c0, d10, d20, q0, l0, k0, m0):
  f = logistic(s(c0, d10, d20, q0), l0, k0, m0)
  return f

```

```
# Define files for data export and column labels
fl1 = open('/content/drive/MyDrive/ind-edu-full-ped'+str(int(p_ed*10))+'-
q'+str(int(q*10))+'-'+str(datetime.date.today())+'.csv', mode='a',
newline="")
wf1 = csv.writer(fl1)
label_fu = ['r', 'n', 't', 'i', 'j', 'c', 'dij', 'dji', 'pr', 'suc', 'iedu',
'jedu']
wf1.writerow(label_fu)

fl2 = open('/content/drive/MyDrive/ind-edu-te'+str(int(te))+'-
ped'+str(int(p_ed*10))+'-q'+str(int(q*10))+'-
'+str(datetime.date.today())+'.csv', mode='a', newline="")
wf2 = csv.writer(fl2)
label_reg = ['y', 'n', 'r', 't', 'i', 'q', 'edu']
wf2.writerow(label_reg)

# r: the run number in the collection
r = 0

for n in range(n_min, n_max+1,2):
# tr: the number of runs for each n
   tr = int(tbc/n)

# run: the run number for each n
   for run in range(tr):

# Define variables for data storage
# a_ra: list of agents with randomized orders for each period
# set_id: list of sets of agents' ideas
# idea: name of each idea
     a_ra = list(range(n))
     set_id = [set() for i in range(n)]
     idea = 0
     exdata1 = []
     exdata2 = []

# Additional code for education 2 starts
# Set each agent the same idea 10000 with probability p_ed
```

```
     r_ed = np.random.rand(n)
     edu = []
     for i in range (n):
      if p_ed > r_ed[i]:
       set_id[i].add(10000)
       edu.append(1)
      else:
       edu.append(0)
# Additional code for education 2 ends

# t: each period of run
     for t in range(te):

# Within a randomized order of agents, the 1st and 2nd agents conduct a joint
creation, 3rd and 4th agents conduct the same, and so on
# c: the number of ideas that the two agents have in common
# d1, d2: the number of ideas that 1st agent knows but 2nd agent does not
know, and vice versa
# With probability calculated via the logistic BF function, the knowledge
creation is a success and add a new idea to the two agents' sets of ideas
       random.shuffle(a_ra)
       rdm = np.random.rand(int(n/2))
       for i in range(0,n-1,2):
         c = len(set_id[a_ra[i]].intersection(set_id[a_ra[i+1]]))
         d1 =  len(set_id[a_ra[i]].difference(set_id[a_ra[i+1]]))
         d2 =  len(set_id[a_ra[i+1]].difference(set_id[a_ra[i]]))
         pr = p(c,d1,d2,q,l,k,m)
         if pr > rdm[int(i/2)]:
           set_id[a_ra[i]].add(idea)
           set_id[a_ra[i+1]].add(idea)
           idea += 1
           ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 1,
edu[i], edu[i+1]]
           wf1.writerow(ex_da1)
         else:
           ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 0]
           wf1.writerow(ex_da1)

```

```
      if t == te - 1:
        for i in range(n):
          ex_da2 = [len(set_id[i])-edu[i], n, r, t, i, q, edu[i]]
          wf2.writerow(ex_da2)

    r += 1

fl1.close()
fl2.close()
```

```
# Code for the RndTmM
import numpy as np
import random
import csv
import datetime
# Define parameters
# n_min: the minimum number of agents in simulation
# n_max: the maximal number of agents in simulation
# q, l, k, m: parameters in the logistic Berliant-Fujita function
# te: the total number of periods
# tbc: the basic parameter for calculation of the number of runs
n_min = 4
n_max = 40
q = 0.9
l = 0.8
k = 1
m = 1
te = 20
tbc = 100000

# Additional code for transmission 1 starts
# p_tm: success rate of transmission of idea
p_tm = 0.3
# Additional code for transmission 1 ends

# Define the Berliant-Fujita function
def s(c0, d10, d20, q0):
  f = (c0**q0)*((d10*d20)**((1-q0)/2))
  return f
# Define the logistic function
def logistic(x, l0, k0, m0):
    f = l0 / (1 + np.exp(-k0 * (x - m0)))
    return f
# Define the logistic BF function
def p(c0, d10, d20, q0, l0, k0, m0):
  f = logistic(s(c0, d10, d20, q0), l0, k0, m0)
  return f

```

```
# Define files for data export and column labels
fl1 = open('/content/drive/MyDrive/ind-tm-full-ptm'+str(int(p_tm*10))+'-
q'+str(int(q*10))+'-'+str(datetime.date.today())+'.csv', mode='a',
newline="")
wf1 = csv.writer(fl1)
label_fu = ['r', 'n', 't', 'i', 'j', 'c', 'dij', 'dji', 'pr', 'suc', 'itm',
'jtm']
wf1.writerow(label_fu)

fl2 = open('/content/drive/MyDrive/ind-tm-te'+str(int(te))+'-
ptm'+str(int(p_tm*10))+'-q'+str(int(q*10))+'-
'+str(datetime.date.today())+'.csv', mode='a', newline="")
wf2 = csv.writer(fl2)
label_reg = ['y', 'n', 'r', 't', 'i', 'q', 'tm']
wf2.writerow(label_reg)

# r: the run number in the collection
r = 0

for n in range(n_min, n_max+1,2):
# tr: the number of runs for each n
   tr = int(tbc/n)

# run: the run number for each n
   for run in range(tr):

# Define variables for data storage
# a_ra: list of agents with randomized orders for each period
# set_id: list of sets of agents' ideas
# idea: name of each idea
     a_ra = list(range(n))
     set_id = [set() for i in range(n)]
     idea = 0
     exdata1 = []
     exdata2 = []

# Additional code for transmission 2 starts
```

```
# id_tm: list of the number of ideas that each agent does not create but
studies from interaction
    id_tm = [0 for i in range(n)]
# Additional code for transmission 2 ends

# t: each period of run
    for t in range(te):

# Within a randomized order of agents, the 1st and 2nd agents conduct a joint
creation, 3rd and 4th agents conduct the same, and so on
# c: the number of ideas that the two agents have in common
# d1, d2: the number of ideas that 1st agent knows but 2nd agent does not
know, and vice versa
# With probability calculated via the logistic BF function, the knowledge
creation is a success and add a new idea to the two agents' sets of ideas
      random.shuffle(a_ra)
      rdm = np.random.rand(int(n/2))
      for i in range(0,n-1,2):
        c = len(set_id[a_ra[i]].intersection(set_id[a_ra[i+1]]))
        d1 =  len(set_id[a_ra[i]].difference(set_id[a_ra[i+1]]))
        d2 =  len(set_id[a_ra[i+1]].difference(set_id[a_ra[i]]))
        pr = p(c,d1,d2,q,l,k,m)
        if pr > rdm[int(i/2)]:
          set_id[a_ra[i]].add(idea)
          set_id[a_ra[i+1]].add(idea)
          idea += 1
          ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 1,
id_tm[a_ra[i]], id_tm[a_ra[i+1]]]
          wf1.writerow(ex_da1)

        else:
          ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 0,
id_tm[a_ra[i]], id_tm[a_ra[i+1]]]
          wf1.writerow(ex_da1)

# Additional code for transmission 3 starts
        if t != te - 1:
         if d1 != 0 and p_tm > np.random.rand():
```

```
          set_id[a_ra[i+1]].add(random.choice(list(set_id[a_ra[i]].difference
(set_id[a_ra[i+1]]))))
          id_tm[a_ra[i+1]] += 1

         if d2 != 0 and p_tm > np.random.rand():
          set_id[a_ra[i]].add(random.choice(list(set_id[a_ra[i+1]].difference
(set_id[a_ra[i]]))))
          id_tm[a_ra[i]] += 1
# Additional code for transmission 3 ends

      if t == te - 1:
        for i in range(n):
          ex_da2 = [len(set_id[i])-id_tm[i], n, r, t, i, q, id_tm[i]]
          wf2.writerow(ex_da2)

    r += 1

fl1.close()
fl2.close()
```

```
# Code for the AbiBM
import numpy as np
import random
import csv
import datetime
# Define parameters
# n_min: the minimum number of agents
# n_max: the maximum number of agents
# q, l, k, m: parameters in the logistic Berliant-Fujita function
# te: the total number of periods
# tbc: the basic parameter for calculating the number of runs
n_min = 4
n_max = 40
q = 0.9
l = 0.8
k = 1
m = 1
te = 20
tbc = 100000

# Define the Berliant-Fujita function
def s(c0, d10, d20, q0):
  f = (c0**q0)*((d10*d20)**((1-q0)/2))
  return f
# Define the logistic function
def logistic(x, l0, k0, m0):
    f = l0 / (1 + np.exp(-k0 * (x - m0)))
    return f
# Define the logistic BF function
def p(c0, d10, d20, q0, l0, k0, m0):
  f = logistic(s(c0, d10, d20, q0), l0, k0, m0)
  return f

# Define files for data export and column labels
fl1 = open('/content/drive/MyDrive/ind-abi-basic-full-q'+str(int(q*10))+'-
'+str(datetime.date.today())+'.csv', mode='a', newline="")
wf1 = csv.writer(fl1)
label_fu = ['r', 'n', 't', 'i', 'j', 'c', 'dij', 'dji', 'pr', 'suc']
```

```
wf1.writerow(label_fu)

fl2 = open('/content/drive/MyDrive/ind-abi-basic-te'+str(int(te))+'-q'+str(int(q*10))+'-'+str(datetime.date.today())+'.csv', mode='a', newline="")
wf2 = csv.writer(fl2)
label_reg = ['y', 'n', 'r', 't', 'i', 'q']
wf2.writerow(label_reg)

# r: the run number in the collection
r = 0

for n in range(n_min, n_max+1,2):
# tr: the tnumber of runs for each n
   tr = int(tbc/n)

# run: the run number for each n
   for run in range(tr):

# Define variables for data storage
# a_ra: list of agents with randomized orders for each period
# set_id: list of sets of agents' ideas
# idea: name of each idea
     a_ra = list(range(n))
     set_id = [set() for i in range(n)]
     idea = 0
     exdata1 = []
     exdata2 = []

# t: each period of run
     for t in range(te):

# Within the ability order of agents, the 1st and 2nd agents conduct a joint creation, 3rd and 4th agents conduct the same, and so on
# c: the number of ideas that the two agents have in common
# d1, d2: the number of ideas that 1st agent knows but 2nd agent does not know, and vice versa
```

```
# With probability calculated via the logistic BF function, the knowledge
creation is a success and add a new idea to the two agents' sets of ideas
      random.shuffle(a_ra)
      rdm = np.random.rand(int(n/2))

# Additional code for sorting the order of matching starts
      las = []
      for i in range(n):
       las.append([a_ra[i] ,len(set_id[a_ra[i]])])
      las.sort(reverse=True, key=lambda x:x[1])
      a_ra = [i[0] for i in las]
# Additional code for sorting the order of matching ends

      for i in range(0,n-1,2):
        c = len(set_id[a_ra[i]].intersection(set_id[a_ra[i+1]]))
        d1 =  len(set_id[a_ra[i]].difference(set_id[a_ra[i+1]]))
        d2 =  len(set_id[a_ra[i+1]].difference(set_id[a_ra[i]]))
        pr = p(c,d1,d2,q,l,k,m)
        if pr > rdm[int(i/2)]:
          set_id[a_ra[i]].add(idea)
          set_id[a_ra[i+1]].add(idea)
          idea += 1
          ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 1]
          wf1.writerow(ex_da1)
        else:
          ex_da1 = [r, n, t, a_ra[i], a_ra[i+1], c, d1, d2, round(pr, 5), 0]
          wf1.writerow(ex_da1)

      if t == te - 1:
        for i in range(n):
          ex_da2 = [len(set_id[i]), n, r, t, i, q]
          wf2.writerow(ex_da2)

    r += 1

fl1.close()
fl2.close()
```