

Specialization and Cooperation in Research

Waka Cheung^a and Yew-Kwang Ng^{a*}

^a*Monash University, Australia*

Abstract

An optimal decision model is developed to explore the impact of the structure of knowledge on specialization and cooperation in research. It shows that, (1) the degree of specialization in knowledge increases with the depth of knowledge and cooperation efficiency, and decreases with the average number of disciplines that are required for the solution of the problem; (2) the problem which requires the knowledge of more disciplines is likely to be solved by more cooperative researchers. Conclusion (2) and part of conclusion (1) are tested by empirical evidence in economic study, where the sample comes from the papers published in the *Journal of Political Economy*.

JEL Classification: D80, J24

Keywords: Specialization, cooperation, research, co-authorship, division of labor

1. Introduction

During the last few decades, collaboration in research has become more and more popular. In economic research, for example, from 1976 to 1979 there were a total of 217 papers published in the article category of the *Journal of Political Economy* (JPE).¹ Among them are 141 papers (65%) by one author, 70 papers (32%) by two authors, and 6 papers (3%) by three authors. From 2000 to 2003 there were a total of 183 papers in the same category. Among them are 61 papers (33%) by one author, 88 papers (48%) by two authors, 30 papers (16%) by three authors, and 4 papers (2%) by four authors. If the author is regarded as the researcher of the relevant problem, it shows that not only the proportion of cooperative research, but also the number of cooperative researchers of a problem has increased.

According to Barnett et al. (1988), there are four reasons for the rise in multi-authorship. The first is the increase in the opportunity for specialization and division of labor. The second is the increase in the opportunity costs of time for reviewers of manuscripts, and

* We are grateful to an anonymous referee for some helpful suggestions.

^a Department of Economics, Monash University, Clayton, Victoria, Australia 3800. Email: kwang.ng@buseco.monash.edu.au. Fax: 613-99055476.

¹ The sample excludes papers in monographic volumes.

hence it may be necessary to offer co-authorship in order to elicit the level of effort required for a thorough review of a paper. The third is that co-authorship may increase the quality of a paper. The fourth is that research is risky and co-authorship can save effort (and time), thus co-authorship is a means of risk diversification. These factors, except the second, may be considered as the reasons for cooperative research.

The purpose of this paper is to construct a model to explore the relationship between specialization and cooperation in research, and to test the conclusions with some empirical evidence. Assume that each researcher has two stages, each endowed with one unit of time. In stage 1, the researcher gains the knowledge, in one or more disciplines. Assuming that the knowledge reveals a pyramidal structure (knowledge needs a wider background at a lower level), the more a researcher studies, the more knowledge he or she masters. In stage 2, he or she will solve a problem only (with a cooperative researcher) if he or she has all (part of) the knowledge to which the problem is related, otherwise he or she will fail to solve the problem. The trade-off is that if he or she specializes in acquiring knowledge in one discipline, he or she will have an advantage in solving a problem which solely requires knowledge in one discipline, but a disadvantage in solving a problem which requires the knowledge of more disciplines; in this case he or she needs to pay the cooperation cost for cooperative research. In any situation, a researcher knows the distribution of the necessary knowledge before the study commences. His or her optimal decision is decided by the maximization of expected utility by choosing the structures of knowledge in stage 1 and solving the problem in stage 2.

The model shows that, (1) the degree of specialization in knowledge increases with depth of knowledge and cooperation efficiency, and decreases with the average number of disciplines with which the problem is related; and (2) problems which require knowledge of more disciplines are likely to be solved by more cooperative researchers.

The approach developed in this article is related to a number of previous studies. In McDowell and Melvin's (1983) model (referred to here as the "MM model"), a researcher can choose sole authorship or co-authorship for a short paper or long paper, in which different choices are made in different times with different utilities. Fafchamps et al. (2006) develop a game model (referred to here as the "F model") to investigate co-authorship. They show that the likelihood of co-authorship (or matching) is related to the efficiency of the scientific network, the absolute difference between researchers' abilities and how well the different abilities complement each other.

The model discussed in this article makes the following changes to the MM and F models. Firstly, the MM model is relatively simple – with the linear restricted condition and equivalent linear utility function, the researcher has only one optimal choice and all choices are indifferent in special cases. In reality, a researcher has various choices across papers; he or she may be the sole author of one paper but a co-author of another. This option is incorporated into the new model. Secondly, compared with the F model, the new model emphasizes more varied aspects of co-authorship. This article provides a model to examine specialization and cooperation in research, but ignores the problem of matching. Thirdly, in the MM and F models, there is no allowance for the situation where a researcher fails to solve a problem. In the new model, if a researcher does not have the necessary knowledge, he or she fails to solve the problem. This article discusses only those problems which have been solved; there are, however, unsolved problems besides these.

It seems that all papers discussing co-authorship, except the MM and F models, focus on empirical analysis.² With various methods and from different points of view, McDowell and Melvin, Barnett et al., Piette and Ross (1992), Hudson (1996), and Laband and Tollison (2000) all validate the relationship between specialization and co-authorship.³

Unlike the previous empirical models, the empirical results in this study support the results that the depth of the necessary knowledge is positively correlated with co-authorship and the degree of specialization in knowledge. Among the co-authorships, the results show that problems related to more than one discipline are likely to be solved by more cooperative researchers.

This article is organized as follows. Section 2 provides a model, and Section 3 provides empirical analyses. Section 4 concludes the article, and all the proof of the propositions are contained in Section 5.

2. Model

In this section a model is developed to show the impact of the structure of knowledge on specialization and cooperation in research.

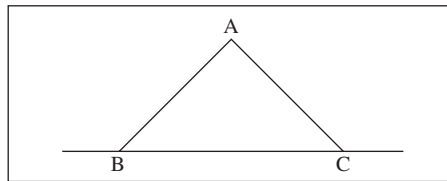
2.1 Assumptions

Assume that every researcher goes through two stages, each endowed with one unit of time. The researcher learns the knowledge in stage 1, and solves the problem in stage 2.

In this model, there are two disciplines denoted by **the space of knowledge**, K_1 and K_2 . For $i = 1$ and 2 , the space of knowledge $K_i = X_i \times R_i^+ = \{k_i \mid k_i = (x_i, d_i), -M < x_i < +M, 0 < d_i < +\infty\}$, where k_i represents the (element of) **knowledge** with **characteristic** (x_i) and **depth** (d_i), and M is a large enough positive number.

It is assumed that if a researcher wants to learn the knowledge (x, d) , he or she needs to learn all the background of (x, d) – the knowledge in the area of triangle $\triangle ABC$, where $A = (x, d)$, $B = (x - d, 0)$, and $C = (x + d, 0)$ (see Figure 1). The assumption reveals a pyramidal structure in knowledge: the knowledge needs a wider background at the lower level.

Figure 1: The background of knowledge A



² Fox and Faver (1984) discuss specialization and cooperation in research, but no model and empirical methods are used.

³ Some of these papers and others, such as Presser (1980), Schweser (1983), Hollis (2001), Sutter and Kocher (2004), Goyal et al., and Fafchamps et al. test other causalities of co-authorship.

The **cost function of learning** is defined as the amount of time spent learning, which relates to the area of knowledge (including the background research). In learning the knowledge of ΔABC (see Figure 1), for example, a researcher needs to spend the amount of time that is equal to the area of ΔABC .

This article explores research (the solving of a problem) only through the relationship between the problem and the necessary knowledge. The **problem** (q) is defined as the function of necessary knowledge. There are two situations. First, $q(k_i)$ denotes that the solution of the problem needs only knowledge k_i ($i=1$ and 2). Next, $q(k_1, k_2)$ denotes that the solution of the problem needs knowledge k_1 and k_2 , where $k_1 \in K_1$ and $k_2 \in K_2$. For the sake of simplicity, it is assumed that, to a researcher, the problems have the same depth, i.e. k_1 and k_2 have the same depth ($d_1 = d_2$).

In the **space of problem** $Q(d) = \{q \mid q = q(k_i) \text{ or } q = q(k_1, k_2), \text{ where } k_1 \in K_1, k_2 \in K_2, \text{ and } d_1 = d_2 = d\}$, it is assumed that the probability:

$$p\{q \mid q = q(k_1), k_1 \in K_1 \text{ and } d_1 = d\} = p_1 \quad (2.1a)$$

$$p\{q \mid q = q(k_2), k_2 \in K_2 \text{ and } d_2 = d\} = p_2 \quad (2.1b)$$

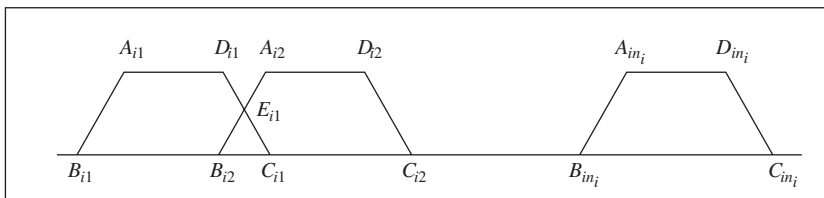
$$p\{q \mid q = q(k_1, k_2), k_1 \in K_1, k_2 \in K_2 \text{ and } d_1 = d_2 = d\} = p_3 \quad (2.1c)$$

where $p_1 + p_2 + p_3 = 1$.

It is assumed that the spaces of knowledge, K_1 and K_2 , are symmetric, and the knowledge in each space is symmetric: it is assumed that $p_1 = p_2$; for problem $q(k_i) \in Q(d)$, k_i is evenly distributed on X_i , where $i = 1$ and 2 ; and for problem $q(k_1, k_2) \in Q(d)$, (k_1, k_2) is evenly distributed on $X_1 \times X_2$. Under the symmetric conditions, a researcher would not consider the position of his or her knowledge. Thus researchers may choose knowledge across the space of knowledge, and hence it can be assumed further that a researcher can solve the problem by cooperating with another researcher who has complementary knowledge.

In stage 1 a researcher needs to learn what is needed for solving problems in stage 2. Since it is assumed that for a researcher the problems he or she faces have the same depth, say d , and assumed that the learning involves learning the background, the knowledge the researcher learns in space K_1 and K_2 satisfies Figure 2. For $i = 1$ and 2 , the union of segments $\bigcup_{j=1}^{n_i} A_{ij} D_{ij}$ represents the knowledge for potential problems, n_i is the number of segments in discipline K_i , $m_i = \sum_{j=1}^{n_i} (D_{ij} - A_{ij})$ represents the **width of the knowledge in K_i** , and $m = m_1 + m_2$ is the **width of the knowledge**. For $1 \leq j \leq n_i$, trapezium $A_{ij} B_{ij} C_{ij} D_{ij}$ represents the background of the knowledge denoted by segment $A_{ij} D_{ij}$.

Figure 2: The structures of the knowledge in K_i , for $i = 1$ and 2



Lemma 1: Given the depth of the knowledge d , assuming that a researcher spends l_i units of time learning the knowledge in K_i , then his or her potential structures of knowledge in K_i (see Figure 2) satisfy

$$m_i = \frac{l_i - n_i d^2 + \sum_{j=1}^{n_i-1} (\max \{d - 0.5s_{ij}, 0\})^2}{d} \quad (2.2)$$

where $i = 1$ and 2 , m_i is the width of the knowledge, $s_{ij} (= A_{ij+1} - D_{ij})$ represents the distance between the segments $A_{ij}D_{ij}$ and $A_{ij+1}D_{ij+1}$, and n_i is the number of the segments.

Lemma 1 shows that, given the learning time, he or she will achieve wider knowledge by choosing one segment, compared to choosing more than one segment. Since it is assumed that a researcher needs to learn the background of the knowledge which displays a triangle structure, when he or she chooses more than one segment - for example, segments $A_{i1}D_{i1}$ and $A_{i2}D_{i2}$ in Figure 2 - these two segments can share the background $\Delta E_{i1}B_{i2}C_{i1}$. As distance $D_{i1}A_{i2}$ decreases, point E_{i1} moves up and hence the two segments share more background. Given the time spent learning, the more the segments are shared, the wider the knowledge the researcher has. Thus, the researcher has the widest knowledge when points D_{i1} and A_{i2} overlap.

After learning the knowledge in stage 1, a researcher will solve the problem in stage 2. The **utility function of research** is defined as follows: if a researcher solves the problem solely, he or she will get one unit of utility. Next, if he or she solves the problem with another researcher, he or she will get 0.5 units of utility.⁴ Last, if he or she fails to solve the problem, he or she will get nothing.

If a researcher acquired the knowledge sets $S_1 \subset K_1$ and $S_2 \subset K_2$ in stage 1, he or she could solve solely the problem $q(k_i)$ if $k_i \in S_i$ ($i = 1$ and 2) and the problem $q(k_1, k_2)$ if $(k_1, k_2) \in S_1 \times S_2$. In addition, he or she could cooperate with another researcher to solve the problem $q(k_1, k_2)$ if $(k_1, k_2) \in S_1 \times S_2^c \cup S_1^c \times S_2$, where S^c denotes the complement set of S . However, he or she could not solve the problem $q(k_i)$ if $k_i \in S_i^c$ ($i = 1$ and 2) or the problem $q(k_1, k_2)$ if $(k_1, k_2) \in S_1^c \times S_2^c$.

The **cost function of research** can be defined in three situations. First, when a researcher solves a problem solely, he or she needs l units of time to spend solving the problem. Next, when he or she solves a problem with another researcher, he or she needs to spend αl ($\alpha > 0.5$) units of time, which includes the time spent researching together with the other researchers. Thus α is an inverse index of **cooperation efficiency**. Last, when he or she fails to solve a problem, he or she still needs βl ($\beta < 1$) units of time to spend investigating.

In stage 2 a researcher has one unit of time for research. Thus the average expected utility per unit of time is used as the researcher's objective function in the optimal decision.

⁴ Sauer's (1988) empirical results show that an individual's return from a co-authored paper with r authors is approximately $1/r$ times that of a single-authored paper.

$$\begin{aligned}
 AEU &= \frac{1}{l} \{ p_1 p \{ k_1 \in S_1 \} + p_2 p \{ k_2 \in S_2 \} + p_3 p \{ (k_1, k_2) \in S_1 \times S_2 \} \} \\
 &\quad + \frac{0.5}{\alpha l} p_3 p \{ (k_1, k_2) \in S_1 \times S_2^c \cup S_1^c \times S_2 \} \\
 &\quad + \frac{0}{\beta l} \{ p_1 p \{ k_1 \in S_1^c \} + p_2 p \{ k_2 \in S_2^c \} + p_3 p \{ (k_1, k_2) \in S_1^c \times S_2^c \} \} \\
 &= \frac{1}{l} \left\{ \left(p_1 + \frac{p_3}{2\alpha} \right) p \{ k_1 \in S_1 \} + \left(p_2 + \frac{p_3}{2\alpha} \right) p \{ k_2 \in S_2 \} + p_3 \left(1 - \frac{1}{\alpha} \right) p \{ (k_1, k_2) \in S_1 \times S_2 \} \right\} \\
 &= \frac{2\alpha p_1 + p_3}{4\alpha M l} \int_{S_1} dx_1 + \frac{2\alpha p_2 + p_3}{4\alpha M l} \int_{S_2} dx_2 + \frac{p_3(\alpha-1)}{4\alpha M^2 l} \iint_{S_1 \times S_2} dx_1 dx_2 \quad (2.3)
 \end{aligned}$$

2.2 The Optimal Decision

For a researcher, the optimal decision is choosing the structures of knowledge in stage 1 for the purpose of the maximum average expected utility in stage 2. From (2.2) and (2.3), it is given:

$$\max \left\{ \frac{2\alpha p_1 + p_3}{4\alpha M l} \int_{S_1} dx_1 + \frac{2\alpha p_2 + p_3}{4\alpha M l} \int_{S_2} dx_2 + \frac{p_3(\alpha-1)}{4\alpha M^2 l} \iint_{S_1 \times S_2} dx_1 dx_2 \right\} \quad (2.4)$$

st: $l_1 + l_2 = 1$, where $l_1 \geq 0$ and $l_2 \geq 0$

S_1 and S_2 , and m_1 and m_2 satisfy (2.2)

$$4M(2\alpha p_1 + p_3)d^3 \leq p_3(\alpha-1)(1-2d^2)^2 \quad (2.5a)$$

$$4M(2\alpha p_1 + p_3)d^3 \geq p_3(\alpha-1)(1-2d^2)^2 \quad (2.5b)$$

Proposition 1: The researcher has two potential choices in the optimal decision (2.4). When (2.5a) is satisfied, the researcher will acquire knowledge in two disciplines, K_1 and K_2 , in stage 1. When (2.5b) is satisfied, the researcher will specialize in acquiring knowledge in one discipline, K_1 or K_2 , in stage 1. In the optimal states, his or her average expected utility (AEU), the number of segments in discipline K_i (n_i), and the width of knowledge (m) satisfy Table 1.

Table 1
The optimal decision

| Structure | The variables in optimal states |
|----------------------------------|---|
| Generalize in K_1 and K_2 | $AEU = \frac{(2\alpha p_1 + p_3)(1 - 2d^2)}{4\alpha Mld} + \frac{p_3(\alpha - 1)}{4\alpha M^2 l} \left(\frac{1 - 2d^2}{2d}\right)^2$ $n_1 = n_2 = 1$ $m = m_1 + m_2 = (1 - 2d^2) / d, \text{ where } m_1 = m_2$ |
| Specialize in K_1 or K_2 | $AEU = \frac{(2\alpha p_1 + p_3)(1 - d^2)}{4\alpha Mld}$ $n_i = 1, \text{ where } i = 1 \text{ or } 2$ $m = m_i = (1 - d^2) / d, \text{ where } i = 1 \text{ or } 2$ |

Proposition 1 reveals several conclusions.

- (1) The number of disciplines a researcher chooses is an inverse index of the degree of specialization in knowledge. The number of disciplines a researcher chooses decreases with the depth of the knowledge and the cooperation efficiency, and increases with the average number of disciplines required for the solution of the problem.

This model shows that when the depth of the knowledge (d) and the cooperation efficiency ($1/\alpha$) are small enough to satisfy (2.5a), the researcher will learn two disciplines (the knowledge in K_1 and K_2). On the other hand, when the depth of the knowledge and the cooperation efficiency are large enough to satisfy (2.5b), the researcher will learn one discipline (the knowledge in K_1 or K_2).

Since it is assumed that a researcher needs to learn the background of the knowledge which displays a triangular structure (see Figure 1), there is a trade-off between the economies of specialization in knowledge and the cooperation cost. Compared with the situation where two disciplines are learnt, when only one discipline is learnt, a researcher saves time, but at the cost of increasing the probability of cooperation, in which he or she needs to pay the cost of cooperation. Thus, as the depth of the knowledge increases when the researcher needs to spend more time doing background research, or as the cooperation efficiency increases, he or she prefers more in the specialization of knowledge.

From (2.5a) and (2.5b), the larger the value of p_1 or the smaller the value of p_3 , the more likely it is that a researcher learns the knowledge in one discipline only. In addition, from (2.1a), (2.1b) and (2.1c) the average number of disciplines with which a problem is solved is

$$EI = p_1 + p_2 + 2p_3 = 2(1 - p_1) = 1 + p_3 \quad (2.6)$$

Thus, the number of disciplines a researcher learns increases with the average number of disciplines with which the problem is solved. From the previous analysis comes conclusion (1).

- (2) Besides the kind of knowledge, the width of knowledge (m) is an inverse index of the degree of specialization in knowledge. It is shown that m decreases with the depth of the knowledge (d) (see Table 1). This is mainly due to the limited time spent learning. The deeper the knowledge, the more time the researcher spends doing background research, and hence the narrower the knowledge gained is.

Conclusions (1) and (2) reveal that the degree of specialization in knowledge is positively correlated with the depth of knowledge. Since conclusions (1) or (2) cannot be tested directly, for convenience, a corollary of conclusions (1) and (2) will now be given.

Corollary 1: Under the condition that the problem has been solved, the condition probability in which the problem is solved by cooperative researchers is positively correlated with the depth of knowledge d .

Why is the conditional probability used instead of the probability in Corollary 1? There are two reasons. The first is that there is only a sample based on the problem having been solved (the paper having been published), but there is no sample based on the problem failing to be solved. Thus the probability in empirical analysis is the conditional probability given the condition that the problem has been solved. Next, from (2.3), the probability of the problem being solved by cooperative researchers is

$$p_3 p\{(k_1, k_2) \in S_1 \times S_2^c \cup S_1^c \times S_2^c\} \quad (2.7)$$

As depth d increases, set S_i decreases but set S_i^c increases, where $i = 1$ and 2 . Thus probability (2.7) may increase or decrease. However, under the condition that the problem has been solved, the conditional probability in which the problem is solved by cooperative researchers is

$$1 - \frac{p_1 p\{k_1 \in S_1\} + p_2 p\{k_2 \in S_2\} + p_3 p\{(k_1, k_2) \in S_1 \times S_2\}}{1 - p_1 p\{k_1 \in S_1^c\} - p_2 p\{k_2 \in S_2^c\} - p_3 p\{(k_1, k_2) \in S_1^c \times S_2^c\}} \quad (2.8)$$

As depth d increases, the conditional probability (2.8) increases. This conclusion includes the situation where d changes from structure 1 to structure 2.

- (3) Another characteristic of the specialization of knowledge is that a researcher prefers to choose an interconnected set (a segment) in one discipline (see Table 1).

As shown in Lemma 1, because of reducing the amount of background, a researcher will get a wider spectrum of knowledge by choosing one segment than by choosing more than one segment. Under the condition of the knowledge of a problem being evenly distributed in the space, there will be a higher probability of the researcher solving the problem when he or she has a wider area of knowledge. Thus, in the optimal decision, the researcher will choose one segment of knowledge in one discipline.

- (4) This model shows that the problems that need knowledge in more disciplines are likely to be solved by more cooperative researchers.

For the convenience of testing, this conclusion is represented by Corollary 2. Denote r the number of cooperative researchers, $q(I)$ the problem which needs I kinds

of knowledge to be solved, $p\{q(I) \mid I, r\}$ the condition probability in which the problem $q(I)$ is solved by r cooperative researchers under the condition of I and r , and $I(r)$ the optimal solution of (2.9).

$$\max_I p\{q(I) \mid I, r\} \quad (2.9)$$

Corollary 2: $I(r)$ increases with r .

$I(r)$ is the number of disciplines with which the problem is solved in maximum conditional probability. That $I(r)$ increases with r means that as the number of researchers increases, the number of disciplines with which the problem is solved in maximum conditional probability increases. It implies that the problem that needs more disciplines is likely to be solved by more cooperative researchers.

3. Evidence

In this section empirical evidence is used in an economic study to test the impact of the structure of knowledge on the specialization and cooperation in research.

Generally, economic study needs three disciplines: economics, mathematics, and statistics. Here, other relative knowledge is classified into one of these three disciplines. For example, some areas of social science are classified as economics, and software knowledge is classified as mathematics or statistics. Usually mathematics is used in model construction and statistics is used in empirical analysis.

A published paper represents a problem which has been solved. Since problems that have not been solved cannot be taken into account, the probability in this section is the condition probability under the published papers. The author / authors of a published paper is/are regarded as the researcher / cooperative researchers.

From 2000 to 2003 there were a total of 183 papers in the article category in JPE. Among them there are 61 papers (33%) with one author, 88 papers (48%) with two authors, 30 papers (16%) with three authors, and four papers (2%) with four authors. Divided by usage method, there is only one paper (0.5%) that does not use the model method and empirical method, 77 papers (42%) that use the model method but not the empirical method, 68 papers (37%) that use the empirical method but not the model method, and 37 papers (20%) that use these two methods at the same time.

Table 2
Definition of Variables

| Variable | Definition |
|--------------|--|
| $p_{\geq 1}$ | The conditional probability in which the paper has more than one author |
| $p_{\geq 2}$ | The conditional probability in which the paper has more than two authors |
| mod | Binary variable equals 1 if the modeling method is used |
| emp | Binary variable equals 1 if the empirical method is used |
| dyn | Binary variable equals 1 if the dynamic programming method is used |
| $nondyn$ | $Nondyn = mod - dyn$ |

In this model, variable $p_{\rightarrow 1}$ represents the conditional probability in which the paper has more than one author, and $p_{\rightarrow 2}$ the conditional probability in which the paper has more than two authors under the condition that the problem has been solved. Variables *mod* and *emp* are defined as follows: *mod* = 1 if the paper develops, not just cites, a microeconomic and / or macroeconomic, but not pure econometric, model, otherwise *mod* = 0; for the variable *emp*, *emp* = 1 if the paper includes empirical evidence (part of the conclusion is tested by an econometrics model and / or experiment), otherwise *emp* = 0. Moreover, the variable *mod* is divided into two parts, *dyn* and *nondyn*. First, *dyn* = 1 if the paper develops, not just cites, a dynamic program model, otherwise *dyn* = 0; next, *nondyn* = *mod* - *dyn*. This means that if the paper develops, not just cites, a non-dynamic program model, *nondyn* = 1, otherwise *nondyn* = 0.

Table 3
Probit Regression Models

| | (1) | (2) | (3) | (4) | (5) | (6)* |
|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|-------------------|
| Dependent variable | $p_{\rightarrow 1}$ | $p_{\rightarrow 1}$ | $p_{\rightarrow 1}$ | $p_{\rightarrow 2}$ | $p_{\rightarrow 2}$ | constant |
| Constant | -0.13 (-0.44) | | | | | |
| <i>mod</i> | 0.50 (1.84) | 0.38 (2.61) | 0.40 (3.06) | -1.13 (-6.22) | | 1.00 (117.7) |
| <i>nondyn</i> | | | | | -1.36 (-5.18) | |
| <i>dyn</i> | | | | | -0.86 (-3.35) | |
| <i>emp</i> | 0.46 (1.69) | 0.34 (2.18) | 0.35 (2.62) | -1.05 (-5.62) | -1.05 (-5.62) | 1.00 (110.6) |
| <i>mod*emp</i> | | 0.06 (0.20) | | 1.79 (5.35) | | -1.00 (-57.35) |
| <i>nondyn*emp</i> | | | | | 1.69 (3.63) | |
| <i>dyn*emp</i> | | | | | 1.79 (4.21) | |
| Log likelihood | -114.54 | -114.62 | -114.64 | -82.81 | -80.95 | $R^2=0.99$ |

Note.-Numbers in parentheses are *t*-statistics.

* Equation (6) is an ordinary least squares regression which tests the multicollinearity between constant and independent variables.

Probit regression equation (1) shows that constant is not a significant variable. This is because there is only one sample that doesn't use either the modeling or the empirical method in these 183 samples, and hence there is multicollinearity between constant, *mod*, and *emp*. The degree of multicollinearity is more serious when other variables, such as *mod*emp* (see regression (6)), *nondyn*emp*, or *dyn*emp*, are introduced into the model, thus the constant is omitted in the regression.

First, Corollary 2 will be tested. Equation (2) shows that variable *mod*emp* is not a significant variable. Equation (3) shows that the marginal contributions of *mod* and *emp* to $p_{\geq 1}$ are 0.40 and 0.35, respectively. This reveals that $p_{\geq 1}$ (the conditional probability in which the paper has more than one author) is positively correlated to the usage of the modeling method and / or the empirical method.

Equation (4) shows that, when the modeling method is the only used method in the paper, the marginal contribution of *mod* to $p_{\geq 2}$ is -1.13, and when the empirical method is the only used method in the paper, the marginal contribution of *emp* to $p_{\geq 2}$ is -1.05. This reveals that $p_{\geq 2}$ (the conditional probability in which the paper has more than two authors) is negatively correlated to the usage of only one method. However, when both methods are used in the paper, the marginal contributions of *mod* and *emp* to $p_{\geq 2}$ are 0.66 and 0.74, respectively. This reveals that $p_{\geq 2}$ is positively correlated to the usage of these two methods at the same time.

Comparing equations (3) and (4), it can be seen that the use of only one method in the paper is positively correlated with the conditional probability that the paper has two authors, but negatively correlated with the conditional probability that the paper has more than two authors, which, however, is positively correlated with the use of two methods at the same time. These empirical results support Corollary 2 – the problem that needs more disciplines is likely to be solved by more cooperative researchers.

Next, we will test the impact of the depth of knowledge on the degree of specialization in research (Corollary 1). Equation (5) shows that when the modeling method is the only used method in the paper, the marginal contributions of *nondyn* and *dyn* to $p_{\geq 2}$ are -1.36 and -0.86, respectively. On the other hand, when the empirical method is also used, the marginal contributions of *nondyn* and *dyn* to $p_{\geq 2}$ are 0.33 and 0.93, respectively. In these two situations the marginal contributions of *dyn* to $p_{\geq 2}$ are larger than those of *nondyn*. Since models of dynamic programming are on average “deeper” than models of non-dynamic programming, these empirical results show that the maximum conditional probability $p_{\geq 2}$ is positively correlated with the depth of the knowledge. Thus, it supports Corollary 1.

4. Conclusion

In this paper an optimal decision model has been developed to explore the impact of the structure of knowledge on specialization and cooperation in research. A researcher will solve a problem solely (with a cooperative researcher) if he or she has all (part of) the knowledge needed to solve the problem, otherwise he or she fails to solve the problem. With any problem, a researcher knows the distribution of the necessary knowledge before study commences. His or her optimal decision is the average expected utility maximization, achieved by choosing the structures of knowledge in stage 1 and studying the problem in stage 2.

The model shows that, (1) the degree of specialization in knowledge increases with the depth of the knowledge and the cooperation efficiency, and decreases with the average number of disciplines with which the problem is solved; and (2) the problem which requires knowledge of more disciplines is likely to be solved by more cooperative researchers.

Moreover, part of conclusion (1) – the degree of specialization in knowledge increasing with the depth of the knowledge – and conclusion (2) are supported by the empirical evidence of economic study, where the sample comes from the papers published in JPE from 2000 to 2003.

Appendices

Proof of **Lemma 1**

For $1 \leq j \leq n_i - 1$, there are two situations. First, when $s_{ij} < 2d$, trapezium $A_{ij}B_{ij}C_{ij}D_{ij}$ and trapezium $A_{ij+1}B_{ij+1}C_{ij+1}D_{ij+1}$ share the triangle $\Delta E_{ij}B_{ij+1}C_{ij}$ which has an area of $(d - 0.5s_{ij})^2$ (see Figure 2). Next, when $s_{ij} \geq 2d$, trapezium $A_{ij}B_{ij}C_{ij}D_{ij}$ and trapezium $A_{ij+1}B_{ij+1}C_{ij+1}D_{ij+1}$ do not share any triangle. Since it is assumed that the time spent learning equals the area of knowledge (including the background), this results in:

$$l_i = m_i d + n_i d^2 - \sum_{j=1}^{n_i-1} (\max \{d - 0.5s_{ij}, 0\})^2.$$

Thus

$$m_i = \frac{l_i - n_i d^2 + \sum_{j=1}^{n_i-1} (\max \{d - 0.5s_{ij}, 0\})^2}{d}$$

Proof of **Proposition 1**

There are two structures in the optimal decision (2.4).

- (1) The researcher will learn two disciplines (knowledge in K_1 and K_2) in stage 1. Lemma 1 shows that, given the time spent learning, the researcher will get wider knowledge by focussing on one segment, compared with choosing more than one segment. Under the condition that the knowledge is evenly distributed, he or she will have a higher probability of solving the problem when he or she has a wider area of knowledge. Thus, in the optimal decision he or she will choose one segment of knowledge in a discipline. Thus $n_1 = n_2 = 1$ and hence $m = m_1 + m_2 = (1 - 2d^2)/d$, where $m_1 = m_2$. Thus

$$AEU = \frac{(2\alpha p_1 + p_3)(1 - 2d^2)}{4\alpha M l d} + \frac{p_3(\alpha - 1)}{4\alpha M^2 l} \left(\frac{1 - 2d^2}{2d}\right)^2. \quad (5.1)$$

- (2) The researcher will learn the knowledge in K_1 or K_2 (specialize in knowledge in one discipline) in stage 1. With the same reason as in structure 1, this results in $n_i = 1$, where $i = 1$ or 2 and hence $m = m_i = (1 - d^2)/d$, where $i = 1$ or 2. Thus

$$AEU = \frac{(2\alpha p_1 + p_3)(1 - d^2)}{4\alpha M l d}. \quad (5.2)$$

Comparing (5.1) with (5.2), when

$$4M(2\alpha p_1 + p_3)d^3 \leq p_3(\alpha - 1)(1 - 2d^2)^2$$

is satisfied structure 1 should be chosen. On the other hand, when

$$4M(2\alpha p_1 + p_3)d^3 \geq p_3(\alpha - 1)(1 - 2d^2)^2$$

is satisfied, structure 2 should be chosen.

Proof of **Corollary 1**

From (2.3) the condition probability in which the problem is solved by two cooperative researchers under condition that the problem has been solved is

$$p\{r = 2 | \text{published}\} = 1 - \frac{p_1 p\{k_1 \in S_1\} + p_2 p\{k_2 \in S_2\} + p_3 p\{(k_1, k_2) \in S_1 \times S_2\}}{1 - p_1 p\{k_1 \in S_1^c\} - p_2 p\{k_2 \in S_2^c\} - p_3 p\{(k_1, k_2) \in S_1^c \times S_2^c\}}$$

When (2.5a) is satisfied, structure 1 is the best structure, thus

$$\begin{aligned} p\{r = 2 | \text{published}\} &= 1 - \frac{p_1 \frac{1-2d^2}{2Md} + p_3 \left(\frac{1-2d^2}{4Md}\right)^2}{(1-2p_1-p_3) + (p_1+p_3) \frac{1-2d^2}{2Md} - p_3 \left(\frac{1-2d^2}{4Md}\right)^2} \\ &= 1 - \frac{p_1 \frac{1-2d^2}{2Md} + p_3 \left(\frac{1-2d^2}{4Md}\right)^2}{(p_1+p_3) \frac{1-2d^2}{2Md} - p_3 \left(\frac{1-2d^2}{4Md}\right)^2} \end{aligned} \quad (5.3)$$

The condition probability $p\{r = 2 | \text{published}\}$ decreases with d .

When (2.5b) is satisfied, structure 2 is the best structure, thus

$$p\{r = 2 | \text{published}\} = 1 - \frac{p_1 \frac{1-d^2}{2Md}}{(1-2p_1-p_3) + (p_1+p_3) \frac{1-d^2}{2Md}} = \frac{p_3}{p_1+p_3} \quad (5.4)$$

The condition probability $p\{r = 2 | \text{published}\}$ does not change with d .

Given k , d is larger in (2.5b) than that in (2.5a). Comparing $p\{r=2 | \text{published}\}$ in these two structures:

$$1 - \frac{p_1 \frac{1-2d^2}{2Md} + p_3 \left(\frac{1-2d^2}{4Md}\right)^2}{(p_1+p_3) \frac{1-2d^2}{2Md} - p_3 \left(\frac{1-2d^2}{4Md}\right)^2} \leq 1 - \frac{p_1 \frac{1-2d^2}{2Md}}{(p_1+p_3) \frac{1-2d^2}{2Md}} = \frac{p_3}{p_1+p_3} \quad (5.5)$$

This means that $p\{r = 2 \mid \text{published}\}$ is larger in structure 2 than that in structure 1. Thus from (5.3), (5.4) and (5.5) it is shown that $p\{r = 2 \mid \text{published}\}$ decreases with depth d .

Proof of **Corollary 2**

From (2.3), it is shown that:

$$p\{q(1) \mid 1, 1\} = p\{k_1 \in S_1\} + p\{k_2 \in S_2\}$$

$$p\{q(2) \mid 2, 1\} = p\{(k_1, k_2) \in S_1 \times S_2\}$$

$$p\{q(1) \mid 1, 2\} = 0$$

$$p\{q(2) \mid 2, 2\} = p\{(k_1, k_2) \in S_1 \times S_2^c \cup S_1^c \times S_2\}$$

since

$$p\{(k_1, k_2) \in S_1 \times S_2\} \leq p\{k_1 \in S_1\} + p\{k_2 \in S_2\}$$

$$p\{(k_1, k_2) \in S_1 \times S_2^c \cup S_1^c \times S_2\} \geq 0.$$

Thus $I(1)=1$ and $I(2)=2$, and from there comes Corollary 2.

References

- Barnett, A.H., R.W. Ault and D.L. Kaserman, 1988, “The Rising Incidence of Co-authorship in Economics: Further Evidence”, *Review of Economics and Statistics* 70, 539–543.
- Fafchamps, M., M. van der Leij and S. Goyal, 2006, “Scientific Network and Co-authorship”, Discussion Paper Series No 256, Department of Economics, Oxford University.
- Fox, M.F. and C.A. Faver, 1984, “Independent and Cooperation in Research: The Motivations and Costs of Collaboration”, *The Journal of Higher Education* 55, 347–359.
- Goyal, S., M. van der Leij and J.L. Moraga-Gonzalez, 2006, “Economics: An Emerging Small World”, *The Journal of Political Economy* 114, 403–413.
- Hollis, A., 2001, “Co-authorship and the Output of Academic Economists,” *Labour Economics* 8, 503–530.
- Hudson, J., 1996, “Trends in Multi-authored Papers in Economics”, *Journal of Economic Perspective* 10, 153–158.
- Laband, D.N. and R.D. Tollison, 2000, “Intellectual Collaboration.” *Journal of Political Economy* 108, 632–622.
- McDowell, J.M. and M. Melvin, 1983, “The Determinants of Co-authorship: An Analysis of the Economics Literature”, *Review of Economics and Statistics* 65, 155–160.
- Piette, M.J. and K.L. Ross, 1992, “An Analysis of the Determinants of Co-authorship in Economics”, *Journal of Economics Education* 23, 277–283.

- Presser, S., 1980, "Collaboration and the Quality of Research," *Social Studies of Science* 10, 95–101.
- Sauer, R.D., 1988, "Estimates of the Returns to Quality and Co-authorship in Economic Academia", *The Journal of Political Economy* 96, 855–866.
- Schweser, C., 1983, "The Economics of Academic Publishing." *Journal of Economics Education* 14, 60–64.
- Sutter, M. and M. Kocher, 2004, "Patterns of Co-authorship among Economics Departments in the USA", *Applied Economics* 36, 327–333.