Developing Shared Knowledge in Growing Firms

Hongyi Li*
UNSW Australia

I develop a theory of knowledge sharing in organizations where coordinated activity requires shared knowledge, and knowledge sharing is *local* and *costly*. Because knowledge sharing is local, knowledge diffuses gradually across an organization. Because knowledge sharing is costly, diffusion may stall, resulting in inefficiently fragmented knowledge. The theory suggests that excessively rapid organizational growth may result in fragmentation or in the abandonment of the organization's early knowledge, and that these effects may persist in the long-run, even after the initial period of growth has ended. To avoid fragmentation, highly productive firms should deliberately constrain firm growth and avoid acquisition-based growth strategies. (*JEL* D21, D83, C73, J24)

1. Introduction

Coordination between individuals requires shared knowledge. A basketball team's performance improves with practice, as members of the team develop shared knowledge of how to coordinate and communicate on the court. Organizations that lack shared knowledge are often fragmented, so that individuals cannot coordinate effectively with each other. Some organizations recognize the costs of coordination failure and take actions to develop shared knowledge. For example, a senior executive at Southwest Airlines explained that Southwest chose to constrain its growth rate so as to maintain the firm's shared knowledge: "We have to grow in a controlled way . . . to make sure the way we conduct business in a city we enter is consistent with the way we conduct business throughout our system."[1]

This article proposes a theory of how shared knowledge develops, or fails to develop, within organizations. The focus of the analysis is on the

* School of Economics, UNSW Business School, UNSW Australia.
Email: hongyi@gmail.com

I thank Robert Gibbons, Oliver Hart, and Eric Van den Steen for their guidance and encouragement; the editor, Wouter Dessein, and three anonymous referees for their suggestions and criticisms; Philippe Aghion, Yidong Chong, Luis Garicano, Richard Holden, Anil Jain, Anton Kolotilin, Niko Matouschek, Michael Powell, Karen Sum-Ping, Tim Van Zandt, Birger Wernerfelt, Julie Wulf, M. Bumin Yenmez and especially Robert Akerlof for comments and advice; and MIT Sloan's Program on Innovation in Markets and Organizations for financial support.

1. See Heskett and Hallowell (1993).

The Journal of Law, Economics, and Organization, Vol. 33, No. 2
doi:10.1093/jleo/eww016
Advance Access published December 24, 2016

impact of an organization's growth strategy on the development of shared knowledge.

Ingredients of the Model. A key ingredient of the theory is the familiar idea that *tacit* knowledge is difficult to communicate (see, e.g., Polanyi 1966). This theory is not about knowledge that is easy to codify and replicate, such as formal procedures and codes; instead, it focuses on less tangible knowledge that one may think broadly of as implicit assumptions about appropriate behavior. Such knowledge may be of conventions of behavior. For example, an organization's decision-making process is more than the formal organizational chart; it also includes implicit assumptions and norms about how subordinates and superiors communicate, how conflict is resolved, and the extent to which dissent is tolerated. Another relevant form of tacit organizational knowledge is a shared understanding of the organization's values: that is, its goals and strategy. Organizational values are often rich and nuanced, and codified statements of values (e.g. "achieve the right balance between customer service and cost efficiency") are often vague, so that understanding of these values is necessarily implicit. Note that this conception of shared knowledge overlaps substantially with the literature on organizational routines[2] (e.g. Nelson and Winter 1982) and organizational culture (e.g. Schein 1985).

I use the term *method* to denote a coherent set of organizational assumptions. Shared knowledge of a method provides a common ground for collective decision-making, and allows groups to coordinate tacitly without having to reach explicit agreement; thus, a method serves as a coordination device for the organization. There may be multiple distinct methods (i.e. distinct sets of values and assumptions) that are effective in a given setting, but coordination is most effective when group members have shared knowledge of the same method.[3] A lack of shared knowledge within a group manifests as fragmentation into multiple cliques of individuals, whereby individuals can coordinate within, but not across, cliques.

For a group to use a method, knowledge of that method has to be disseminated to each individual in the group. It is difficult to quickly disseminate tacit knowledge across a large group, as Penrose (1959: 42–43) points out:

> It is impossible for a firm to expand efficiently beyond a certain point merely by drawing up a management "blueprint" . . . [and] . . . detailed "job descriptions" . . . An administrative group is . . . a collection of individuals who have had

2. Nelson and Winter (1982) define a routine to be *a regular and predictable pattern of behavior*, emphasize the role of routines in enabling coordination within organizations, and highlight the tacit nature of routinized knowledge.

3. As a language-based analogy, a French-speaking team may be as effective as a Spanish-speaking team, but a mix of French and Spanish speakers will be less effective.

> experience working together, for only in this way can "teamwork" be developed ....

Instead, as Penrose explains, shared knowledge is developed through interpersonal interaction during the production process. In other words, individuals engage in learning-by-doing-*together*.[4] This implies that knowledge propagation is local, with knowledge spreading from individual to individual, in a process analogous to diffusion.[5]

However, unlike the diffusion of a fluid, the process of propagation is not purely mechanical. Tacit knowledge is difficult to communicate; so knowledge sharing is time-consuming for the individuals involved, and diffusion may not take place if individuals choose not to incur the opportunity costs of knowledge sharing. Inefficient outcomes may arise because individuals make self-interested knowledge sharing decisions: an individual who teaches or learns a method generates benefits for everyone who knows that method, but internalizes only the benefit to himself while ignoring the effect on others.

To study these issues, I develop a simple discrete-time model of production and learning. In each period, workers undertake two types of activities. First, they engage in joint production with others who know the same method, and share the payoffs from production. Second, they may teach their method to another worker, or learn another worker's method. Knowledge sharing is local: each worker may only teach one other worker in a given period. Knowledge sharing is also costly: if a worker teaches or learns, he loses a fraction of his payoff from production in that period.

In this setting, I analyze a stylized organization that, growing over three periods, eventually hires a total of four workers. In particular, I compare two growth paths for the organization: slow growth where the hiring of workers is staggered over time, versus fast growth where worker hiring is front-loaded. (Later in the article, I also consider a version of the model that accommodates more than four workers within the organization.)

Implications of the Model. An organization's growth rate has a long-run effect on the development of shared knowledge: a slowly growing organization is more likely to successfully develop shared knowledge and avoid persistent fragmentation. In a rapidly growing organization, a single method cannot diffuse quickly enough to keep up with the expansion of

4. Relatedly, Weber and Camerer (2003), in summarizing the literature on corporate culture, state: "Culture is developed in an organization through joint experience, usually over long periods of time."

5. Consider the department store chain Nordstrom. For a long time, Nordstrom's employee handbook stated only to "Use best judgment in all situations. There will be no additional rules. Please feel free to ask your department manager, store manager, or division general manager any question at any time." My interpretation of this statement is that Nordstrom recognized that tacit shared knowledge could not be communicated formally, and instead relied on local knowledge sharing between employees.

the organization. As a result, multiple distinct methods diffuse simultaneously in equilibrium; in the short run, the organization becomes fragmented into multiple productive cliques. Each individual may then be unwilling to incur the opportunity cost of further knowledge sharing. Consequently, propagation stalls and fragmentation persists even after organizational growth has leveled off. On the other hand, when an organization grows slowly, a single method can spread across the organization, keeping pace with the expansion of the organization, so that new hires learn the incumbent method rather than form new cliques.

Put another way: for a given increase in size over a given time period, an organization that follows a steady, moderate growth path will be more successful at developing shared knowledge than an organization that pursues rapid and abrupt expansion that then levels off. In the latter case, the fragmentation induced by rapid growth cannot be undone merely by subsequently keeping organizational size constant.

As a corollary, rapid organizational growth may also result in loss of the organization's early knowledge. With slow growth, an initial method pervades the organization at every step, and subsequently persists. In contrast, fast growth allows competing cliques to emerge in the short run. The organization may then end up coordinating on an alternative method in the long run, even if it is less productive than the organization's initial method.

An organization's optimal growth rate depends on the organization's initial productivity (which, in the model, corresponds to the quality of the organization's initial knowledge endowment). Organizations that are endowed with highly productive methods benefit more from developing shared knowledge than less well-endowed organizations, and thus may deliberately constrain growth to improve workers' incentives to share knowledge. They may also choose to grow organically, by hiring new employees with no prior shared experience, rather than acquiring other organizations whose workers have already developed shared knowledge.

Because the development of shared knowledge—in particular, the local nature of knowledge sharing and the costs of teaching and learning—is modeled in reduced form, alternative interpretations may be given for the economic mechanisms that emerge from the model. For example, one may think of methods as representing an individual's social identity (see, e.g., Akerlof 2016). Under this interpretation, shared identity enables cooperation (and thus joint production), and may be transmitted through interpersonal interaction. Corresponding to the learning costs in our model, an individual who abandons an old identity and acquire a new one incurs an emotional cost.[6] Thus, our results may be relevant to the development of shared identity within a growing community or organization.

6. For example, in the 1998 Daimler-Chrysler merger, Chrysler employees expressed anger at having to adopt the Daimler culture (Vlasic and Stertz 2002).

Related Literature. A growing literature studies the process of "learning to coordinate"; this literature includes Crawford and Haller (1990), Ellison and Fudenberg (1995), and Blume and Franco (2007). These papers typically model how dyads may, through repeated interaction, work out the details of coordination; for example, in Ellison and Holden (2014), a principal and an agent develop a set of rules for effective communication over time. In contrast, my paper studies how patterns of coordination (modeled as fragmentation patterns) form *across* larger groups over time.[7]

A number of papers study knowledge transmission and acquisition in large groups. Some papers (e.g. Ellison and Fudenberg 1995; Lazear 1995) study how behavior or knowledge may diffuse across a large group when individuals do not optimize over transmission decisions, but instead follow exogenously specified rules. In particular, Lazear (1995) models behavior as evolving through a "mating" process whereby individuals may adopt the behavior of other people that they encounter. Compared to these papers, the present paper models knowledge transmission as a strategic decision subject to moral hazard; this logic drives a number of key and distinctive results such as the persistence of inefficient fragmentation. Lazear (1999) presents a model of shared knowledge where the incentives for knowledge transmission depends on the existing pattern of knowledge. His analysis is static and focuses on the stability (or lack thereof) of existing patterns. The present model builds on Lazear (1999) by studying the dynamics of shared knowledge, and models knowledge transmission as being local; which allows for an analysis of the evolution of knowledge patterns as well as implications about organizational growth.

This article's premise that shared knowledge enables organizational coordination draws from the literature on organizational culture.[8] Schein (1985) provides an influential description and analysis of organizational culture as shared assumptions and values. Van den Steen (2005, 2010a, 2010b) defines organizational culture as shared beliefs between workers, and shows how shared beliefs may emerge over time through selection and learning effects within an organization. Closer to the present article's perspective, Crémer (1993) models organizational culture in terms of shared knowledge, and points out that cultural homogeneity facilitates coordination at the cost of more frequent organizational mistakes. Relatedly, Kreps (1990) views culture as a focal equilibrium in an environment with multiple equilibria, and provides an informal discussion of culture change in terms of "switching the equilibrium." These analyses, while

7. Relatedly, Weber and Camerer (2003), Feiler and Camerer (2010), Weber (2006), and Selten and Warglien (2007) present experimental studies of how dyads and groups learn to coordinate over time. These experiments focus on the resolution of strategic uncertainty in team-theoretic settings, whereas the current paper considers a moral-hazard-in-teams setting where individuals may not have the appropriate incentives to resolve strategic uncertainty.

8. See Hermalin (2013) for a survey of the economic literature on corporate culture.

sharing the present article's premise regarding the benefits of shared knowledge, are orthogonal to my focus on the diffusion of cultural knowledge, and thus do not address issues such as persistent inefficiencies or the effect of organizational growth.

A number of recent papers analyze the economics of organizational knowledge.[9] Crémer et al. (2007) introduce a model of optimal communication codes and analyze the optimal code in a static, team-theoretic setting; they point out that fragmentation may be optimal when different divisions of an organization have to perform different tasks that have different communication requirements. In contrast, this article assumes that uniformity of shared knowledge is always optimal, then provides a dynamic account of how fragmentation may arise due to the local nature of knowledge sharing, and highlights the role of organizational growth in the fragmentation of organizational knowledge. Prescott and Visscher (1980) argues that the need to learn about the match quality of new employees constrains organizational growth. The present article, while sharing a focus on organizational growth, models a very different mechanism and produces distinctive implications: about fragmentation, adaptation, and organic versus acquisitive growth.

Section 2 introduces the model and Section 3 characterizes equilibrium knowledge diffusion in the model. Section 4 studies how an organization's early knowledge affects subsequent knowledge diffusion. Section 5 analyzes an organization's choice of growth strategy; Section 5.1 analyzes the optimal growth rate, and Section 5.2 analyzes the trade-off between organic and acquisitive growth. Section 6 modifies the basic model from Section 2 to accommodate more than four workers, and shows that the main insights from Section 3 are preserved. Section 7 summarizes. Appendix A contains the proofs.

2. The Model

A single organization contains up to four workers, indexed by $w_i : i \in \{1, 2, 3, 4\}$, who interact over three periods, $t = 1, 2, 3$. The organization grows over time by hiring new workers from outside. In each period t, W_t denotes the set of workers in the organization. To study how an organization's growth path affects its long-run outcomes, I restrict attention to two growth paths which differ in the initial rate of growth. Under *fast growth* (f), the organization hires four workers at the start of $t = 1$, and ceases hiring thereafter. Under *slow growth* (s), the organization hires two workers at the start of $t = 1$, and two more workers at the start of $t = 2$. Workers hired at $t = 1$ are called *early hires*, and workers hired at $t = 2$ are called *late hires*. The two growth paths are illustrated in Figure 1.

9. See, e.g., Meagher and Prasad (2016).

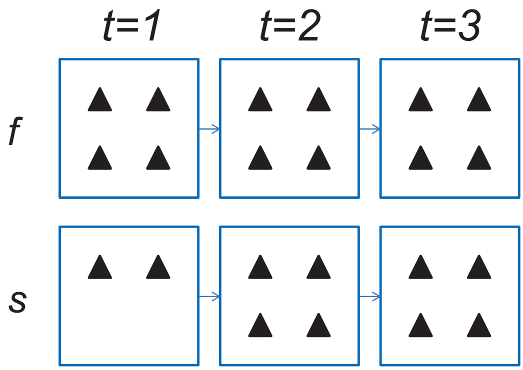

Figure 1. Fast Growth (*f*) versus Slow Growth (*s*).

At the start of the game, the growth path $g \in \{f, s\}$ is fixed.[10]

Knowledge. The relevant domain of knowledge is a set of methods. Each worker w_i knows one method $m_{i,t}$; he discards his old method whenever he learns a new method. Each new worker is endowed with a distinct method when he joins the organization.

Note that in any period, workers may be partitioned, based on their knowledge, into *cliques*. So, each clique is associated with a method m, and is the subset of workers that know m:

$$C_{m,t} = \{w_i \in W_t : m_{i,t} = m\}. \tag{1}$$

In each period t, I call the distribution $\Gamma_t = \{|C_{m_1,t}|, \ldots, |C_{m_k,t}|\}$ of clique sizes the fragmentation *pattern*. (See Figure 2.)

Knowledge Sharing. In each of the first two periods, each worker may (if he wishes) teach one other worker his method, or learn one other worker's method. If worker w_i learns from worker w_j (equivalently, worker w_j teaches worker w_i) in period t, then worker w_i updates his method in the subsequent period:

$$m_{i,t+1} = \begin{cases} m_{j,t} & : i \text{ learns from } j \text{ in period } t \\ m_{i,t} & : i \text{ does not learn in period } t \end{cases}. \tag{2}$$

In each period, each worker w_i is randomly either available ($A_i = 1$) or unavailable ($A_i = 0$) for production. Teaching or learning reduces the

10. Note that there is no dynamic commitment problem over the choice of growth path; after all, the difference between growth paths boils down to how many workers to hire in the first period.

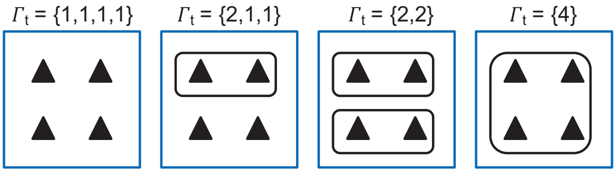

Figure 2. Examples of Fragmentation Patterns Γ_t with Four Workers ($|W_t| = 4$). Rectangles with Rounded Corners Represent Cliques. Note that $\Gamma_t = \{3, 1\}$ is Omitted from the Figure.

probability that he is available, and learning is more costly than teaching in terms of reducing availability:

$$Pr[A_i = 1] = \begin{cases} 1: & i \text{ neither teaches nor learns} \\ 1 - \alpha_{\text{teach}}: & i \text{ teaches} \\ 1 - \alpha_{\text{learn}}: & i \text{ learns} \end{cases}, \tag{3}$$

with $0 \leqslant \alpha_{\text{teach}} < \alpha_{\text{learn}} \leqslant 1$.

Production and Payoffs. In each period t, available workers engage in production. To capture the premise that production requires coordination via shared knowledge, each worker's production payoff is proportional to the number of available coworkers who know the same method: if i is in clique C, then[11]

$$\pi_{i,t} = A_i \sum_{j \in C \setminus i} A_j. \tag{4}$$

Notice that teaching and learning impose opportunity costs: each worker's expected payoff is increasing in the likelihood that he is available.

For a worker w_i who joins the organization in period t, his full-game payoff is the discounted sum of his stage payoffs:

$$V_i = \sum_{\tau=t}^{3} \Delta_\tau \pi_{i,\tau}, \tag{5}$$

where $\Delta_1 = 1$, $\Delta_2 = \delta$, $\Delta_3 = \frac{\delta^2}{1-\delta}$ with $\delta \in (0, 1)$. I interpret the third period as the *long-run* outcome for the organization; think of it as representing an infinite number of future periods, with organizational productivity in each of these "long-run" periods determined by knowledge sharing choices made in the first two periods.

11. Note that (4) corresponds to each worker receiving the Shapley value of his contribution to total period-t production, so we may interpret payoffs as deriving from a period-by-period bargaining process. Alternatively, we may think of this payoff function as capturing aspects of incentive schemes that motivate workers to contribute to organizational output.

Stage Game. The stage timing is as follows: (a) Workers make teaching/learning choices. (b) Each worker's availability is realized. (c) Production takes place. Payoffs are realized. (d) Workers' methods $\{m_{i,t+1}\}_{i \in W_t}$ are updated.

In step (a), teaching and learning choices are made sequentially via the following extensive form (all actions are publicly observed):

At the start of the period, a random ordering of all the workers is publicly realized; each possible ordering is equally likely. One or more rounds are then played, where each worker takes a turn in the realized order. At his turn, each worker w_i either does nothing or chooses another worker w_j and offers to teach him. The offer recipient w_j immediately chooses whether to accept or reject w_i's offer before the next player takes his turn.

The round ends once an offer is accepted, or if all players have taken a turn. If an offer was accepted in the previous round, and if there are players remaining who have not accepted an offer or had an offer accepted, then an additional round is played by the remaining players in the same order. Otherwise, teaching and learning choices are fixed, and production takes place. (In the current setting where there are at most four players, at most two rounds are played in each period.)

I analyze pure-strategy subgame-perfect equilibria of this game.[12] By convention, asterisks denote equilibrium outcomes, e.g., Γ_2^* denotes the equilibrium $t = 2$ pattern.

2.1 Discussion of Modeling Choices

Before proceeding to the analysis, let me discuss some aspects of the model.

I model the costs of knowledge sharing as opportunity costs, and I assume that learning incurs higher opportunity costs than teaching ($\alpha_{\text{learn}} > \alpha_{\text{teach}}$). The premise is that acquiring tacit knowledge largely involves learning-by-doing. So, in the current context where tacit knowledge is used in coordination, one learns a method over time by adopting it to interact with others who use that method. A learner is unproductive during the learning process because he is inept at the method he is adopting,[13] and he is using his existing method (which he is adept at) less, or not at all. On the other hand, a teacher continues to use his existing method; his role is merely to set examples and provide feedback to the learner. Teachers thus lose little, if any, of their productivity during the learning process.

12. Although this solution concept admits multiple equilibria, these equilibria are identical with respect to outcomes of interest (payoffs, fragmentation patterns, etc.). The exposition will abstract from the details of equilibrium multiplicity.

13. In fact, I implicitly assume that the learner is *utterly* inept at using this new method: in the period when he is learning, his payoff is derived solely from joint production with his existing clique-mates, using his existing method.

The assumption $\alpha_{\text{learn}} > \alpha_{\text{teach}}$ is important. If instead $\alpha_{\text{learn}} < \alpha_{\text{teach}}$, then the long-run knowledge fragmentation is independent of the growth path. I discuss this further in Section 3 (Remark 1).

I model *local* knowledge sharing by assuming that in each period, each player can only teach one other player. The underlying premise is that unlike formal knowledge, tacit knowledge cannot be transmitted in a one-to-many, lecture-style format, but instead requires intimate interaction between teachers and learners. The specific one-teacher-to-one-learner assumption is simply a convenient way to represent this premise in a four-player setup.

I model an organization that initially lacks the shared knowledge necessary for coordinated activity by assuming that each worker is endowed with a distinct method, so that joint production is impossible unless players share knowledge with each other. This reflects the premise that the key obstacle to coordination is the cost of sharing existing knowledge, rather than creating new knowledge (in the form of rules, language, etc.). It seems straightforward to extend the model to incorporate the cost of knowledge creation, by assuming that workers are not endowed with methods, but have to create them at some cost.

Workers have limited memory (i.e. whenever a worker learns a new method, he discards his old method) in the model. This assumption seems strong, but does not drive the main insights; I obtain similar results in a model where workers retain memory of old methods and, in each period, choose which method to use for production. The assumption dramatically simplifies the details of the analysis by reducing the complexity of the state space (the set of possible configurations of organizational knowledge).

In the model, teaching and learning decisions cannot be contracted upon (which rules out Coasian bargains, or centralization of teaching/learning decisions). This reflects the premise that it is difficult to verify the transmission or acquisition of tacit knowledge. The essential insights would be preserved if we allowed contracts to be written over knowledge sharing outcomes, as long as the costs of writing or verifying such contracts was sufficiently high.

I consider a stylized model with four players and three periods. As we will see in the course of analyzing the model, this is the simplest setup that allows us to analyze how knowledge fragmentation may arise and persist under local knowledge sharing. In Section 6, I consider a modification of the model which accommodates many workers and many periods, and show that the main insights extend to this setting.

Teaching and learning decisions are made sequentially in the model. Alternatively, I could specify that workers move simultaneously; however, such a setup produces multiple equilibria due to strategic complementarities between potential teachers and between potential learners. Such complications are eliminated with the sequential-move structure, at least in the four-worker structure. In Section 6, with more than four workers, additional equilibrium selection rules are introduced to reduce equilibrium multiplicity to a manageable level. Relatedly, the inclusion of multiple rounds in the sequence of offers may seem curious at first glance. This feature does not affect the results in any substantive way; however, it is included because it dramatically simplifies the details of the proofs.[14]

3. Equilibrium Fragmentation Patterns

This section derives equilibrium outcomes for each growth path. The analysis makes two points. First, it demonstrates that an organization's choice of growth path has significant and persistent effects on the long-run extent of shared knowledge. Second, it identifies a wedge between efficiency and equilibrium arising from the public-good nature of knowledge sharing. Some terminology: players in the same clique are *clique-mates*. A clique of size $|C|$ is a $|C|$-clique.

I start by describing how knowledge sharing takes place over time for each growth path (as illustrated in Figure 3). I will then explain the equilibrium logic after stating the main result. To compare outcomes between growth paths, I say that the long-run pattern Γ_3 is *unfragmented* if every worker knows the same method (i.e. $\Gamma_3 = \{4\}$); otherwise, Γ_3 is *fragmented.*

In the first period, for both growth paths and for all discount factors, every worker either teaches or learns. Under slow growth, there are two early hires; one early hire teaches the other at $t = 1$, leading to $\Gamma_2 = \{2, 1, 1\}$. Under fast growth, there are four early hires; two early hires teach and two early hires learn at $t = 1$, leading to $\Gamma_2 = \{2, 2\}$.

In the second period, for both growth paths, further knowledge sharing that leads to unfragmented knowledge ($\Gamma_3 = \{4\}$) occurs if and only if workers have sufficiently low discount factor δ. Under slow growth, this entails the workers in the 2-clique teaching the workers in the 1-clique. Under fast growth, this entails the workers in one of the 2-cliques teaching the workers in the other two cliques. The main result of this section is that the δ-threshold, above which unfragmented knowledge is produced, is lower under slow growth than under fast growth. In other words, faster growth leads to more knowledge fragmentation, strictly so for intermediate discount factors.

Proposition 1. (Illustrated in Figure 3)

- If $\frac{\delta}{1-\delta} < \frac{\tilde{\alpha}_{\text{teach}}}{2}$, then the long-run pattern is fragmented ($\Gamma_3^* = \{2, 2\}$) under both growth paths.
- If $\frac{\tilde{\alpha}_{\text{teach}}}{2} < \frac{\delta}{1-\delta} < \frac{\tilde{\alpha}_{\text{learn}}}{2}$, then the long-run pattern is fragmented ($\Gamma_3^* = \{2, 2\}$) under fast growth but unfragmented ($\Gamma_3^* = \{4\}$) under slow growth.

14. Specifically, it eliminates incentives for a worker to strategically choose his offer recipient based on the order in which workers move.

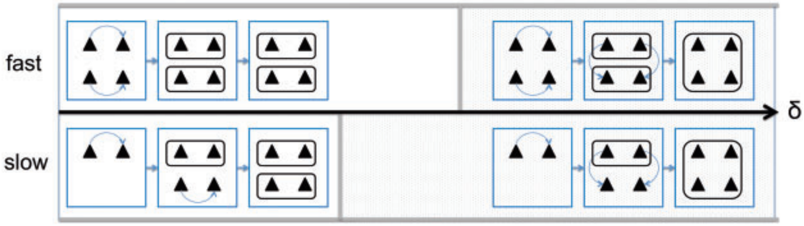

Figure 3. Long-run Outcomes: Fast Growth versus Slow Growth.

- If $\frac{\delta}{1-\delta} > \frac{\tilde{\alpha}_{\text{learn}}}{2}$, then the long-run pattern is unfragmented ($\Gamma_3^* = \{4\}$) under both growth paths.

I defer the explanation for why every worker either teachers or learns at $t = 1$ until after we have analyzed behavior at $t = 2$. For now, simply keep in mind that slow and fast growth lead to different fragmentation patterns at $t = 2$, which in turn will affect incentives for further knowledge sharing. We examine knowledge sharing at $t = 2$ for each growth path in turn; this will give us the essential intuition for Proposition 1.

Fast growth: At the start of $t = 2$, there is an even fragmentation pattern with two equally-sized 2-cliques. Unfragmented knowledge is achieved if both workers from one 2-clique teach both workers from the other 2-clique. As we will derive shortly, this occurs if and only if workers are sufficiently patient: $\frac{\delta}{1-\delta} > \frac{\tilde{\alpha}_{\text{learn}}}{2}$ where $\tilde{\alpha}_{\text{learn}} = 1 - (1 - \alpha_{\text{learn}})^2$. This equation highlights the basic benefit-cost trade-off that workers face. Consider a worker w_i who is deciding whether to offer to teach worker w_j from another clique. Worker w_i increases his $t = 3$ payoff by teaching and thus expanding his $t = 3$ clique, but reduces his availability for production at $t = 2$, and thus incurs an opportunity cost. This tradeoff is favorable for w_i if and only if he is sufficiently patient.

Note that for each worker, the costs and benefits of teaching/learning depend also on his clique-mate's teaching/learning decision. Conveniently, clique-mates' choices are strategic complements: if a worker teaches (resp. learns), then his clique-mate's benefit from teaching (resp. learning) increases.[15] Consequently, each worker anticipates that his clique-mate will

15. Why might worker w_1's and his clique-mate w_2's teaching choices be strategic complements? If worker w_1 chooses to teach, then he becomes less likely to be available for joint production with worker w_2, which in turn reduces the opportunity cost of teaching for worker w_2. On the other hand, worker w_2's $t = 3$ payoff increases from teaching—remember that worker w_2 increases his $t = 3$ scale by one if he teaches—is unchanged by worker w_1's teaching decision, because worker w_2's payoff is *linearly* increasing in scale. Combining these two observations, worker w_2 gains more (on net) from teaching if worker w_1 also teaches than if worker w_1 does nothing. Similarly, consider the complementarity between worker w_3's and worker w_4's learning choices. If worker w_3 learns, then the opportunity cost of learning for worker w_4 is reduced. Further, worker w_4 benefits more from learning if worker w_3 also learns: if worker w_3 does nothing, then worker w_4 experiences a + 1 increase in scale from learning (in

make the same choice as him on the equilibrium path. So a worker in a 2-clique prefers to teach (anticipating that his clique-mate will subsequently also teach) rather than do nothing if and only if his *teaching incentive constraint* is satisfied:

$$2\frac{\delta}{1-\delta} > 1-(1-\alpha_{\text{teach}})^2. \tag{6}$$

The right-hand side of (6) is the worker's $t = 3$ gain if he and his clique-mate both teach rather than do nothing. The left-hand side of (6) is the corresponding $t = 2$ loss; that is the opportunity cost of teaching. Analogously, a worker in a 2-clique prefers to learn from the other 2-clique (anticipating that his clique-mate will subsequently also learn) rather than do nothing if and only if his *learning incentive constraint* is satisfied:

$$2\frac{\delta}{1-\delta} > 1-(1-\alpha_{\text{learn}})^2. \tag{7}$$

Note that the opportunity cost of learning is higher than the opportunity cost of teaching, that is, $1-(1-\alpha_{\text{learn}})^2 > 1-(1-\alpha_{\text{teach}})^2$, so the learning incentive constraint is tighter than the teaching incentive constraint. Thus, one of the 2-cliques will teach the other 2-clique if and only if equation (7) holds.[16] Otherwise, no teaching or learning will take place.

Slow growth: At the start of $t = 2$, there is an uneven fragmentation pattern consisting of one 2-clique and two 1-cliques. Unfragmented knowledge is achieved if and only if the 2-cliquers teach the 1-cliquers. The 1-cliques' workers are unproductive and incur zero opportunity costs, thus are always willing to learn from the 2-clique. Meanwhile, the workers in the 2-clique face identical teaching incentive constraints as their counterparts under fast growth. Consequently, long-run knowledge is unfragmented if and only if Equation (6) is satisfied.

Comparing slow versus fast growth: at $t = 2$, slow growth and fast growth induce identical teaching incentive constraints, but fast growth induces tighter learning incentive constraints than fast growth. If the learning incentive constraints (under fast growth) bind but the teaching incentive constraints do not, then the two growth paths produce different knowledge sharing outcomes: specifically, unfragmented knowledge is

doing so he moves from $\{w_3, w_4\}$ to $\{w_1, w_2, w_4\}$), whereas if worker w_3 learns from worker w_1, then then worker w_4 experiences a $+3$ increase in scale from learning (in doing so he moves from $\{w_4\}$ to $\{w_1, w_2, w_3, w_4\}$).

16. The discussion so far ignores the possibility that worker w_i might reject worker w_j's offer because w_i anticipates that he will subsequently have the opportunity to offer to teach worker w_j. But in that case, we will be back in exactly the same situation as presented above, except with the identities of workers permuted (i.e. worker w_i is relabeled as worker w_j, etc.). Thus, our analysis is without loss of generality.

achieved only under slow growth. This is indeed the case for intermediate discount factors.

Put another way: incentives for learning are strong under uneven fragmentation patterns with large and small cliques coexisting, and weak under even fragmentation patterns that consist solely of large cliques. Slow growth produces less fragmentation than fast growth because it generates a more uneven fragmentation pattern, and thus stronger learning incentives, at $t = 2$. We will return to this insight, in a setting with more than four workers and richer fragmentation patterns, in Proposition 6 of Section 6.

As promised, I now return to $t = 1$ and discuss why every early hire either teaches or learns at $t = 1$. Consider what happens to a worker w_i who does nothing at $t = 1$ instead of teaching or learning. In the short term, he forgoes the payoffs he would receive from being in a 2-clique at $t = 2$. On the other hand, there are potential benefits to doing nothing under fast growth: by doing nothing, w_i induces an uneven distribution of clique sizes in period two ($\Gamma_2 = \{2, 1, 1\}$ instead of $\Gamma_2 = \{2, 2\}$), which may result in larger clique sizes—and thus higher payoffs—in the last period. It turns out that the short-term loss from doing nothing always outweighs the long-run gain, so each worker always prefers to teach/learn rather than do nothing at $t = 1$. To see why, note that doing nothing (and inducing $\Gamma_2 = \{2, 1, 1\}$) rather than teaching/learning (and inducing $\Gamma_2 = \{2, 2\}$) increases long-run payoffs only if knowledge sharing stalls when $\Gamma_2 = \{2, 2\}$; this requires that the opportunity cost of learning as part of a 2-clique outweighs the long-run benefit of increased $t = 3$ clique size (which is exactly the long-run gain from doing nothing at $t = 1$). But this opportunity cost is (by definition) strictly less than the payoff from being in a 2-clique at $t = 2$; which in turn is precisely the short-term loss from doing nothing at $t = 1$. Therefore, this short-term loss must exceed the long-run gain from doing nothing at $t = 1$.

To sum up the discussion so far, and describe the intuition for Proposition 1: fast growth induces long-run fragmentation because, when an organization grows so rapidly that new hires cannot assimilate into the organization's "incumbent" method, the new hires choose to form new cliques rather than remain unproductive. Further sharing then stalls once the organization is fragmented into multiple productive cliques. On the other hand, if knowledge diffusion can keep up with the organization's expansion, every new hire assimilates into the incumbent clique, so other large cliques do not form and long-run fragmentation is avoided.

Remark 1. Proposition 1 was derived under the assumption that $\alpha_{\text{teach}} < \alpha_{\text{learn}}$. In an alternative setting where teaching is more costly than learning ($\alpha_{\text{teach}} > \alpha_{\text{learn}}$), the long-run ($t = 3$) fragmentation pattern is independent of the growth path. Why? Start by recalling that at $t = 2$, the teaching incentive constraint (6) is the same across growth paths: it is determined by the size of the largest cliques, which is where the potential teachers reside. On the other hand, the learning incentive constraint depends on the disparity in clique size, and is tighter under fast growth.

When $\alpha_{\text{teach}} < \alpha_{\text{learn}}$, the learning constraint sometimes limits the extent of knowledge sharing. Consequently, differences in the learning constraints across growth paths drive the differences in knowledge sharing outcomes.

In contrast, when $\alpha_{\text{teach}} > \alpha_{\text{learn}}$, the teaching constraint is always tighter than the learning constraint, and solely determines the extent of knowledge sharing. Slow and fast growth generate the same long-run fragmentation pattern because they induce the same maximum clique size ($\equiv 2$), and thus the same teaching incentive constraints, at $t = 2$. In fact, long-run knowledge is unfragmented if and only if the teaching constraint (6) is satisfied.

3.1 Inefficient Fragmentation

Because per-worker payoffs are increasing in clique size, increased fragmentation results in lower total production. More precisely, given two partitions over a set of workers W and the corresponding fragmentation patterns Γ and Γ', suppose that Γ is more fragmented than Γ' in the sense that each clique corresponding to an element of Γ is a subset of some clique corresponding to an element of Γ' (e.g. $\Gamma = \{2, 2\}$ is more fragmented than $\Gamma' = \{4\}$). Then, holding fixed the teaching and learning decisions of workers, total stage payoffs are higher under Γ' than under Γ.

For each growth path, define the *first-best* outcome to be the sequence of knowledge sharing decisions that maximize total value $\mathbb{E}[\sum_{t=1}^{3} \sum_{i \in W_t} \Delta_t \pi_{i,t}]$ for that growth path. Then, holding fixed the growth path, the long-run pattern Γ_3 is more fragmented (and thus less productive) in equilibrium than in the first-best:

Proposition 2. For each growth path $g \in \{f, s\}$, there exist thresholds $0 < \lambda'(g) < \lambda(g) < 1$ such that

- For low discount factors ($\frac{\delta}{1-\delta} < \lambda'(g)$), the long-run pattern is fragmented both in equilibrium and under the first-best: $\Gamma_3^* = \Gamma_3^{FB} = \{2, 2\}$.
- For intermediate discount factors ($\lambda'(g) < \frac{\delta}{1-\delta} < \lambda(g)$), the long-run pattern is fragmented in equilibrium but unfragmented under the first-best: $\Gamma_3^* = \{2, 2\}$, $\Gamma_3^{FB} = \{4\}$.
- For high discount factors ($\frac{\delta}{1-\delta} > \lambda(g)$), the long-run pattern is unfragmented both in equilibrium and under the first-best: $\Gamma_3^* = \Gamma_3^{FB} = \{4\}$.

Loosely speaking, individuals do not fully internalize the social benefits of knowledge sharing, and do not share equally in the opportunity costs of these actions. Consequently, knowledge sharing that would confer a net social benefit may not be not undertaken by the individuals involved, and workers end up more fragmented and less productive than in the first-best.

4. Persistence of Organizational Knowledge

Organizational knowledge is a key organizational resource, and may have persistent effects on productivity (e.g. Stinchcombe 1965; Schein 1985). In this section, I consider an organization that starts off with unique and valuable organizational knowledge, and study how its growth path may affect its ability to retain this knowledge over the long-run. The analysis will highlight an additional drawback of fast growth: fast growth may result in the abandonment of superior knowledge.

For this analysis, I modify the basic model to introduce differences in productivity between methods. The payoff of a worker in a clique C with method m now depends not just on the usual factors, but also on the *quality* $q(m)$ of his clique's method:

$$\pi_{i,t} = q(m)A_i \sum_{j \in C \setminus i} A_j. \tag{8}$$

The following setup captures the premise that organizational knowledge is not easily replicated by outsiders or newcomers. One early hire, who I call the *originator*, is endowed with an *original* method with quality $q^o > 1$. Every other worker is endowed with a distinct *generic* method with quality $q = 1$. To interpret this setup, think of the originator as being with the organization before the start of the game, having acquired or created organization-specific knowledge (in the form of the original method) at some point. On the other hand, all other workers do not have any useful organization-specific knowledge when they join the organization; the premise is that, unlike the original method, a generic method can be easily created without any special knowledge or expertise.[17]

As an initial observation, note that the logic of Proposition 1 continues to hold in this setting: there is more long-run fragmentation under fast growth than under slow growth.

Proposition 3a.

- If $\frac{\delta}{1-\delta} < \frac{\tilde{\alpha}_{\text{teach}}}{2}$, then the long-run pattern is fragmented ($\Gamma_3^* = \{2, 2\}$) under both growth paths.
- If $\frac{\tilde{\alpha}_{\text{teach}}}{2} < \frac{\delta}{1-\delta} < \max\left\{\frac{\tilde{\alpha}_{\text{teach}}}{2}, \frac{\tilde{\alpha}_{\text{learn}}}{3q^o-1}\right\}$, then the long-run pattern is fragmented ($\Gamma_3^* = \{2, 2\}$) under fast growth but unfragmented ($\Gamma_3^* = \{4\}$) under slow growth.
- If $\frac{\delta}{1-\delta} > \max\left\{\frac{\tilde{\alpha}_{\text{teach}}}{2}, \frac{\tilde{\alpha}_{\text{learn}}}{3q^o-1}\right\}$, then the long-run pattern is unfragmented ($\Gamma_3^* = \{4\}$) under both growth paths.

17. The case $q^o < 1$, where the original method is inferior to generic methods, does not capture the premise of valuable organization-specific knowledge. Further, the analysis turns out to be uninteresting: the originator would immediately learn another early hire's generic method at $t = 1$, so that the original method is abandoned by the start of period two.

The main result of this Section is that even if unfragmented knowledge is developed in the long-run, slow growth ensures that the original method dominates, whereas fast growth may result in the original method being abandoned.

Proposition 3b. (Illustrated in Figures 4 and 5)

- Under slow growth, if there is unfragmented knowledge of some method m in the long-run, then m is the original method.
- Under fast growth, if

$$\frac{\delta}{1-\delta} > \max\left\{\frac{\tilde{\alpha}_{\text{teach}}}{2}, \frac{\tilde{\alpha}_{\text{learn}}}{3q^o - 1}\right\} \text{ and}$$

$$q^o - 1 < \min\left\{\frac{2\delta - (1-\delta)\tilde{\alpha}_{\text{learn}}}{\delta + (1-\delta)\tilde{\alpha}_{\text{learn}}}, \frac{1-\delta}{3\delta}(\tilde{\alpha}_{\text{learn}} - \tilde{\alpha}_{\text{teach}})\right\},$$

  then (a) there is unfragmented knowledge of a method m in the long-run, and (b) m is either generic or original, with equal probability 1/2 of either outcome.

Slow growth ensures that the original method is dominant in each period: the early hires develop shared knowledge of the original method at $t = 1$, then teach it to the late hires at $t = 2$. On the other hand, fast growth produces a "culture clash" between the two 2-cliques—one original, one generic—at $t = 2$. If q^o is not too large, so that learning the original method is not too valuable, then both clique may prefer to teach rather than learn. Either clique may end up teaching, depending on the random order in which workers move at $t = 2$. If the generic clique teaches the original clique, then the original method is abandoned.

This result resonates with the Hambrick and Crozier (1985) observation that when firms grow too rapidly,

> . . . the employee and managerial populations are no longer as homogeneous as they once were. In many cases, the core vision has been blurred or even lost . . ..

The premise of Proposition 3b is that in an organization with equal-sized cliques competing for dominance, no individual or subgroup has an advantageous bargaining position ex ante, so either clique may come out on top in the long-run. (This is driven by the assumption that players move in random order in each period.) One notable exception to this premise is that some organizations are dominated by a highly influential or charismatic founder, who often also is the source of valuable organizational knowledge. If we model an influential founder as an originator who has all the bargaining power (e.g. we may assume that the originator always moves last in the first decision round of $t = 2$), then his clique (which knows the original method) always teaches the other clique at $t = 2$.

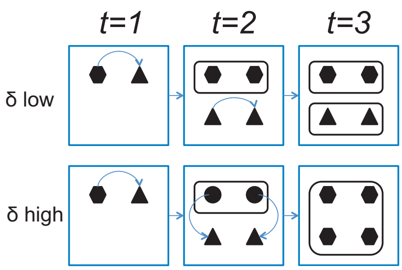

Figure 4. Knowledge Sharing under Slow Growth (hexagon = original method; triangle = generic method).

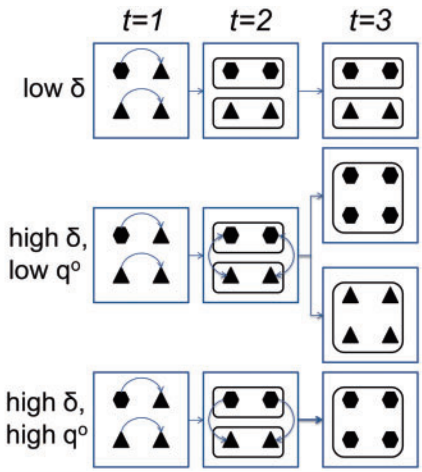

Figure 5. Knowledge Sharing under Fast Growth.

Consequently, whenever knowledge is unfragmented in the long-run, it will always be of the original method, even under fast growth. In other words, an influential founder may be able to induce organization-wide adoption of his method.[18]

5. Optimal Growth

This section considers the observed growth choices of organizations. In particular, we address why some successful firms make a deliberate choice to (a) grow slowly, and to (b) avoid acquiring other firms.

18. For example, the strong culture of organizations such as Hewlett-Packard and Disney was often attributed to the presence of charismatic founders. In fact, these organizations struggled with cultural conflict after the departure of their founders.

For example, Southwest Airlines had (until the late 1990s) a policy of conservative expansion into new cities. Historically, Southwest targeted a growth rate of 8–9%, despite the huge demand for Southwest flights on many routes. Further, Southwest expanded *organically*, by hiring new employees instead of acquiring other airlines.[19] In contrast, many of Southwest's competitors pursued more aggressive and acquisitive growth strategies.

I define the optimal growth path as the growth path that maximizes total expected output

$$E\left[\sum_{t=1}^{3} \Delta_t \left(\sum_{i \in W_t} \pi_{i,t}\right)\right]. \tag{9}$$

We may think of the growth path as being chosen at the start of the game by a principal who does not participate directly in production but seeks to maximize total output. In Section 5.1, I consider the choice between fast and slow growth paths. In Section 5.2, I consider the choice between acquisitive and organic growth.

5.1 Growth Rates

Start with the optimal *rate* of growth. I will focus on the idea that differences in growth rates are driven by differences in early organizational knowledge.[20] Retain the setting of Section 4, so that all workers except the originators are endowed with distinct generic methods, and the organization's initial state is characterized by the productivity q^o of the originator's method. This allows us to study how the optimal growth path depends on q^o; we may think of q^o as the initial quality of the organization.

In choosing the optimal growth rate, the key trade-off is between short-run and long-run output. The slow growth path sacrifices short-run ($t = 2$) output relative to fast growth; the two late hires are completely unproductive at $t = 2$ under slow growth, while all four workers are productive under fast growth. On the other hand, slow growth provides better incentives (for the late hires) to learn, potentially resulting in higher $t = 3$ output.

The main result of this section is that the relationship between an organization's quality q^o and its optimal growth rate may be nonmonotone.

19. However, Southwest seems to have changed its growth strategy recently: it announced an acquisition of AirTran Airways in late 2010.

20. An alternative approach might be to look at how differences in growth rate are driven by differences in the discount factor of the firm's principal. In brief, the answer is that a more patient principal will choose a slower growth rate, because he cares more about developing shared knowledge in the long-run. In this article, I choose to focus instead on differences in organizational knowledge as a source of heterogeneity, partly because I consider this analysis to produce more interesting insights, and partly because variation in organizational discount factors seems harder to motivate empirically.

Fast growth is optimal for both low-quality and high-quality organizations, albeit for different reasons; while slow growth is optimal for organizations of intermediate quality.

Proposition 4. (Illustrated in Figures 6 and 7) For each set of parameter values ($\alpha_{\text{teach}}, \alpha_{\text{learn}}, \delta$), there exist thresholds $1 \leqslant \hat{q}_1 \leqslant \hat{q}_2 \leqslant \infty$ (with strict inequality for some parameter values) such that

- For $1 < q^o < \hat{q}_1$, fast growth is optimal. If $0 < \frac{\delta}{1-\delta} < \frac{\tilde{\alpha}_{\text{learn}}}{2}$, the long-run pattern is fragmented. If $\frac{\delta}{1-\delta} > \frac{\tilde{\alpha}_{\text{learn}}}{2}$, long-run pattern is unfragmented, but sometimes all workers know a generic method.
- For $\hat{q}_1 < q^o < \hat{q}_2$, slow growth is optimal. The long-run pattern is unfragmented and all workers always know the original method.
- For $q^o > \hat{q}_2$, fast growth is optimal. The long-run state is unfragmented and all workers always know the original method.

For high firm quality ($q^o > \hat{q}_2$), incentives to learn the original method are strong enough that unfragmented long-run knowledge of the original method is developed under both growth paths, so that both growth paths produce identical long-run output; consequently, P chooses fast growth because it maximizes short-run output.

For $q^o < \hat{q}_2$, the short-run versus long-run trade-off becomes relevant: slow growth induces strictly higher long-run output, while fast growth induces strictly higher short-run output. Crucially, within this region, an increase in q^o shifts the trade-off in favor of slow growth. All production under slow growth utilizes the more-productive original method, whereas production under fast growth is split between the original method and a generic method.[21] Consequently, relative to fast growth, slow growth becomes more productive as q^o increases, so (a) for low firm quality, fast growth, whereby the organization tolerates some use of generic methods, is optimal; (b) for medium firm quality, slow growth is optimal because it maximizes adoption of the original method.

Our analysis suggests that deliberately constraining growth rates to develop shared knowledge may be optimal only if the organization's early knowledge is sufficiently unique and productive. This may explain why Southwest Airlines pursued a conservative growth strategy relative to its competitors. Southwest was historically known for a unique cultural emphasis on customer service that is regarded as a key source of competitive advantage (Gittell 2003), and thus had more to gain from maintaining and fostering its shared culture; whereas competitors with less productive

21. Long-run knowledge is fragmented between a generic method and the original method for low δ, while long-run knowledge is randomly of either a generic method or the original method for high δ.

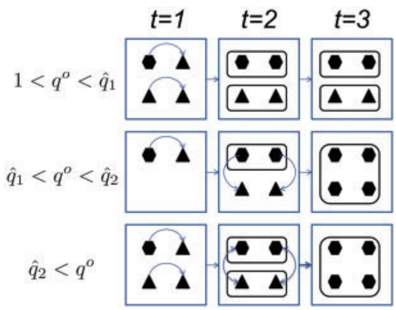

Figure 6. Optimal Growth Path for Low δ ($0 < \frac{\delta}{1-\delta} < \frac{\tilde{\alpha}_{\text{learn}}}{2}$).

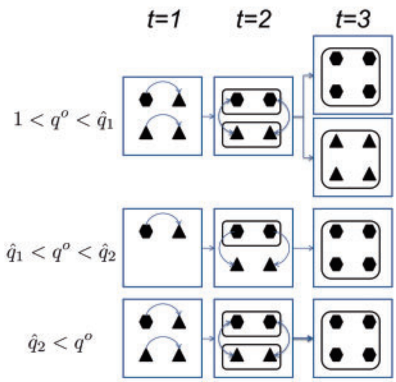

Figure 7. Optimal Growth Path for High δ ($\frac{\delta}{1-\delta} > \frac{\tilde{\alpha}_{\text{learn}}}{2}$).

cultures were more willing to tolerate cultural fragmentation or dilution,[22] and thus pursued faster growth.

5.2 Acquisitive versus Organic Growth

In the analysis so far, each worker joins the organization endowed with a distinct method; this reflects the premise of *organic* growth, whereby new workers are hired separately, and thus do not have prior shared knowledge. This section considers acquisitive growth, whereby an organization grows by acquiring other firms. Workers from the same acquired firm may join the organization with shared knowledge, as they have prior shared experience from working together. To model acquisitive growth, consider a slow growth path where the two late hires are endowed with *shared* knowledge of a method.

Acquisitive growth involves the same trade-offs as fast growth: it increases payoffs in the short run but may stymie subsequent knowledge sharing, resulting in lower long-run payoffs. To highlight this point, we

22. One example of fast growth leading to fragmentation is People Express, an airline which encountered persistent problems coordinating ticketing, luggage and customer service after an aggressive growth and acquisition strategy. See Beer (1990) for more details.

return to the basic setting where all methods have the same quality: $q \equiv 1$. The following proposition then states that acquisitive growth produces outcomes identical to fast organic growth.

Proposition 5. In each of the second and third periods, fast organic growth and acquisitive growth produce the same fragmentation patterns (up to a permutation of worker identities), and the same total expected output.

Acquisitive growth and fast organic growth both induce the same $t = 2$ fragmentation pattern $\Gamma_2^* = \{2, 2\}$ (Figure 8). Thus, both growth paths generate the same $t = 2$ knowledge sharing incentives and the same $t = 3$ patterns.

Proposition 5 may explain why organizations that deliberately grow slowly to develop shared knowledge also favor organic growth over acquisitive growth. They do so to avoid either long-run fragmentation or the abandonment of their organizational knowledge (as in Proposition 4).[23]

6. Many Workers

So far, the analysis has been performed in a stylized setting with $\leqslant 4$ players and three periods. I now modify the model from Section 2 to accommodate more workers and more periods. The goal is to reproduce the main insights from Section 3: that knowledge sharing takes place when fragmentation patterns are uneven, and that fast growth results in persistent fragmentation.

Time is discrete and infinite ($t = 1, 2, 3, \ldots$). Workers are myopic ($\delta = 0$). Unlike Section 2, workers do not trade off today's costs versus tomorrow's benefits in making their teaching and learning decisions; instead, both costs and benefits are realized in the current period. Specifically, each period now consists of two sub-periods. Worker i's total stage payoff is the weighted sum of sub-period payoffs: $(1 - \lambda)\pi_{i,t;1} + \lambda\pi_{i,t;2}$, where $\pi_{i,t;k}$ is his production payoff from sub-period k and $\lambda \in (0, 1)$ is a weighting parameter. The production function for each sub-period is unchanged from Section 2: worker i's unweighted

23. This emphasis on organic growth is often observed in organizations for which corporate culture is a competitive advantage. Historically, Nordstrom expanded by opening new stores and carefully training new hires rather than acquiring other department store chains (see, e.g., Spector and McCarthy 2005); in contrast, competitors like Macy's aggressively acquired other major chains. Some elite law firms and consulting firms also deliberately restrict themselves to organic growth. For example, the law firm Knobbe, Martens, Olsen and Bear, which is known for its cohesive firm culture, mostly hires new associates who are fresh out of law school. It rarely hires established partners from other firms; such lateral hires would often bring along a team of associates that the partner nurtured at his previous firm. In particular, it seems that Knobbe's growth strategy was intended to preserve its shared culture. In a 2008 interview, its managing partner stated that "we want to stick to our core values and grow internally."

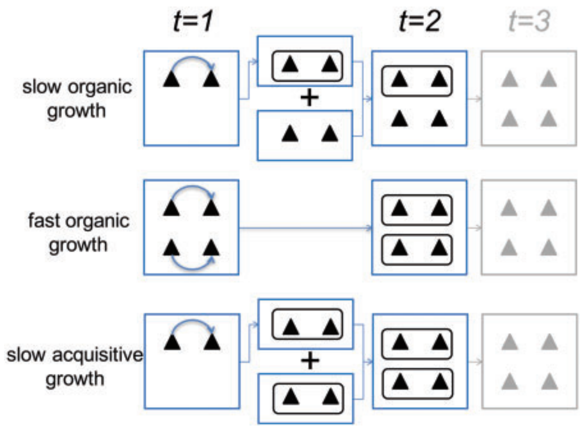

Figure 8. Fragmentation Patterns at $t = 2$.

payoff in subperiod k is

$$\pi_{i,t;k} \equiv A_{i,k} \sum_{j \in C \setminus i} A_{j,k},$$

where $A_{i,k}$ is w_i's availability, and C is w_i's clique, in subperiod k. In the first sub-period, the full stage game from Section 2 is played. Workers' methods are updated at the end of the first sub-period based on their teaching and learning choices. In the second sub-period, a second round of joint production takes place using the updated methods, but workers do not teach or learn and are always available. So, no opportunity costs are incurred and methods are not further updated in that sub-period.

Modify the knowledge sharing process from Section 2 as follows: in the first sub-period, only one worker per clique can *teach*. So, if a worker has an offer accepted, then his clique-mates are ineligible to make offers for the rest of the period. This constraint ensures that cliques expand "smoothly", by no more than one worker per period.[24]

Define a growth path with *growth rate* $\gamma > 0$ and *ceiling* $N \geq 1$ as follows: the organization adds workers at the start of each period, so that in period $t \geq 1$, it has $\min\{\lfloor\gamma(t+1)\rfloor, N\}$ workers. In other words, the organization initially experiences a growth phase where it add $\lfloor\gamma(t+1)\rfloor - \lfloor\gamma t\rfloor \approx \gamma$ workers per period, then stops growing in period $T = \lceil N/\gamma - 1\rceil$

24. In contrast, under the assumption from Section 2 that all workers in a clique can simultaneously teach, cliques can double in size each period. Such dramatic period-by-period jumps in clique size result in complicated dynamics that are difficult to characterize. I regard such problems as an artifact of the discrete-time setting rather than a robust feature of the economic setting. Such difficulties are mostly avoided in the current setting by ensuring that cliques grow gradually, by no more than one worker per period. One potential interpretation of this assumption is that opportunities to engage in learning-by-doing (such as apprenticeships and internships) remain limited even within large groups.

when it reaches long-run size N. As before, in each period, new workers join at the start of the period endowed with a distinct method.

I focus on the case where teaching is costless and learning is maximally costly ($\alpha_{\text{teach}} = 0$ and $\alpha_{\text{learn}} = 1$). Here, teaching costs do not impose any constraints on clique expansion. This allows for clean exposition of the results: as discussed in Remark 1, learning rather than teaching costs drive the effect of growth rates on fragmentation.

What fragmentation patterns are conducive to knowledge sharing? A key step toward Proposition 1 was the insight that incentives for knowledge sharing are stronger when fragmentation patterns are relatively uneven. The following proposition presents a simple version of this point.

Proposition 6. Let $\alpha_{\text{teach}} = 0$ and $\alpha_{\text{learn}} = 1$. At the start of period t, let the size of the largest clique be n, and let the size of the smallest clique be $m \leqslant n$. Then knowledge sharing takes place in period t if and only if there are at least two cliques, and

$$\lambda > \frac{m-1}{n}. \tag{10}$$

Proposition 6 states that *some* knowledge sharing takes place if and only if workers in the *smallest* clique are willing to learn from workers in the *largest* clique. A worker in an m-clique who learns from a worker in an n-clique faces an opportunity cost of $(1-\lambda)(m-1)$ from the first sub-period, and gains $\lambda(n-m+1)$ in the second sub-period. Comparing costs and benefits, we get Equation (10). Clearly, incentives to learn are large when potential learners (teachers) are in small (large) cliques, so that (a) the opportunity costs of learning are small, and (b) the gains from learning are large. On the other hand, a fragmentation pattern consisting of multiple similarly-sized cliques is stable if the cliques are sufficiently large. In the special case where all cliques are of equal size k, Equation (10) gives us the stability condition $k > \frac{1}{1-\lambda}$.

Proposition 6 does not tell us "how much" knowledge sharing takes place. In fact, there may be nontrivial equilibrium multiplicity, for two reasons. First, equilibria may be supported by off-path punishments which condition on history in arbitrary ways. Second, potential teachers are indifferent regarding the identity of the learner, so different equilibria may simply correspond to different choices by workers about whom to teach.

To pin down the model's dynamics and precisely characterize long-run outcomes, I introduce some equilibrium selection rules. To eliminate history dependence, I restrict attention to pure-strategy Markov perfect equilibria.[25] I also impose the following tiebreaking rules, which produce

25. This restriction has the following bite. Given a history up to a node where worker w_i takes his turn, consider two alternative events: (a) w_i does not make an offer, and (b) w_i makes

"greedy" dynamics whereby the largest cliques preferentially teach the smallest cliques. If a worker is otherwise indifferent between accepting or rejecting an offer, then he accepts the offer. Similarly, if a worker is otherwise indifferent at his turn between not making any offer and making a offer that will be accepted on the continuation path (call this a *meaningful* offer), then he makes the offer. Finally, if a worker is otherwise indifferent at his turn between multiple meaningful offers, then he tie-breaks in favor of the offer to the worker in the smallest clique at the start of the period (among all his meaningful targets).

We now turn to the effect of growth rate on long-run fragmentation. In stating the result, two observations are convenient. First, equilibrium dynamics are deterministic: for each growth path, the fragmentation pattern (up to permutations of workers) in each period is completely determined by the equilibrium selection rules. Second, for each growth path, knowledge sharing ceases in finite time, after which the same fragmentation pattern persists forever. The number of *cliques* in this *long-run* pattern is a natural measure of the degree of persistent fragmentation. The following proposition is analogous to Proposition 1. It states that fast growth results in long-run fragmentation, in the form of multiple cliques of similar size, given that the ceiling N is sufficiently high relative to the growth rate γ. On the other hand, long-run fragmentation is avoided under sufficiently low growth rates.

Proposition 7. Let $\alpha_{\text{teach}} = 0$ and $\alpha_{\text{learn}} = 1$, and suppose $N > \frac{2\lceil\gamma\rceil}{1-\lambda}$. Then the number of cliques in the long-run pattern is weakly increasing in γ. For integer γ, the number of long-run cliques exactly equals γ. For $\gamma \leqslant 1$, there is exactly one long-run clique. Further, if there is more than one long-run clique, then the difference in size between the largest and the second-smallest long-run clique is at most one.

The special case where γ and N/γ are integers highlights the essential intuition. At $t = 1$, the organization starts with exactly 2γ workers who immediately pair up, resulting in γ cliques of two at the end of the period.[26] Further, during the growth phase, each clique absorbs exactly one newcomer per period, so that at the end of every period t, there are γ equally-sized cliques of size $t + 1$. At the end of the growth phase, the condition $N > \frac{2\lceil\gamma\rceil}{1-\lambda}$ ensures that each clique is large enough that workers are unwilling to incur the opportunity costs of learning to join another (equally-sized) clique. Consequently, knowledge sharing stalls after the

an offer which is rejected. Note that the subgames following either event are indistinguishable. Thus, in any Markov perfect equilibrium, both events must produce identical continuation equilibria. Further, in the case where w_i's offer is rejected, the continuation equilibrium does not depend on the offer recipient's identity.

26. The assumption that the organization starts with exactly $\lfloor 2\gamma \rfloor$ workers renders the dynamics transparent, but is not essential for Proposition 7. The key assumptions are that (a) the number of new hires in each period $t \geq 2$ is approximately γ, and (b) the number of workers at $t = 1$ is small compared to $\lceil\gamma\rceil/(1 - \lambda)$ and N.

ceiling is attained and the γ equally-sized cliques persist forever. In other words, the number of long-run cliques is proportional to the growth rate γ.

The general case of Proposition 7 introduces some integer problems that complicate the analysis. For example, if γ is not an integer, then clique growth is slightly asymmetric: in the end of each period t of the growth phase, there will be $\lfloor\gamma\rfloor$ *regular* cliques of size $t + 1$, and one smaller clique of size $\lfloor\gamma(t+1)\rfloor - \lfloor\gamma\rfloor(t+1)$. After the growth phase ends, the smaller clique will persist if and only if it is large enough relative to the regular cliques that its workers are unwilling to learn; specifically, if (approximately) $\gamma - \lfloor\gamma\rfloor > \lambda$. Otherwise, the smaller clique is absorbed by the regular cliques, so only the $\lfloor\gamma\rfloor$ regular cliques persist in the long run. Figure 9 illustrates the second case, whereby a leftover clique forms during the growth phase, then is absorbed by the regular cliques after the growth phase ends.

7. Conclusion

This article shows how imperfect coordination may arise when knowledge sharing is local and costly. Within an organization, local knowledge sharing means that short-run knowledge diffusion must necessarily be incomplete. The resultant fragmentation patterns may inhibit or distort individuals' incentives for further knowledge sharing, in which case incomplete diffusion persists. In much of the analysis, such persistence arises when the organization is fragmented into multiple cliques of relatively equal size, so that workers lack sufficient incentive to further assimilate into other cliques. Persistent fragmentation becomes more likely when organizations expand too quickly and prioritize acquisitive growth over organic growth.

The article's main contribution is a tractable model of organizational knowledge that takes seriously the Penrose (1959) insights that such knowledge must be developed (a) over time and (b) across large groups of individuals. Importantly, the article seeks to take a step toward providing economic foundations for the literature on the "knowledge-based theory of the firm" (e.g. Grant 1996). Consistent with this literature, the article emphasizes the point that coordination failures within organizations may arise not because individuals refuse to coordinate (as in economic theories that involve differing beliefs or misaligned incentives), but because individuals simply *do not know how to coordinate*; it does so with a moral-hazard-in-teams model where coordination failures arise because of a reluctance by individuals to invest in shared knowledge.

Beyond the analysis in the article, the fact that local knowledge sharing plays such an important role in the logic of the model suggests that firms that share knowledge nonlocally may avoid forming undesirable patterns of coordination and thus efficiently develop shared knowledge. One way to do so may be by formalizing relevant organizational knowledge into

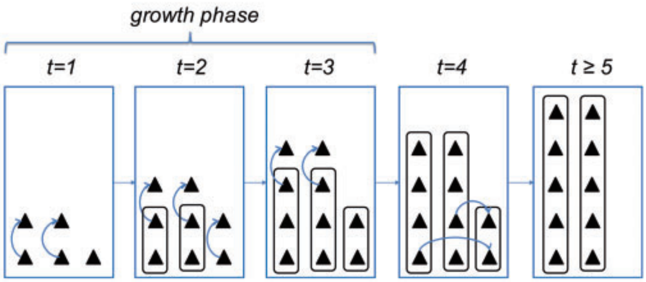

Figure 9. Long-run Fragmentation with $\gamma = 2.7$, $N = 10$, $\lambda = 0.3$.

easy-to-disseminate forms such as standard operating procedures, cookbooks, etc. An extension of the existing theory to incorporate knowledge formalization should generate useful predictions.[27] Let me briefly mention a few examples. First, firms that engage in more knowledge formalization should, ceteris paribus, grow more rapidly. Second, organizational growth rates should be higher in industries where organizational knowledge is easier to formalize. Third, in industries where returns to scale are strongly increasing, we should observe more formalization of organizational knowledge (concurrently with higher organizational growth rates).

Another potential application of this framework is to the adaptation problem. Consider an organization that faces a change in environment, such as the introduction of a disruptive technology or a change in the regulatory environment. The organization's existing shared knowledge is unsuited for the new environment; consequently, it has to adapt by developing new shared knowledge. In this setting, the premise of local knowledge sharing may produce interesting insights about the adaptation process. In particular, parallel to the logic of Proposition 1, one might expect that large organizations may inadvertently foster multiple independent attempts to develop new methods among different groups or divisions, leading to post-adaptation fragmentation in the form of multiple large cliques. On the other hand: small, growing organizations may more successfully focus on a single source of innovation and thus successfully develop shared knowledge of a new method.

An important step in developing this article's ideas is empirical: to develop measures of knowledge fragmentation within organizations, which would enable tests of the model's predictions. One possible strategy may be to use data about the structure of communication networks within organizations to tease out fragmentation patterns. In particular,

27. The implications of knowledge formalization have been explored, in other directions, within the management literature. For example, Kogut and Zander (1992), in a wide-ranging discussion of organizational knowledge, points out that knowledge formalization has the drawback that it facilitates imitation by competitors.

knowledge fragmentation may be characterized by relatively informal communication (reflecting shared tacit knowledge) within cliques, versus relatively formal communication across cliques.[28]

Funding

This work was supported by MIT Sloan's Program on Innovation in Markets and Organizations.

Appendix A Proofs

Lemma A1. Suppose that there are two 2-cliques at $t = 2$, that is, $\Gamma_2^* = \{2, 2\}$. If $\frac{\delta}{1-\delta} > \frac{\tilde{\alpha}_{\text{learn}}}{2}$ where $\tilde{\alpha}_{\text{learn}} = 1 - (1 - \alpha_{\text{learn}})^2$, then the workers from one 2-clique will teach the workers from the other 2-clique, so $\Gamma_3^* = \{4\}$. Otherwise, no teaching and learning takes place, so $\Gamma_3^* = \{2, 2\}$.

			$t = 2$ continuation value for worker w_i ($v_{i,2}$)			
	$t = 2$ step	Γ_3	Worker w_1	Worker w_2	Worker w_3	Worker w_4
#1	No sharing	$\{2, 2\}$	$1+\frac{\delta}{1-\delta}$			
#2	$w_1 \to w_3$, $w_2 \to w_4$	$\{2, 2\}$	$(1-\alpha_{\text{teach}})(1-\alpha_{\text{learn}})+\frac{\delta}{1-\delta}$			
#3	$w_1 \to w_3$	$\{3, 1\}$	$1-\alpha_{\text{teach}}+2\frac{\delta}{1-\delta}$		$1-\alpha_{\text{learn}}+2\frac{\delta}{1-\delta}$	$1-\alpha_{\text{learn}}$
#4	$w_3 \to w_1$	$\{3, 1\}$	$1-\alpha_{\text{learn}}+2\frac{\delta}{1-\delta}$	$1-\alpha_{\text{learn}}$	$1-\alpha_{\text{teach}}+2\frac{\delta}{1-\delta}$	
#5	$w_1 \to w_2$, $w_3 \to w_4$	$\{4\}$	$(1-\alpha_{\text{teach}})^2+3\frac{\delta}{1-\delta}$		$(1-\alpha_{\text{learn}})^2+3\frac{\delta}{1-\delta}$	
#6	$w_3 \to w_1$, $w_4 \to w_2$	$\{4\}$	$(1-\alpha_{\text{learn}})^2+3\frac{\delta}{1-\delta}$		$(1-\alpha_{\text{teach}})^2+3\frac{\delta}{1-\delta}$	

Proof. WLOG, $\{w_1, w_2\}$ are in one 2-clique and $\{w_3, w_4\}$ are in the other 2-clique at $t = 2$. Start by calculating and listing the $t = 2$ continuation value to each worker under each $t = 2$ diffusion step (some isomorphic diffusion steps are omitted):

28. Datasets of email communications within firms, such as the Enron email corpus (Klimt and Yang 2004; see also Palacios-Huerta and Prat 2012), may be a promising avenue for measuring the intensity and nature of communication within organizations. Tantalizingly, data about the content of emails, such as the frequency of words and phrases, may allow us to construct measures of the informality of communication.

With these calculations in hand, I verify some partial results about the learning process at $t = 2$; combined, these results prove the Lemma.

(i) Assume that at worker w_2's turn, worker w_1 has made an accepted offer to worker w_3. Then worker w_2 will make an offer to worker w_4, and worker w_4 will accept, if and only if $\frac{\delta}{1-\delta} > \alpha_{\text{teach}}(1 - \alpha_{\text{teach}})$.
(ii) $\Gamma_3^* \neq \{3, 1\}$ if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2}$.
(iii) $\Gamma_3^* = \{4\}$ if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2}$.
(iv) $\Gamma_3^* = \{2, 2\}$ if $\frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{2}$.

Step (i): First, note that worker w_2 will not accept any offer from worker w_4 if worker w_1 has already made an accepted offer to worker w_3: if worker w_2 accepts an offer from worker w_4, then outcome #2 results, while if worker w_2 neither makes not accepts any offers in the period, we get outcome #4. Worker w_2 is strictly better off under #4 than #2, so prefers not to learn from worker w_4. Next, consider an offer from worker w_2 to worker w_4 after worker w_1 has accepted an offer from worker w_3. If worker w_4 accepts worker w_2's offer, then outcome #5 results; if no offer from worker w_2 to worker w_4 is accepted, then outcome #3 results. Worker w_4 prefers #5 to #3 if and only if $\frac{\delta}{1-\delta} > \alpha_{\text{teach}}(1 - \alpha_{\text{teach}})$ while worker w_4 always prefers 5 to #3, so worker w_2 teaches worker w_4 if and only if $\frac{\delta}{1-\delta} > \alpha_{\text{teach}}(1 - \alpha_{\text{teach}})$.

Step (ii): Assume, toward a contradiction, that $\Gamma_3^* = \{3, 1\}$ and $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{learn})^2}{2}$. WLOG, $\Gamma_3^* = \{3, 1\}$ results from worker w_1 teaching worker w_3 at $t=2$. Then, by step (ii), and since $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2} > \alpha_{\text{teach}}(1 - \alpha_{\text{teach}})$, worker w_2 and worker w_4 will prefer to deviate from the equilibrium path, so that worker w_2 makes an accepted offer to worker w_4. In light of step (iv), this implies that $\Gamma_3^* = \{4\}$ when $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{learn})^2}{2}$.

Step (iii): Assume, toward a contradiction, that $\Gamma_3^* = \{2, 2\}$ and $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{learn})^2}{2}$. WLOG, let worker w_1 be the last worker in the first round; by assumption, worker w_1 does not make an accepted offer (outcome #1). Consider a deviation from the equilibrium path where worker w_1 makes an accepted offer to worker w_3; then (from step (ii), and since $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2} > \alpha_{\text{teach}}(1 - \alpha_{\text{teach}})$), another round will be played wherein worker w_2 will make an offer to worker w_4, who will accept (outcome 5). Since $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2}$, every worker prefers outcome 5 to #1; so worker w_1 and worker w_3 strictly prefer to deviate. This contradicts the claim that worker w_1 does nothing at his turn on the equilibrium path.

Step (iv): See step (i) from the proof of Lemma A2. □

Lemma A2. Suppose that $\Gamma_2^* = \{2, 1, 1\}$. If $\frac{\delta}{1-\delta} > \frac{\tilde{\alpha}_{\text{teach}}}{2}$ where $\tilde{\alpha}_{\text{teach}} = 1 -(1 - \alpha_{\text{teach}})^2$, then each 2-cliquer teaches a 1-cliquer, resulting in $\Gamma_3^* = \{4\}$. Otherwise, one 1-cliquer teaches the other 1-cliquer while the 2-cliquers do nothing, resulting in $\Gamma_3^* = \{2, 2\}$.

Proof. Assume, without loss of generality, that at $t = 2$, the 2-clique is $\{w_1, w_2\}$ and the 1-cliquers are worker w_3 and worker w_4. I proceed in the following steps:

(i) Show that worker w_1 and worker w_2 will not accept any offers from worker w_3 or worker w_4 on the equilibrium path.

(ii) Show that on the equilibrium path, worker w_3 and worker w_4 (who are each in a 1-clique) will accept any offer made by worker w_1 and worker w_2 (who are in the 2-clique); also, if worker w_3 (or worker w_4) anticipates that he will receive an offer from worker w_1 or worker w_2, then he will neither make or accept an offer to or from worker w_4.

(iii) Show that worker w_1 and worker w_2 will each make an offer to worker w_3 and worker w_4 (which will be accepted) if and only if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{teach}})^2}{2}$.

(iv) Show that worker w_1 and worker w_2 will make no offers, while (without loss of generality) worker w_3 will make an offer to worker w_4 (which will be accepted), if and only if $\frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{teach}})^2}{2}$. The claim of the lemma then follows from the combination of the intermediate results.

Some notation: let $1_{\{Z\}}$ be the indicator function that equals unity if event Z occurs and equals zero otherwise. Also, let $|C(i, t)|$ be the size of worker w_i's clique at period t.

Step (i): Consider the choices of each 2-cliquer (WLOG, worker w_1). First, if neither worker w_1 nor worker w_2 learn from worker w_3 or worker w_4 at $t = 2$, then worker w_1's $t = 2$ continuation value is weakly greater than $\left(1 - \alpha_{\text{teach}} \cdot 1_{\{B \text{ teaches } C \text{ or } D\}}\right) + \left(1 + 1_{\{B \text{ teaches } C \text{ or } D\}}\right)\frac{\delta}{1-\delta}$, which is his continuation value *if he does nothing at* $t = 2$. On the other hand, if worker w_1 learns from worker w_3 or worker w_4 at $t = 2$, then his $t = 2$ continuation value is $(1 - \alpha_{\text{learn}})\left(1 - \alpha_{\text{teach}} \cdot 1_{\{B \text{ makes accepted offer}\}}\right) + \frac{\delta}{1-\delta}$. The latter is strictly smaller than the former, so worker w_1 will not learn from worker w_3 or worker w_4 at $t = 2$ if he anticipates that worker w_2 will not learn from worker w_3 or worker w_4 at $t = 2$; similarly, worker w_2 will not learn from worker w_3 or worker w_4 if he anticipates that worker w_1 will not learn from worker w_3 or worker w_4. Applying an induction argument to the statement "Whenever worker w_1 (or worker w_2) receives an offer to learn from worker w_3 or worker w_4, he rejects the offer and anticipates correctly that any subsequent offers from worker w_3 or worker w_4 to worker w_1 or worker w_2 will be rejected" then establishes the claim.

Step (ii): Consider, WLOG, an offer from worker w_1 to worker w_3. Note from step (i) that worker w_1 and worker w_2 never accept any offers. The continuation value to worker w_3, besides depending on whether he accepts the offer, also depends on whether worker w_4 accepts an offer from worker w_2 (either before or after the offer from worker w_1 to worker w_3), and on whether worker w_3 accepts an offer from worker w_4 (or vice versa). If worker w_3 accepts the offer from worker w_1, then his continuation value is $0+\left(2+1_{\{D\text{ accepts offer from }B\}}\right)\frac{\delta}{1-\delta}$, while if worker w_3 rejects the offer from worker w_1, then his continuation value is $0+1_{\{C/D\text{ accepts offer from }D/C\}}\frac{\delta}{1-\delta}$. We can verify that worker w_3's continuation value from accepting is higher, regardless of worker w_4's actions. Thus implies that worker w_3 always accepts an offer that worker w_1 makes, assuming that worker w_2 had not previously accepted an offer. By the same reasoning, if worker w_3 (or worker w_4) anticipates that he will receive an offer from worker w_1 or worker w_2, then he will neither make or accept an offer to or from worker w_4.

Step (iii): Without loss of generality, assume that at worker w_1's turn, worker w_2 made an accepted offer to worker w_3 (or worker w_4). Then worker w_1's continuation value from teaching worker w_4 (or worker w_3) equals$(1-\alpha_{\text{teach}})^2+3\frac{\delta}{1-\delta}$, while his continuation value from doing nothing equals $(1-\alpha_{\text{teach}})+2\frac{\delta}{1-\delta}$. Thus, to maximize his continuation value, worker w_1 makes an offer to worker w_4 (which worker w_4 accepts) if and only if $\frac{\delta}{1-\delta}>\alpha_{\text{teach}}(1-\alpha_{\text{teach}})$.

Next, consider the case where at worker w_1's turn, worker w_2's turn has passed without worker w_2 teaching anyone; so that the $t=2$ diffusion process will end, and worker w_1 and worker w_2 will do nothing in this period, if worker w_1 does not teach worker w_3 (or worker w_4). If worker w_1 teaches worker w_3 (or worker w_4), then (from step (ii)) worker w_2 will subsequently teach worker w_4 (or worker w_3) if and only if $\frac{\delta}{1-\delta}>\alpha_{\text{teach}}(1-\alpha_{\text{teach}})$. Thus worker w_1's continuation value if he teaches worker w_3 is

$$\begin{cases}(1-\alpha_{\text{teach}})^2+3\dfrac{\delta}{1-\delta} \text{ if } \dfrac{\delta}{1-\delta}>\alpha_{\text{teach}}(1-\alpha_{\text{teach}}) \\ (1-\alpha_{\text{teach}})+2\dfrac{\delta}{1-\delta} \text{ if } \dfrac{\delta}{1-\delta}<\alpha_{\text{teach}}(1-\alpha_{\text{teach}})\end{cases};$$

while his continuation value if he does nothing is $\frac{1}{1-\delta}$. Maximizing his continuation value, worker w_1 offers to teach worker w_4 (which worker w_4 accepts), anticipating that worker w_2 will offer to teach worker w_3 (which worker w_3 accepts), if and only $\frac{\delta}{1-\delta}>\alpha_{\text{teach}}\left(1-\frac{\alpha_{\text{teach}}}{2}\right)$.

Finally, combining the two cases above, assume WLOG that worker w_2 moves before worker w_1 in the round. Then, from the above cases, we may infer that the same $t=2$ diffusion step results regardless of whether worker w_2 makes an offer to worker w_3 or worker w_4: both worker w_1 and worker w_2 will teach, resulting in $\Gamma_3=\{4\}$ if and only if $\frac{\delta}{1-\delta}>\alpha_{\text{teach}}\left(1-\frac{\alpha_{\text{teach}}}{2}\right)$.

Step (iv): If $\frac{\delta}{1-\delta}>\alpha_{\text{teach}}\left(1-\frac{\alpha_{\text{teach}}}{2}\right)$, then worker w_3 and worker w_4 both anticipate that they will not receive any offers from worker w_1 or worker w_2. If worker w_3 makes an accepted offer to worker w_4, then each of worker w_3 and worker w_4's $t=2$ continuation values equal $\frac{\delta}{1-\delta}$; while if worker w_3 and worker w_4 reject all offers from each other, their $t=2$

continuation value is zero. Thus one of worker w_3 or worker w_4 will make an accepted offer to the other. □

Lemma A3.

- Under slow growth, one early hire teaches the other at $t = 1$, so $\Gamma_2^* = \{2, 1, 1\}$.
- Under fast growth, two early hires teach and two early hires learn at $t = 1$, so $\Gamma_2^* = \{2, 2\}$.

Proof. I will use Lemmas A2 and A1 in the proof.

(i) I start by analyzing the slow growth path s. The two possible $t = 2$ fragmentation patterns are $\Gamma_2 = \{2, 1, 1\}$ and $\Gamma_2 = \{1, 1, 1, 1\}$. If $\Gamma_2 = \{2, 1, 1\}$, then by Lemma A2, worker w_1 and worker w_2 each have $t = 1$ continuation value of $\delta(1 - \alpha_{\text{teach}})^2 + 3\frac{\delta^2}{1-\delta}$ if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{teach}})^2}{2}$, while they each have $t = 1$ continuation value of $\frac{\delta}{1-\delta}$ if $\frac{\delta}{1-\delta} \leqslant \frac{1-(1-\alpha_{\text{teach}})^2}{2}$. If $\Gamma_2^* = \{1, 1, 1, 1\}$, then each of worker w_1 and worker w_2 have $t = 1$ continuation value of no more than $\frac{\delta^2}{1-\delta}$ (since they will have zero stage payoff at $t = 2$, and their maximum possible potential clique size at $t = 3$ is two). Since $\delta(1 - \alpha_{\text{teach}}) + 3\frac{\delta^2}{1-\delta} > \frac{\delta^2}{1-\delta}$ and $\delta + \frac{\delta^2}{1-\delta} > \frac{\delta^2}{1-\delta}$, worker w_1 and worker w_2 always prefer $\Gamma_2 = \{2, 1, 1\}$ to $\Gamma_2 = \{1, 1, 1, 1\}$. This immediately implies that worker w_1 will teach worker w_2 (or worker w_2 will teach worker w_1) at $t = 1$.

(iia) Next, I consider the fast growth path f. First, assume, toward a contradiction that $\Gamma_2^* = \{2, 1, 1\}$: WLOG, worker w_3 and worker w_4 do nothing while worker w_1 teaches worker w_2 (or worker w_2 teaches worker w_1) at $t = 1$. From Lemma A2, in the continuation equilibrium, worker w_3 and worker w_4 each get $t = 1$ continuation value of $\frac{\delta^2}{1-\delta}$ if $\frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{teach}})^2}{2}$, and of $3\frac{\delta^2}{1-\delta}$ if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{teach}})^2}{2}$. Compare the equilibrium to a deviation where worker w_3 makes an accepted offer to worker w_4, so that $\Gamma_2 = \{2, 2\}$. Under this deviation, worker w_3 and worker w_4 each get $t = 1$ continuation value of $\frac{\delta}{1-\delta}$ if $\frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{2}$, and of $1 - \frac{(1-\alpha_{\text{teach}})^2 + (1-\alpha_{\text{learn}})^2}{2} + 3\frac{\delta^2}{1-\delta}$ if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2}$ (in the latter case, $\Gamma_3 = \{4\}$, and at $t = 2$ it is equally likely to have worker w_1 and worker w_2 teach worker w_3 and worker w_4 or to have worker w_3 and worker w_4 teach worker w_1 and worker w_2). It is straightforward to check that worker w_3 and worker w_4 strictly prefer to deviate, contradicting our original assumption.

(iib) Now, assume towards a contradiction that $\Gamma_2^* = \{1, 1, 1, 1\}$ under fast growth f. WLOG, let worker w_4 be the last worker in the first round at $t = 1$. Then, in equilibrium, each worker's $t = 1$ continuation payoff is at most $\frac{\delta^2}{1-\delta}$. Compare this outcome to a deviation where worker w_4 makes an accepted offer to worker w_2. From (iia), in the continuation equilibrium following this deviation, each worker gets a $t = 1$ continuation value strictly greater than $\frac{\delta^2}{1-\delta}$. It follows that $\Gamma_2^* \neq \{1, 1, 1, 1\}$.

Combining (iia) and (iib), we conclude that $\Gamma_2^* = \{2, 2\}$ under f. □

I now prove Lemmas A8 and A9 via a set of intermediate results.

Lemma A4. Let $\Gamma_2^* = \{2, 1, 1\}$, and assume that the two workers in the 2-clique know the original method, while each worker w_i in the two 1-cliques knows a distinct generic method. If $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{teach}})^2}{2}$, then $\Gamma_3^* = \{4\}$. Otherwise, $\Gamma_3^* = \{2, 2\}$.

Proof. The proof is very similar to that of Lemma A2. For brevity, I omit the details, and only outline the steps of the proof. Assume, without loss of generality, that $\{1, 2\}$ form the 2-clique while worker w_3 and worker w_4 are each in a 1-clique. Then the proof proceeds in the following steps:

(i) Show that worker w_1 and worker w_2 will not accept any offers from worker w_3 or worker w_4 on the equilibrium path.

(ii) Show that on the equilibrium path, worker w_3 and worker w_4 (who are each in a 1-clique) will accept any offer made by worker w_1 or worker w_2 (who are in the 2-clique); also, if worker w_3 (or worker w_4) anticipates that he will receive an offer from worker w_1 or worker w_2, then he will neither make or accept an offer to or from worker w_4.

(iii) Show that worker w_1 and worker w_2 will each make an offer to worker w_3 and worker w_4 (which will be accepted) if and only if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{teach}})^2}{2}$.

(iv) Show that worker w_1 and worker w_2 will make no offers, while (without loss of generality) worker w_3 will make an offer to worker w_4 (which will be accepted), if and only if $\frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{teach}})^2}{2}$. The claim of the lemma then follows from the combination of the intermediate results. □

Lemma A5. Let $\Gamma_2^* = \{2, 2\}$, and assume that one 2-clique knows the original method m^{orig} while the other 2-clique knows a generic method m.

- If $\frac{\delta}{1-\delta} < \max\left\{\frac{1-(1-\alpha_{\text{teach}})^2}{2}, \frac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}\right\}$, then $\Gamma_3^* = \{2, 2\}$, and one 2-clique has shared knowledge of the original method while the other 2-clique has shared knowledge of a generic method m at $t = 3$.
- If $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{3-q^o}$ and $\frac{\delta}{1-\delta} > \frac{(1-\alpha_{\text{teach}})^2}{3(q^o-1)}$, or if $\max\left\{\frac{1-(1-\alpha_{\text{teach}})^2}{2}, \frac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}\right\} < \frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{3-q^o}$, then $\Gamma_3^* = \{4\}$, and all workers have shared knowledge of m^{orig} at $t = 3$.
- If $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{3-q^o}$ and $\frac{\delta}{1-\delta} < \frac{(1-\alpha_{\text{teach}})^2}{3(q^o-1)}$, then $\Gamma_3^* = \{4\}$, and with probability 1/2, all workers have shared knowledge of m^{orig}. Otherwise, with probability 1/2, they have shared knowledge of a generic method m at $t = 3$.

	$t=2$ step	Γ_3	$t=2$ continuation value for worker w_i ($v_{i,2}$)			
			Worker w_1	Worker w_2	Worker w_3	Worker w_4
#1	No sharing	{2, 2}	$q^o\left(1+\frac{\delta}{1-\delta}\right)$		$1+\frac{\delta}{1-\delta}$	
#2	$w_1 \rightarrow w_3$, $w_4 \rightarrow w_2$	{2, 2}	$q^o\left(1+\frac{\delta}{1-\delta}\right)$	$1+\frac{\delta}{1-\delta}$	$q^o\left(1+\frac{\delta}{1-\delta}\right)$	$1+\frac{\delta}{1-\delta}$
#3	$w_1 \rightarrow w_3$	{3, 1}	$q^o\left(1-\alpha_{\text{teach}}+2\frac{\delta}{1-\delta}\right)$		$1-\alpha_{\text{learn}}+2\frac{\delta}{1-\delta}$	$1-\alpha_{\text{learn}}$
#4	$w_3 \rightarrow w_1$	{3, 1}	$1-\alpha_{\text{learn}}+2\frac{\delta}{1-\delta}$	$1-\alpha_{\text{learn}}$	$q^o\left(1-\alpha_{\text{teach}}+2\frac{\delta}{1-\delta}\right)$	
#5	$w_1 \rightarrow w_2$, $w_3 \rightarrow w_4$	{4}	$q^o\left((1-\alpha_{\text{teach}})^2+3\frac{\delta}{1-\delta}\right)$		$(1-\alpha_{\text{learn}})^2+3q^o\frac{\delta}{1-\delta}$	
#6	$w_3 \rightarrow w_1$, $w_4 \rightarrow w_2$	{4}	$q^o(1-\alpha_{\text{learn}})^2+3\frac{\delta}{1-\delta}$		$(1-\alpha_{\text{teach}})^2+3\frac{\delta}{1-\delta}$	

Proof. WLOG, let $\{w_1, w_2\}$ be the 2-clique that knows the original method while $\{w_3, w_4\}$ is the 2-clique that knows a generic method. Start by calculating and listing the $t = 2$ continuation value to each worker under each $t = 2$ diffusion step (isomorphic diffusion steps are omitted).

With these calculations in hand, I take the following steps to prove the Lemma. The following claims are about the $t = 2$ diffusion step.

(i) Assume that worker w_1 makes an accepted offer to worker w_3. Then in the continuation equilibrium, worker w_2 does not accept any offer from worker w_4.

(ii) The following does not occur on the equilibrium path: worker w_3 makes an accepted offer to worker w_1, followed by worker w_2 making an accepted offer to worker w_4. (Combined with step (i), this implies that the $t = 3$ state is neither $\{\{w_1, w_3\}, \{w_2, w_4\}\}$ nor $\{\{w_1, w_4\}, \{w_2, w_3\}\}$.

(iii) Assume that on the equilibrium path, at worker w_2's turn, worker w_1 has previously made an accepted offer to worker w_3. Then worker w_2 will make an offer to worker w_4, and worker w_4 will accept, if and only if $\frac{\delta}{1-\delta} > \max\left\{\alpha_{\text{teach}}(1-\alpha_{\text{teach}}), \frac{\alpha_{\text{learn}}(1-\alpha_{\text{learn}})}{3q^o}\right\}$.

(iv) Assume that, on the equilibrium path, at worker w_4's turn, worker w_3 has previously made an accepted offer to worker w_1. Then worker w_4 will make an offer to worker w_2, and worker w_2 will accept, if and only if $\frac{\delta}{1-\delta} > \max\left\{\alpha_{\text{teach}}(1-\alpha_{\text{teach}}), \frac{q^o\alpha_{\text{learn}}(1-\alpha_{\text{learn}})}{3}\right\}$.

(v) Assume that worker w_1 is the last worker to move, and that no offers have been accepted prior to worker w_1's turn. Then worker w_1 will make an offer to worker w_3, and worker w_3 will accept, if and only if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{teach}})^2}{2}$.

(vi) Assume that worker w_4 is the last worker to move, and that no offers have been accepted prior to worker w_4's turn. Then worker w_4 will make an offer to worker w_2, and worker w_2 will accept, if and only if $\frac{\delta}{1-\delta} > \max\left\{\frac{1-(1-\alpha_{\text{teach}})^2}{2}, \frac{q^o\left(1-(1-\alpha_{\text{learn}})^2\right)}{3-q^o}\right\}$.

(vii) $\Gamma_3^* \neq \{2,2\}$ if $\frac{\delta}{1-\delta} > \max\left\{\frac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}, \frac{1-(1-\alpha_{\text{teach}})^2}{2}\right\}$.

(viii) $\Gamma_3^* = \{4\}$, and all workers know the original method at $t = 3$, if $\frac{\delta}{1-\delta} > \max\left\{\frac{q^o\left(1-(1-\alpha_{\text{learn}})^2\right)}{3-q^o}, \frac{1-(1-\alpha_{\text{teach}})^2}{3(q^o-1)}\right\}$.

(ix) $\Gamma_3^* = \{4\}$, and all workers know the original method at $t = 3$, if $\max\left\{\frac{1-(1-\alpha_{\text{teach}})^2}{2}, \frac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}\right\} < \frac{\delta}{1-\delta} < \frac{q^o\left(1-(1-\alpha_{\text{learn}})^2\right)}{3-q^o}$.

(x) $\Gamma_3^* = \{4\}$, and all workers know the original method at $t = 3$ with probability 1/2, if $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{3-q^o}$ and $\frac{\delta}{1-\delta} < \frac{(1-\alpha_{\text{teach}})^2}{3(q^o-1)}$.

(xi) $\Gamma_3^* = \{2,2\}$ if $\frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}$.

Now, I proceed to prove each of these results.

Step (i): If worker w_2 accepts the offer from worker w_4, then his continuation value is $q^o(1-\alpha_{\text{learn}})(1-\alpha_{\text{teach}})+\frac{\delta}{1-\delta}$; if worker w_2 does nothing (and does not make any offers in the period), his continuation value is $q^o\left((1-\alpha_{\text{teach}})+2\frac{\delta}{1-\delta}\right)$. The latter is strictly larger than the former, so he always prefers to reject any offer from worker w_4.

Step (ii): Assume otherwise. Consider a deviation where worker w_1 rejects the offer from worker w_3. If worker w_2 makes an accepted offer to worker w_4, then both worker w_1 and worker w_2 must prefer this outcome to the outcome where both do nothing, which in turn they prefer to the outcome specified in (b); if worker w_2 accepts an offer from worker w_4, then another round will be played, and worker w_1 will accept an offer from worker w_3. Thus (b) cannot occur.

Step (iii): After worker w_3 accepts an offer from worker w_1, if worker w_4 accepts an offer from worker w_2, then we get outcome #5; if no offer from worker w_2 to worker w_4 is accepted, we get outcome #3. We can check that worker w_2 prefers #5 to #3 if and only if $\frac{\delta}{1-\delta} > \alpha_{\text{teach}}(1-\alpha_{\text{teach}})$, while worker w_4 prefers 5 to #3 if and only if $\frac{\delta}{1-\delta} > \frac{q^o\alpha_{\text{learn}}(1-\alpha_{\text{learn}})}{3q^o}$. Combining the two inequalities, worker w_2 makes an accepted offer to worker w_4 if and only if $\frac{\delta}{1-\delta} > \max\left\{\alpha_{\text{teach}}(1-\alpha_{\text{teach}}), \frac{\alpha_{\text{learn}}(1-\alpha_{\text{learn}})}{3q^o}\right\}$.

Step (iv): Similar to step (iii), only with identities permuted and with the factor q^o moved.

Step (v): Note that $\frac{1-(1-\alpha_{\text{teach}})^2}{2} > \max\left\{\alpha_{\text{teach}}(1-\alpha_{\text{teach}}), \frac{\alpha_{\text{learn}}(1-\alpha_{\text{learn}})}{3q^o}\right\}$, and $\frac{1-(1-\alpha_{\text{teach}})^2}{2} > \frac{1-(1-\alpha_{\text{learn}})^2}{2}$. Anticipating that worker w_2 will make a subsequent offer to worker w_4 that worker w_4 will accept, worker w_1 makes an offer to worker w_3 that worker w_3 accepts.

Step (vi): Note that $\frac{1-(1-\alpha_{\text{teach}})^2}{2} > \alpha(1-\alpha_{\text{teach}})$, and $\frac{(q^o-1)\left(1-(1-\alpha_{\text{learn}})^2\right)}{3-q^o} > \frac{(q^o-1)\alpha_{\text{learn}}(1-\alpha_{\text{learn}})}{3q^o}$. Anticipating that worker w_3 will make a subsequent offer to worker w_1 that worker w_1 will accept, worker w_4 makes an offer to worker w_2 that worker w_2 accepts.

Step (vii): Follows from steps (v) and (vi).

Step (viii): Note that when $\frac{\delta}{1-\delta} > \frac{1}{3-q^o}$ and $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{teach}})^2}{3(q^o-1)}$, all four workers prefer outcome 5 to any other outcome. Then the claim holds by induction on the statement "at each worker's turn, he makes a choice that is consistent with outcome #5".

Step (ix): Given the stated conditions, we can restrict attention to outcomes #1,#5,#6. Now, we may check that worker w_3 and worker w_4 prefer $6 \succ 1$ and $5 \succ 1$; while worker w_1 and worker w_2 prefer $5 \succ 1 \succ 6$. worker w_1 and worker w_2 can guarantee themselves either #1 or #5 by rejecting all offers, so #6 does not occur in equilibrium. Step (vii) implies that #1 does not occur on the equilibrium path, which implies that the equilibrium outcome is in fact #5.

Step (x): We know from previous steps that $\Gamma_3^* = \{4\}$. To show that $Pr = \frac{1}{2}$: Let the worker who is last in order be i. Then worker w_i and his clique-mate can jointly induce their desired outcome by rejecting all offers in the first round; worker w_i will then choose to make an offer at his turn which will be accepted, and worker w_i's clique-mate will subsequently also make an accepted offer. Since each worker is equally likely to be last in turn, it follows that $Pr = \frac{1}{2}$.

Step (xi): Note that when $\frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}$, each worker w_i prefers outcome $1 \succ 5$ and $1 \succ 6$; note that #1 is achieved if all offers are rejected in the period. Our claim then holds from induction on the statement "each worker rejects any offer made to him, and anticipates that any subsequent offers made to any worker will be rejected." □

Lemma A6. Under slow growth s, $\Gamma_2^* = \{2, 1, 1\}$. The originator and the early hire both know the original method.

Proof. I will use Lemmas A2 and A1 in the proof.

Let worker w_1 denote the originator and worker w_2 denote the other early hire. If $\Gamma_2 = \{2, 1, 1\}$, then by Lemma A2, worker w_1 and worker w_2 each have $t=1$ continuation value of

$$\begin{cases} \left(1+(q^o-1)\cdot 1_{\{w_1 \text{ teaches } w_2\}}\right)\left[\delta(1-\alpha_{\text{teach}})+3\dfrac{\delta^2}{1-\delta}\right] & \text{if } \dfrac{\delta}{1-\delta} > \dfrac{1-(1-\alpha_{\text{teach}})^2}{2} \\ \left(1+(q^o-1)\cdot 1_{\{w_1 \text{ teaches } w_2\}}\right)\left[\delta+\dfrac{\delta^2}{1-\delta}\right] & \text{if } \dfrac{\delta}{1-\delta} \leqslant \dfrac{1-(1-\alpha_{\text{teach}})^2}{2} \end{cases}.$$

If $\Gamma_2 = \{1, 1, 1, 1\}$, then each of worker w_1 and worker w_2 have $t = 1$ continuation value of no more than $q^o\left[\frac{\delta^2}{1-\delta}\right]$. Comparing $t = 1$

continuation values, worker w_1 and worker w_2 always prefer $\Gamma_2 = \{2, 1, 1\}$ to $\Gamma_2 = \{1, 1, 1, 1\}$; and they prefer to have worker w_1 teach worker w_2, rather than to have worker w_2 teach worker w_1, at $t = 1$. Thus worker w_1 will make an accepted offer to worker w_2 (and reject any offer from worker w_2), resulting in $\Gamma_2^* = \{2, 1, 1\}$ □

Lemma A7. Under fast growth f, $\Gamma_2^* = \{2, 2\}$. One 2-clique knows the original method, while the other 2-clique knows a generic method.

Proof. I will proceed in the following steps.

- $\Gamma_2^* \neq \{1, 1, 1, 1\}$.
- If a first offer is made and accepted at $t = 1$, a second offer will subsequently be made and accepted in the continuation equilibrium. (Combined with step (i), we can conclude that $\Gamma_2^* = \{2, 2\}$.
- Show that the originator (worker w_1) teaches at $t = 1$.

Step (i): Assume, toward a contradiction, that $\Gamma_2^* = \{1, 1, 1, 1\}$. Then, it is optimal for worker w_1 to teach at $t = 2$, corresponding to a $t = 1$ continuation value for worker w_1 of $q^o \frac{\delta^2}{1-\delta}$; while the other workers will each get $t = 1$ continuation value of at most $q^o \frac{\delta^2}{1-\delta}$. On the other hand, assume that worker w_1 deviates and makes an offer to another worker w_i at $t = 1$ that is accepted; then regardless of the remaining workers' $t = 1$ actions, (we may check, using Lemmas A4 and A5) worker w_1 and worker w_i will have $t = 1$ continuation values of at least $q^o \left(\delta + \frac{\delta^2}{1-\delta}\right)$. This implies that $\Gamma_2 = \{1, 1, 1, 1\}$ cannot be an equilibrium outcome.

Step (ii): In light of step (i), we only have to show that after a first offer is made and accepted at $t = 1$, a second offer will subsequently be made and accepted. We may split the analysis into three cases: (a) The originator makes the first accepted offer. (b) The originator accepts the first offer. (c) The originator neither makes nor accepts the first offer. We only go through case (a); the calculations for the remaining cases proceed similarly. In case (a), label the remaining two workers not involved in the first accepted offer as worker w_3 and worker w_4. If worker w_3 teaches worker w_4 (or if worker w_4 teaches worker w_3) after the first offer is made and accepted, then worker w_3's and worker w_4's $t = 1$ continuation values are

$$\begin{cases} \delta + \dfrac{\delta^2}{1-\delta} & \text{if } \dfrac{\delta}{1-\delta} < \max\left\{\dfrac{1-(1-\alpha_{\text{learn}})^2}{3q^o - 1}, \dfrac{1-(1-\alpha_{\text{teach}})^2}{2}\right\} \\ \geq \delta\left(1-(1-\alpha_{\text{learn}})^2\right) + 3\dfrac{\delta^2}{1-\delta}q^o & \text{if } \dfrac{\delta}{1-\delta} > \max\left\{\dfrac{1-(1-\alpha_{\text{learn}})^2}{3q^o - 1}, \dfrac{1-(1-\alpha_{\text{teach}})^2}{2}\right\} \end{cases},$$

where the latter value is weakly larger than their continuation value from learning at $t = 2$ from the other 2-clique. In comparison, their $t = 1$ continuation value if they do nothing at $t = 1$ after the first offer is made or accepted is

$$\begin{cases} \dfrac{\delta^2}{1-\delta} & \text{if } \dfrac{\delta}{1-\delta} < \max\left\{\dfrac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}, \dfrac{1-(1-\alpha_{\text{teach}})^2}{2}\right\} \\ 3\dfrac{\delta}{1-\delta}q^o & \text{if } \dfrac{\delta}{1-\delta} > \max\left\{\dfrac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}, \dfrac{1-(1-\alpha_{\text{teach}})^2}{2}\right\} \end{cases},$$

Comparing $t = 1$ continuation values, we can confirm that worker w_3 and worker w_4 always prefer to have worker w_3 teach worker w_4 (or vice versa) rather than do nothing.

Step (iii): Assume, toward a contradiction, that in equilibrium some worker $i \neq 1$ teaches the originator, worker w_1, at $t = 1$ (so that $\Gamma_2 = \{2, 2\}$ with both 2-cliques knowing a generic method). Consider a deviation where worker w_1 and worker w_i deviate and have worker w_1 teach worker w_i at $t = 1$ instead (so that $\Gamma_2 = \{2, 2\}$ with both worker w_1 and worker w_i knowing the original method, and the other 2-clique knowing a generic method). It is straightforward to compare the $t = 1$ continuation values under the equilibrium and the deviation, and verify that worker w_1 and worker w_i strictly prefer to deviate; I omit the calculations for brevity. Thus, in equilibrium, worker w_1 teaches at $t = 1$. □

We may rewrite Lemmas A4, A5, A6 and A7 as the following two lemmas, which then immediately imply Propositions 3a and 3b.

Lemma A8. Under slow growth,

- If $\frac{\delta}{1-\delta} < \frac{\tilde{\alpha}_{\text{teach}}}{2}$, then the long-run pattern is fragmented: $\Gamma_3^* = \{2, 2\}$. At $t = 3$, the early hires know the original method while the late hires know a generic method.
- If $\frac{\delta}{1-\delta} > \frac{\tilde{\alpha}_{\text{teach}}}{2}$, then the long-run pattern is unfragmented: $\Gamma_3^* = \{4\}$. At $t = 3$, all workers know the original method.

Lemma A9. Under fast growth,

- If $\frac{\delta}{1-\delta} < \max\left\{\frac{\tilde{\alpha}_{\text{teach}}}{2}, \frac{\tilde{\alpha}_{\text{learn}}}{3q^o-1}\right\}$, then knowledge is fragmented in the long run ($\Gamma_3^* = \{2, 2\}$). At $t = 3$, one 2-clique knows the original method and the other 2-clique knows a generic method.
- If $\frac{\delta}{1-\delta} > \max\left\{\frac{\tilde{\alpha}_{\text{teach}}}{2}, \frac{\tilde{\alpha}_{\text{learn}}}{3q^o-1}\right\}$, then there is unfragmented knowledge of a method m^* in the long run ($\Gamma_3^* = \{4\}$). Further,
- If $q^o - 1 < \min\left\{\frac{2\frac{\delta}{1-\delta}-\tilde{\alpha}_{\text{learn}}}{\frac{\delta}{1-\delta}+\tilde{\alpha}_{\text{learn}}}, \frac{1-\delta}{3\delta}(\tilde{\alpha}_{\text{learn}} - \tilde{\alpha}_{\text{teach}})\right\}$, then m may be either generic or original, depending on the order in which players make offers at $t = 2$.
- If $q^o - 1 > \min\left\{\frac{2\frac{\delta}{1-\delta}-\tilde{\alpha}_{\text{learn}}}{\frac{\delta}{1-\delta}+\tilde{\alpha}_{\text{learn}}}, \frac{1-\delta}{3\delta}(\tilde{\alpha}_{\text{learn}} - \tilde{\alpha}_{\text{teach}})\right\}$, then m^* is always the original method.

Proposition 4. Proof. First, note that when $0 < \frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{teach}})^2}{2}$, fragmentation is persistent under both growth paths, so total expected output is maximized under fast growth, in which case fragmentation is persistent.

Second, consider the regime $\frac{1-(1-\alpha_{\text{teach}})^2}{2} < \frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{2}$. Note that

$$q^o - 1 < \min\left\{\frac{2\frac{\delta}{1-\delta} - \tilde{\alpha}_{\text{learn}}}{\frac{\delta}{1-\delta}+\tilde{\alpha}_{\text{learn}}}, \frac{1-\delta}{3\delta}(\tilde{\alpha}_{\text{learn}} - \tilde{\alpha}_{\text{teach}})\right\} \tag{11}$$

(from Lemma A9) does not hold, because $\frac{q^o\left(1-(1-\alpha_{\text{learn}})^2\right)}{3-q^o} > \frac{\delta}{1-\delta}$. When $\frac{1-(1-\alpha_{\text{teach}})^2}{2} < \frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{3q^o-1}$, fragmentation is persistent under fast growth, but knowledge of the original method is fully shared under slow growth. Under these conditions, P's $t = 1$ continuation value under slow growth is $2\delta q^o(1-\alpha_{\text{learn}})^2+12\frac{\delta^2}{1-\delta}q^o$, while her $t = 1$ continuation value under fast growth is $4\delta(1+q^o)+4\frac{\delta^2}{1-\delta}(1+q^o)$. Comparing these two terms, we may calculate that under these conditions, P chooses slow growth if and only if $q^o - 1 > \frac{\frac{1}{1-\delta}}{5\frac{\delta}{1-\delta}-\left(1-(1-\alpha_{\text{teach}})^2\right)} - 1$. It follows that $\hat{q}_1 = \min\left\{\hat{q}_2, \max\left\{1, \frac{\frac{1}{1-\delta}}{5\frac{\delta}{1-\delta}-\left(1-(1-\alpha_{\text{teach}})^2\right)}\right\}\right\}$ and $\hat{q}_2 = \max\left\{1, \frac{1-(1-\alpha_{\text{learn}})^2+\frac{\delta}{1-\delta}}{3\frac{\delta}{1-\delta}}\right\}$ for the regime $\frac{1-(1-\alpha_{\text{teach}})^2}{2} < \frac{\delta}{1-\delta} < \frac{1-(1-\alpha_{\text{learn}})^2}{2}$. Also, if $\alpha_{\text{teach}} \approx 0$, $\alpha_{\text{learn}} \approx 1$, and $0 < \delta < \frac{1}{5}$, then $\left(\hat{q}_1, \hat{q}_2\right) \approx \left(\frac{1}{5\delta}, \frac{1}{3\delta}\right)$, so the intervals $(q, \hat{q}_1)$ and $(\hat{q}_1, \hat{q}_2)$ are nonempty.

Third, consider the regime $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2}$. Under these conditions, knowledge is fully shared under both slow growth and fast growth, but fully shared knowledge may either be of the original method or a generic method (with equal probability) under fast growth. In this case, the principal's $t = 1$ continuation value under slow growth is $\delta(1+q^o)\left((1-\alpha_{\text{teach}})^2+(1-\alpha_{\text{learn}})^2\right)+12\frac{\delta^2}{1-\delta}\frac{1+q^o}{2}$ while her $t = 1$ continuation value under slow growth is $2\delta q^o(1-\alpha_{\text{teach}})^2+12\frac{\delta^2}{1-\delta}q^o$. We may infer that when (11) holds, slow growth maximizes total expected output if and only if $q^o > \frac{2(1-\alpha_{\text{learn}})^2}{(1-\alpha_{\text{teach}})^2-(1-\alpha_{\text{learn}})^2+6\frac{\delta}{1-\delta}}+1$; it follows that $\hat{q}_1 = \min\left\{\hat{q}_2, \frac{2(1-\alpha_{\text{learn}})^2}{(1-\alpha_{\text{teach}})^2-(1-\alpha_{\text{learn}})^2+6\frac{\delta}{1-\delta}}\right\}+1$ and

$$\hat{q}_2 = \min\left\{\frac{3\delta}{1-(1-\delta)(1-\alpha_{\text{learn}})^2}, \frac{(1-\alpha_{\text{teach}})^2-(1-\alpha_{\text{learn}})^2}{3\frac{\delta}{1-\delta}}\right\}+1$$

for the regime $\frac{\delta}{1-\delta} > \frac{1-(1-\alpha_{\text{learn}})^2}{2}$.

Finally, note that if we choose $\alpha_{\text{learn}} \approx 1$, $\alpha_{\text{teach}} \approx 0$, $\frac{1}{3} < \delta < 1$, then $\hat{q}_1 \approx 1$, $\hat{q}_2 \approx \min\left\{3\delta, \frac{1+2\delta}{3\delta}\right\} > 0$; so the interval $(\hat{q}_1, \hat{q}_2)$ is nonempty. □

Lemma A10. Under slow acquisitive growth, $\Gamma_2^* = \{2, 2\}$.

Proof. We may check that the incentives for each early hire to share knowledge at $t = 1$ under slow acq. growth are identical to the incentives for early hires to share knowledge at $t = 1$ under fast org. growth. Specifically, under slow acq. growth: (a) if the originator teaches (respectively, learns) at $t = 1$, then $\Gamma_2 = \{2, 2\}$, with the originator's 2-clique knowing the original method and the other 2-clique knowing a generic method (respectively, with both 2-cliques knowing distinct generic methods). (b) if both early hires do nothing, then $\Gamma_2 = \{2, 1, 1\}$. We may then recycle the calculations from Lemma A5 to conclude that in equilibrium, the originator teaches the other early hire at $t = 1$. □

Proposition 5. *Proof.* Direct implication of Lemmas A7 and A10. □

Proposition 6. *Proof.* **Only if.** Suppose that $\lambda < \frac{m-1}{n}$. Let h_t be a history leading up to a worker w_i's decision whether to accept an offer in period t. We verify the following claim by induction: if no offer has yet been accepted so far in h_t, then no offer will be accepted in the continuation equilibrium following h_t. Induction assumption: at all histories hh_t, the claim holds. Let the size of w_i's clique be $m' \geq m$. By the induction assumption, if w_i stays on the continuation path, nobody else accepts an offer, so his period-t payoff is $(1 - \lambda)(m' - 1) + \lambda(m' - 1) = m' - 1 \geq m - 1$. If w_i deviates and accepts the offer, he joins a clique of size $n' \leqslant n$; so his payoff is $\lambda n' \leqslant \lambda n \leqslant m - 1$. The claim follows.

If. Suppose toward a contradiction that in period t, $\lambda > \frac{m-1}{n}$, and no knowledge sharing takes place in equilibrium. Let i be the last worker in the order such that $\lambda > \frac{m-1}{n'}$ where n' is i's clique size. (Note that i must exist by our assumption.) If n' does not make an offer, then in the continuation equilibrium, no further offers will be accepted in that period. This implies that any offer from i to a worker in an m-clique will be accepted. It immediately follows that i will deviate and make an offer, which will be accepted—a contradiction. □

The next few lemmas are used in the proof of Proposition 7.

Lemma A11. Given growth rate γ and ceiling N, at the end of each period t such that $\lfloor \gamma(t+1) \rfloor < N$, there are $\lfloor \gamma \rfloor$ cliques with size $t + 1$, and (if $\lfloor \gamma(t+1) \rfloor > \lfloor \gamma \rfloor (t+1)$) one more clique of size $\lfloor \gamma(t+1) \rfloor - \lfloor \gamma \rfloor (t+1)$.

Proof. There are a few cases to verify. The claims corresponding to each case are proven in a subsequent series of Lemmas (A12a–A12c).

- If $t = 1$, then at the beginning of the period, for $k = \lfloor \gamma \rfloor$, there are either $2k$ or $2k+1$ *singletons. We seek to show that at the end of the* *period, there are k 2-cliques and* (if $\lfloor 2\gamma \rfloor$ is odd) one additional singleton.

- If $t > 1$ and $\lfloor \gamma(t+1) \rfloor = \lfloor \gamma \rfloor (t+1)$, then at the beginning of the period t, for $k = \lfloor \gamma \rfloor$ and $n = t$, there are k n-cliques and k singletons. We seek to show that at the end of the period, there are k $(n+1)$-cliques.
- If $t > 1$ and $\lfloor \gamma t \rfloor = \lfloor \gamma \rfloor t$ and $\lfloor \gamma(t+1) \rfloor > \lfloor \gamma \rfloor (t+1)$, then at the beginning of the period t, for some $k = \lfloor \gamma \rfloor$ and $n = t$, there are k n-cliques and $k + 1$ singletons. We seek to show that at the end of the period, there are k $(n+1)$-cliques and one additional singleton.
- If $t > 1$ and $\lfloor \gamma t \rfloor > \lfloor \gamma \rfloor t$ and $\lfloor \gamma(t+1) \rfloor - \lfloor \gamma t \rfloor > \lfloor \gamma \rfloor$, then at the beginning of the period t, for $k = \lfloor \gamma \rfloor$ and $m = \lfloor \gamma t \rfloor - \lfloor \gamma \rfloor t$ and $j = \lfloor \gamma(t+1) \rfloor - \lfloor \gamma t \rfloor$, there are k n-cliques, one m-clique, and j singletons. We seek to show that at the end of the period, there are k $(n+1)$-cliques and one clique of size $m+j-k$.

Lemma A12a. Suppose that at the start of period t, there are $2k+1$ cliques, of which k cliques have size $n > 1$, k cliques are singletons, and 1 clique has size m with $1 \leqslant m < n$. Then in period t, each of the n-cliques will teach a singleton worker, and no other teaching takes place. Consequently, at the end of the period, there will remain k $(n+1)$-cliques and one m-clique.

Lemma A12b. Suppose that at the start of period t, there are 2k cliques, of which $k-1$ cliques have size $n > 1$, k cliques are singletons, and 1 clique has size m with $1 \leqslant m \leqslant n$. Then in period t, each of the n-cliques will teach a singleton worker. If $m > 1$, then the m-clique worker will also teach a singleton worker; otherwise, if $m = 1$, then the m-clique worker will either teach or learn from a singleton worker. No other teaching takes place. Consequently, at the end of the period, there will remain $k-1$ $(n+1)$-cliques and one $(m+1)$-clique.

Lemma A12c. Suppose that at the start of period t, there are either 2k or $2k+1$ singleton cliques. Then k workers will teach and k workers will learn, resulting in k 2-cliques (and perhaps one remaining singleton) at the end of the period.

Proof. I prove Lemma A12a here. The proofs of Lemmas A12b and A12c are similar, and thus are omitted.

For this proof, we restrict attention to strategies where, if an offer by w_i will be rejected on the continuation path, then w_i does not make the offer. This restriction is WLOG, but simplifies the exposition.

Consider a history, up till a point in period t, where: (a) whenever an offer was made from an n-clique to a singleton, it was accepted. (b) all offers from workers in n-cliques had been to singletons. (c) No offers from anyone other than workers in n-cliques were made or accepted. (d) if there are no workers from n-cliques left in a round, then an offer from a worker in an n-clique (to a singleton) was previously accepted in that round. Note that if (a)–(d) are satisfied on the equilibrium path, then we are done.

Claim I: given a history that satisfies (a)–(d), the continuation path for the rest of period t also satisfies (a)–(d). We proceed by induction. Given a history h that satisfies (a)–(d), suppose that Claim I holds for all histories $h'h$ that also satisfy (a)–(d); then we will show that Claim I holds for h as well, in sequence from (a) to (d).

(a) Suppose that at h, a singleton worker w_i has received an offer. If he accepts the offer, he achieves his maximum possible payoff of λn; thus by our equilibrium selection rules, he accepts the offer. (b) Suppose that at h, it is the turn of a worker w_i in an eligible n-clique, and that not all singletons have accepted offers. We know that if w_i makes an offer to an eligible singleton, then the offer will be accepted. Then, by our equilibrium selection rule, if w_i makes an offer, it will be to a singleton. (c) Suppose that at h, worker w_i is made an offer by a worker w_j who is not in an n-clique. First, consider the case where w_i is in an n-clique. Then by Claim I, if he rejects the offer, then he or a clique-mate will make an offer to a singleton, which will be accepted, in the continuation path; this outcome strictly maximizes his period-t payoff, and thus he will not accept the offer from w_j. Second, consider the case where w_i is a singleton and w_j was in an m-clique, with $m > 1$. If w_i rejects, then he will join an n-clique in the continuation path, which strictly maximizes his period-t payoff at λn; thus he strictly prefers to reject the offer. Finally, consider the case where both w_i and w_j are singletons. If either w_i rejects the offer or w_j does not make the offer, then in the continuation path, either w_i or w_j will learn from an n-clique. (By the Markovian assumption, the identity of the learner is the same in both cases.) Thus, either w_i strictly prefers not to make the offer, or w_j does not make the offer in the first place. (d) Suppose that at h, it is the turn of a worker w_j in an n-clique, and there are no other workers from n-cliques remaining in the round. If w_j makes an offer to a singleton (which will be accepted), then he weakly maximizes his period-t payoff at $(1-\lambda)(n-1)+\lambda n$; thus, by our equilibrium selection rules, he always makes an offer to a singleton. □

Proposition 7. *Proof*. Let $\bar{t} = \lfloor \frac{N}{\gamma} \rfloor - 1$ be the first period where the number of workers in the organization exceeds $\lfloor \gamma(t+1) \rfloor$. Note that the organization's size hits the ceiling N either in period $\bar{t}-1$ or $\bar{t}$.

First, consider the case where $\gamma \geq 1$ is an integer. By Lemma A11, at the end of period $\bar{t}-1$, there are γ cliques, each with $\bar{t}$ workers. Focus on period $\bar{t}$. At the beginning of the period, there are γ cliques, each with $\bar{t}$ workers, and $N-\gamma\bar{t} \leqslant \gamma$ singletons. Parallel to Lemma A12b, we may show that $N-\gamma\bar{t}$ of the $\bar{t}$-cliques will each teach a singleton, and that no other knowledge sharing takes place, in period $\bar{t}$. Consequently, at the end of period $\bar{t}$, there are a total of γ cliques, of size either $\bar{t}$ or $\bar{t}+1$. By assumption, $\gamma(\bar{t}+1) > N > \frac{2\gamma}{1-\lambda}$, which implies $\gamma < \frac{\bar{t}}{\bar{t}+2}$: Condition 10 does not hold for the period-$\bar{t}$ fragmentation pattern. Thus, knowledge sharing ceases with γ cliques.

Second, consider the case where $\gamma \geq 1$ is not an integer. We seek to show that the number of long-run cliques is weakly increasing in γ for sufficiently large N. By Lemma A11, at the end of period $\bar{t}-1$, there are $\lfloor\gamma\rfloor$ *regular* cliques with $\bar{t}$ workers, and (if $\lfloor\gamma\bar{t}\rfloor > \lfloor\gamma\rfloor\bar{t}$) one *leftover* clique with $\lfloor\gamma\bar{t}\rfloor - \lfloor\gamma\rfloor\bar{t}$ workers. We start by characterizing the long-run fragmentation pattern on a case-by-case basis.

Suppose $N = \lfloor\gamma\bar{t}\rfloor$. Then no more workers join the organization in subsequent periods. What happens from period $\bar{t}$ onwards? If $\lambda > \lfloor\gamma\bar{t}\rfloor - \lfloor\gamma\rfloor\bar{t}\bar{t}$, then workers in the leftover clique are willing to learn from the regular clique, so knowledge sharing continues in the following form: in each period, workers from the leftover clique learn from, and join regular cliques, until all workers have left the leftover clique. At the end of each period t where the leftover clique still has remaining workers, the regular cliques are of the same size $t+1$; at the end of the first period $\bar{\bar{t}}$ where the leftover clique has no workers left, the regular cliques differ in size by at most one. Notice that, in period $\bar{\bar{t}}$, every regular clique has size of at least $\bar{t}$, which ensures that Condition 10 does not hold with respect to regular cliques; so, workers from regular cliques will not learn. Thus knowledge sharing stalls with $\lfloor\gamma\rfloor$ regular cliques at the end of period $\bar{\bar{t}}$. On the other hand, if $\lambda < \frac{\lfloor\gamma\bar{t}\rfloor - \lfloor\gamma\rfloor\bar{t}-1}{\bar{t}}$, then knowledge sharing stalls at the end of period $\bar{t}-1$ with $\lfloor\gamma\rfloor+1$ cliques.

Now suppose $N > \lfloor\gamma\bar{t}\rfloor$. We analyze the case where $\lfloor\gamma\bar{t}\rfloor > \lfloor\gamma\rfloor\bar{t}$; the case $\lfloor\gamma\bar{t}\rfloor = \lfloor\gamma\rfloor\bar{t}$ is trivial. At the start of period $\bar{t}$, $N - \lfloor\gamma\bar{t}\rfloor < \lfloor\gamma\rfloor+1$ singleton workers join the organization. The largest $N - \lfloor\gamma\bar{t}\rfloor$ cliques each teach a singleton, while the leftover clique does not teach because there are not enough singletons. After all singletons have accepted offers, regular cliques have sizes of either $\bar{t}$ or $\bar{t}+1$, while the leftover clique has size $\lfloor\gamma\bar{t}\rfloor - \lfloor\gamma\rfloor\bar{t}$. Similar to the case where $N = \gamma\bar{t}$, if $\lambda > \frac{\lfloor\gamma\bar{t}\rfloor - \lfloor\gamma\rfloor\bar{t}-1}{\bar{t}+1}$, then further knowledge sharing takes place, resulting in $\lfloor\gamma\rfloor$ cliques in the long-run; otherwise knowledge sharing stalls, and $\lfloor\gamma\rfloor+1$ cliques persist in the long-run.

Having characterized the long-run outcome as a function of parameters, I now perform the following exercise. Given $\gamma = \gamma_0$ and N, how does a small change in γ affect the long-run outcome? Again, we proceed on a case-by-case basis, with the goal being to show that the number of cliques in the long run is weakly increasing in γ. To clarify notation, define $\bar{t}_0 = \bar{t}(\gamma_0)$ and $\bar{\bar{t}}_0 = \bar{\bar{t}}(\gamma_0)$ to be the respective thresholds given $\gamma = \gamma_0$.

Start with the case where γ_0 is an integer. We know that for sufficiently large N, the long-run fragmentation pattern always has at least $\lfloor\gamma\rfloor$ cliques. Thus $f(\gamma, N)$ is increasing in a neighborhood of γ_0. For the remaining cases, we suppose γ is not an integer.

Next, suppose $N = \lfloor\gamma_0\bar{t}\rfloor$. Let s be the size of the leftover clique at the end of period $\bar{t}_0 - 2$. As γ increases, so does (weakly) the number of workers $\lfloor\gamma t\rfloor$ in each period t before $\bar{t}_0 - 1$, and thus the number of workers in the leftover clique – but note that the size of each regular clique remains at $t+1$ at the end of each such period t. Eventually, for sufficiently large γ, the number of workers in the leftover clique at the end of period $\bar{t}_0 - 1$, after all singletons have joined other cliques, is $s+1$, while the largest regular clique's size at that point is at most $\bar{t}_0$. Subsequently, the leftover clique is absorbed into the regular cliques only if $\lambda > \frac{s}{\bar{t}_0}$, which is stricter then the corresponding (necessary and sufficient) condition $\lambda > \frac{s-1}{\bar{t}_0}$ when $\gamma = \gamma_0$. Thus, the number of long-run cliques weakly increases.

Now, suppose $N > \lfloor \gamma_0 \bar{t} \rfloor$. Let s be the size of the leftover clique at the end of period $\bar{t}$. First, consider the case where $N > \lfloor \gamma_0 \bar{t} - 1 \rfloor + 1$. Similar to the case $N = \lfloor \gamma_0 \bar{t} \rfloor$, as γ increases, so does (weakly) the size of the leftover clique in each period. Eventually, one of two possibilities is realized. Possibility A: the size of the leftover clique at the end of period $\bar{t} - 1$ becomes $s+1$, but the ceiling N is not achieved in period $\bar{t} - 1$. Then the leftover clique is absorbed into the regular cliques in period $\bar{t}$ only if $\lambda > \frac{s-1}{\bar{t}_0}$, which is stricter then the corresponding (necessary and sufficient) condition $\lambda > \frac{s-1}{\bar{t}_0+1}$ when $\gamma = \gamma_0$. Possibility B: the ceiling N is achieved in period $\bar{t} - 1$. Then the size of the leftover clique at the end of period $\bar{t} - 1$ strictly increases from s to $s' = N - \lfloor \gamma(\bar{t} - 1) \rfloor$, and it is absorbed into the regular cliques if and only if $\lambda > \frac{s'-1}{\bar{t}_0}$, which is stricter than the corresponding (necessary and sufficient) condition $\lambda > \frac{s-1}{\bar{t}_0+1}$ when $\gamma = \gamma_0$. Thus for both A and B, the number of long-run cliques weakly increases. $\square$

References

Akerlof, Robert J. 2016. "*Value Formation: The Role of Esteem.*" Working paper, University of Warwick.

Beer, Michael. 1990. "People Express Airlines: Rise and Decline." Harvard Business School Case 490-012.

Blume, Andreas and April Mitchell Franco. 2007. "Decentralized Learning from Failure," 133 *Journal of Economic Theory* 504–23.

Crawford, Vincent P. and Hans Haller. 1990. "Learning How to Cooperate - Optimal Play in Repeated Coordination Games," 58 *Econometrica* 571–95.

Crémer, Jacques. 1993. "Corporate Culture and Shared Knowledge," 2 *Industrial and Corporate Change* 351–86.

Crémer, Jacques, Garicano Luis, and Prat. Andrea 2007. "Language and the Theory of the Firm," 122 *Quarterly Journal of Economics* 373–407.

Ellison, Glenn and Drew Fudenberg. 1995. "Word-of-Mouth Communication and Social Learning," 110 *Quarterly Journal of Economics* 93–125.

Ellison, Glenn and Richard Holden. 2014. "A Theory of Rule Development," 30 *Journal of Law Economics and Organization* 649–82.

Feiler, Lauren and Colin F. Camerer. 2010. "Code Creation in Endogenous Merger Experiments," 48 *Economic Inquiry* 337–52.

Gittell, Jody Hoffer. 2003. *The Southwest Airlines Way: Using the Power of Relationships to Achieve Higher Performance*. New York, NY: McGraw-Hill.

Grant, Robert M. 1996. "Toward a Knowledge-Based Theory of the Firm," 17 *Strategic Management Journal* 109–22.

Hambrick, Donald C. and Lynn M. Crozier. 1985. "Stumblers and Stars in the Management of Rapid Growth," 1 *Journal of Business Venturing* 31–45.

Hermalin, Benjamin E. 2013. "Leadership and Corporate Culture," in R. Gibbons and J. Roberts, eds., The Handbook of Organizational Economics. Princeton, NJ: Princeton University Press.

Heskett, James L. and Roger H. Hallowell. 1993. "Southwest Airlines: 1993 (A)." Harvard Business School Case 694-023.

Klimt, Bryan and Yiming Yang. 2004. "The Enron Corpus: A New Dataset for Email Classification Research," *Proceedings of ECML'04, 15th European Conference on Machine Learning* 217–26.

Kogut, Bruce and Udo Zander. 1992. "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology," 3 *Organization Science* 383–97.

Kreps, David M. 1990. "Corporate Culture and Economic Theory," in J.E. Alt and K.A. Shepsle, eds., Perspectives on Positive Political Economy. Cambridge, MA: Cambridge University Press.

Lazear, Edward P. 1995. "Corporate Culture and the Diffusion of Values," in H. Siebert , ed., Trends in Business Organization. Tubingen, Germany: J.C.B. Mohr.

———. 1999. "Culture and Language," 107 *Journal of Political Economy* S95–126.

Meagher, Kieron and Suraj Prasad. 2016. "Career Concerns and Team Talent," 129 *Journal of Economic Behavior & Organization* 1–17.

Nelson, Richard and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Palacios-Huerta, Ignacio and Andrea Prat. 2012. "The Impact Factor of Managers." Working paper, Columbia University.

Penrose, Edith Tilton. 1959. *The Theory of the Growth of the Firm*. New York, NY: Wiley.

Polanyi, Michael. 1966. *The Tacit Dimension*. New York, NY: Doubleday.

Prescott, Edward C. and Michael Visscher. 1980. "Organization Capital," 88 *Journal of Political Economy* 446–61.

Schein, Edgar H. 1985. *Organizational Culture and Leadership*. San Francisco, CA: Jossey-Bass.

Selten, Reinhard and Massimo Warglien. 2007. "The Emergence of Simple Languages in an Experimental Coordination Game," 108 *Proceedings of the National Academy of Sciences* 7361–66.

Spector, Robert and Patrick D. McCarthy. 2005. *The Nordstrom Way to Customer Service Excellence*. Hoboken, NJ: John Wiley & Sons.

Stinchcombe, Arthur. 1965. "Social Structure and Organizations," in J.G. March, ed., Handbook of Organizations. Chicago, IL: Rand McNally.

Van den Steen, Eric J. 2005. "Organizational Beliefs and Managerial Vision," 21 *Journal of Law Economics and Organization* 256–83.

———. 2010a. "Culture Clash: The Costs and Benefits of Homogeneity," 56 *Management Science* 1718–38.

———. 2010b. "On the Origin of Shared Beliefs (and Corporate Culture)," 41 *RAND Journal of Economics* 617–48.

Vlasic, Bill and Bradley A. Stertz. 2002. *Taken for a Ride: How Daimler-Benz Drove Off With Chrysler*. New York, NY: Harper Collins.

Weber, Roberto A. 2006. "Managing Growth to Achieve Efficient Coordination in Large Groups," 96 *American Economic Review* 114–26.

Weber, Roberto A. and Colin F. Camerer. 2003. "Cultural Conflict and Merger Failure: An Experimental Approach," 49 *Management Science* 400–15.