Description of the computational model

Hierarchical decision structure

A firm needs to first choose a primary trajectory $i \in \{1,2, \dots, A\}$ and then determine a complementary technology $j \in \{1,2, \dots, B\}$. To facilitate exposition, for the rest of this paper, we will call the A primary technological *trajectories* and the B complementary technologies *alternatives*, and we will simply say *technological combinations* or *choices* when referring to any of the total AB choices without referring to the hierarchical structure.

The payoff from a choice (i, j) follows a normal distribution with the mean $\mu_{i,j}$ and unit variance. The mean of this distribution is determined as follows. The average payoffs from each technological path μ_i are drawn independently from a normal distribution with a zero mean and a variance δ_A^2, while the average payoff for each sub-alternative under the *i*-th alternative $\mu_{i,j}$ is drawn independently from a normal distribution with the mean μ_i and a variance δ_B^2.

Without loss of generality, we order the alternatives by their average payoff, so $\mu_1 > \mu_2 > \dots > \mu_A$ and $\mu_{i,1} > \mu_{i,2} > \dots > \mu_{i,B}$ for $i = 1,2, \dots, A$. In other words, the average payoff of the alternatives decreases with its index, and the average payoffs of the sub-alternatives under each alternative also decreases with its index. The average payoff of the *i*-th alternative μ_i is greater than that from an inferior alternative μ_{i+1}. However, it is still possible that the best alternative in the (i+1)th trajectory has a greater average payoff than the worst alternative in the *i*th trajectory ($\mu_{i,B} < \mu_{i+1,1}$).

The organization forms its beliefs about the expected returns from these alternatives based on its previous experiences. It can choose to exploit its current knowledge by choosing the alternative that is believed to be the best, or to explore other alternatives to obtain more-accurate beliefs, with the hope of identifying a better alternative with a higher payoff.

The state of the environment can be represented as the expected payoffs from each of these alternatives $\boldsymbol{\mu} = [\mu_{1,1}, \dots, \mu_{1,B}, \mu_{2,1}, \dots, \mu_{A,B}]$. Beliefs and decision rules are the two central components of experiential learning. Beliefs at time t, $\boldsymbol{q_t} = [q_{1,1,t}, q_{1,2,t}, \dots, q_{A,B,t}]$, are a subjective assessment of the expected payoffs of the alternatives. Beliefs are initialized as the expected payoffs of the alternatives—a zero vector in our setting. Denote $s = (i, j)$. If a firm receives a payoff R_t from choice s, its belief is updated as follows:

$$q_{s,t+1} = q_{s,t} + a\big(R_t - q_{s,t}\big),$$

where $0 \leq a \leq 1$ is the rate of updating: in light of the new payoff R_t, the belief is updated in the direction of the new payoff. The parameter a represents the weight placed on the most recent experience: if $a = 1$, then the new belief is always the same as the most recent payoff; if $a = 0$, then the belief is never affected by the new experience. If $0 < a < 1$, the new belief is a weighted average of the old belief and the new payoff. In the experiment, we use $a = \frac{1}{k_s+1}$, where k_s is the number of times alternative s has been chosen before, so the belief is the average of all the previous experienced payoffs. For all the other alternatives that are not chosen in period t, beliefs do not change: $q_{s',t+1} = q_{s',t}, \forall s' \neq s$.

In each period t, a firm first chooses an alternative i with the following probability based on the SoftMax rule (Fang & Levinthal, 2009; Posen & Levinthal, 2012)

$$p_i = \frac{e^{\bar{q}_i/\tau}}{\Sigma_\mathrm{k} e^{\bar{q}_k/\tau}},$$

where $\bar{q}_i = \frac{1}{B}\Sigma_j q_{i,j}$ is the belief of the average payoff of all i's sub-alternatives, and τ is a parameter reflecting the firm's tendency for exploration. When $\tau \to 0$, the firm uses a greedy strategy and chooses the alternative it believes to be the best with certainty. When $\tau \to \infty$, all the alternatives are chosen with the same probability, regardless of the beliefs on their expected payoff, and the firms take the time to explore all the alternatives.

After choosing alternative i, the firm uses the same rule in choosing a sub-alternative j with the following probability conditional on i is chosen:

$$p_{i|j} = \frac{e^{q_{i,j}/\tau}}{\sum_{\mathrm{k}} e^{q_{i,k}/\tau}},$$

In addition to intra-firm learning, a firm can also choose to imitate the choice made by other firms. A firm's learning strategy can be characterized by its exploration tendency τ, defined above, and its imitation tendency λ, which denotes the probability that a firm will decide to engage in inter-firm imitation instead of intra-firm learning in a period. As mentioned above, firms have information only on the level-1 choice of other firms and can, therefore, only imitate the level-1 choice. When a firm decides to engage in inter-firm imitation in a period (with probability λ), instead of using the decision rule, it will copy the level-1 choice of the imitation target. This adaptive and iterative process of organizational learning is illustrated in Figure A1.

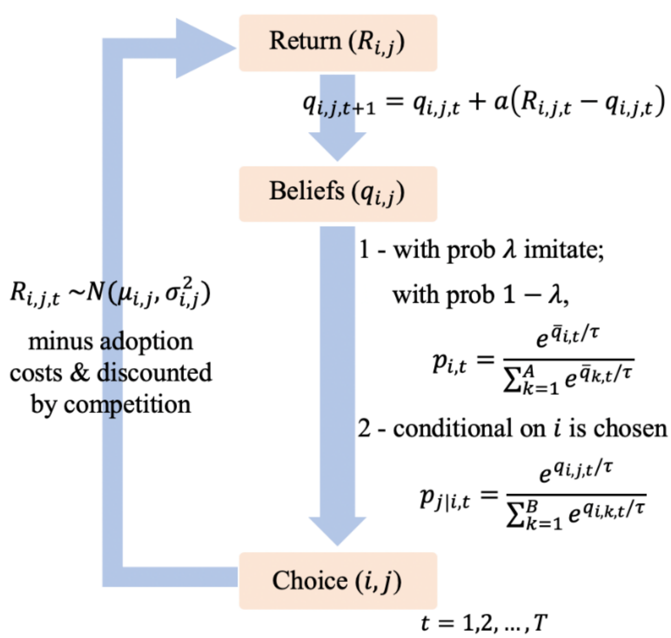

Figure A1. The adaptive process of organizational learning. In each period t, an organization chooses an alternative i, j based on its beliefs $q_{i,j,t}$ (with probability $1-\lambda$) or based on imitation (with probability λ). The return from this alternative is drawn from a normal distribution, minus adoption costs, and discounted by competition. The return influences the belief $q_{i,j,t}$ on the expected return, which, in return, changes the probability that each alternative is chosen.

Adaptability and the experience curve effects

When a new technological trajectory $i \in \{1,2,\dots,A\}$ emerges in an industry, it incurs an adoption cost of c_A for all firms adopting this technology, and this cost decreases with the number of times the technology has been adopted by firms in the industry as the technology has become more mature. This adoption cost decreases in the following manner:

$$c_{i,t}^{A} = c_A\,\beta^{(k_{i,t}-1)},$$

where $k_{i,t}$ represents the number of times technology i has been chosen by all firms in the industry at time t; the parameter β $(0 \leq \beta \leq 1)$ measures the intensity of the experience curve effect, and a smaller β represents steeper learning curves. Similarly, when an organization f starts to adopt a new technology $s \in \{(i,j)|i \in \{1,2,\dots,A\}, j \in \{1,2,\dots,B\}\}$, it bears an adoption cost of c_B, and this cost decreases exponentially over time as an organization becomes more adept in the use of this technology.

$$c_{f,s,t}^{B} = c_A\,\beta^{(k_{f,s,t}-1)},$$

$k_{f,s,t}$ represents the number of times technology s has been chosen by firm f at time t. Based on empirical evidence from previous studies, we set $\beta = 0.8$ (The Boston Consulting Group, 1970), $c_A = 2$, and $c_B = 1$.

Competition

When firms choose the same market, that increases competition with the incumbents. The payoff that a firm receives from the bandit problem is discounted by the number of existing firms in the market.

Suppose that, in time t, firm f chooses alternative s, and a total of $n_{s,t}$ firms choose the same alternative. Without competition, the payoff received by firm f is $R_{f,s,t}$, which is determined from the random draw from bandit s minus implementation costs. Competition would negatively influence the payoff and be [?] discounted by a factor $\delta^{n_{s,t}-1}$, where $0 < \delta < 1$ measures the intensity of competition. In other words, if there is only one firm that chooses s in period t, the payoff will not be influenced, while an increasing number of firms choosing the same action discounts the payoff received by the firm.

To model the first-mover advantage, we also assume that the payoff is partially dependent on how long a firm has been in the corresponding market. Denote $k_{f,s,t}$ as the number of times firm f has chosen alternative s so far; and $\bar{k}_{f,s,t} = \frac{1}{n_{s,t}} \Sigma_{f'} k_{f',s,t}$ as the average number of times alternative s has been used by all $n_{s,t}$ firms. Combining these two effects, the payoff received by firm f in period t can be expressed as

$$R_{f,t} = \frac{k_{f,s,t}}{\bar{k}_{f,s,t}} R_{f,s,t} \delta^{n_{s,t}-1}.$$