Sequential Learning and the Structure of Knowledge

Hongyi Li*†

April 2011‡

Abstract

When learning is sequential, the structure of knowledge restricts what we can learn based on what we know. I describe a model of individual sequential learning where the player faces a tradeoff between knowledge acquisition and knowledge exploitation. Knowledge structure is modeled as a directed tree. I show that when knowledge is noncomplementary, the player specializes over the long run by acquiring knowledge on only one branch of the tree. However, when knowledge is complementary, the player pursues a balanced learning strategy across multiple branches. The model exhibits path dependence: in some learning environments, the player specializes with positive probability and balances with positive probability.

1 Introduction

Learning is often sequential. Cohen and Levinthal (1990) state that "learning is cumulative, and learning performance is greatest when the object of learning is related to what is already known." Ellis (1965) asserts that "students who have thoroughly mastered the principles of algebra find it easier to grasp advanced work in mathematics such as calculus." On the other hand, expertise in literature does not make it easier to learn physics. There is a structure of knowledge that constrains what an individual can learn based on what he previously learned.

As a motivating example, consider the learning problem faced by a budding mathematician who is just starting college. This student seeks to learn mathematics - in our example, he does so by taking courses - so that he can use his acquired knowledge to write mathematics papers. He is constrained by a number of factors. First, he faces a time constraint which limits the number of

*I am particularly grateful to Robert Gibbons for his patient advice and encouragement. I also thank Philippe Aghion, Sandeep Baliga, Luis Garicano, Jerry Green, Jonathan Hall, Heikki Rantakari, Richard Townsend and Birger Wernerfelt for comments and advice.

†MIT Sloan School of Management, hongyili@mit.edu

‡Work in progress. Comments are very welcome, but please do not cite or quote this version.

classes he can take in a given semester. Second, he has a random opportunity set: for example, some classes may be unavailable in each semester. Third, he faces a sequential learning constraint: for example, he cannot study analysis until he has learned calculus. At the beginning of the learning process, the student's initial choices of introductory courses may be essentially random, being determined by the vagaries of scheduling, or even the flip of a coin. Over time, the student will accumulate knowledge of various mathematical topics. Suppose that, at some point, the student is more advanced in geometry than algebra. Because learning is sequential, the student can choose to study an advanced topic in geometry, or a less advanced topic in algebra. If more advanced knowledge is more valuable (for example, because it allows the student to write higher-quality papers), the student will bias his learning towards the subject that he is more advanced in. This effect drives the student towards a specialised skillset.

The idea that sequential learning constraints in academia may drive knowledge specialization is consistent with the observation that researchers in technical and scientific fields tend to be extremely specialised in their academic interests. The logic of sequential learning applies in other non-academic settings, such as business: the idea that firms face sequential learning constraints in knowledge acquisition is the basis of the management literature on absorptive capacity, pioneered by Cohen and Levinthal (1989, 1990). However, knowledge acquisition in non-academic settings tends to be less specialized; for example, Lazear (2005) finds that entrepreneurs tend to be generalists in acquiring knowledge, and argues that complementarities in different branches of knowledge are behind the tendency of entrepreneurs to become generalists. These differences suggest that the structure of knowledge is an important determinant of an individual's optimal learning strategy.

To understand the relationship between the structure of knowledge and the incentives for knowledge acquisition, this paper analyzes a model of learning where a single player repeatedly and randomly encounters opportunities to (i) exploit his existing skillset for productive activity, and (ii) acquire new knowledge. We assume that the set of knowledge is discrete, and we call each piece of knowledge a *skill*, and we call the subset of knowledge that the player knows his *skillset*. We make a strong assumption about knowledge structure: skills correspond to vertices on a directed rooted tree. Our premise of sequential learning then corresponds to the assumption that a skill can only be acquired if it is adjacent to another skill that is already in the player's skillset. Over time, the player's skillset expands as he acquires more and more skills. As an rough illustration, Figure 1 depicts an example of a knowledge tree and a snapshot of the player's skillset (represented by the darkened vertices) .

In our model, there is a tradeoff between knowledge acquisition and knowledge exploration: knowledge acquisition has the opportunity cost of neglecting profitable exploitation activities. As the player's skillset expands, he has more opportunities to exploit his existing knowledge, so the

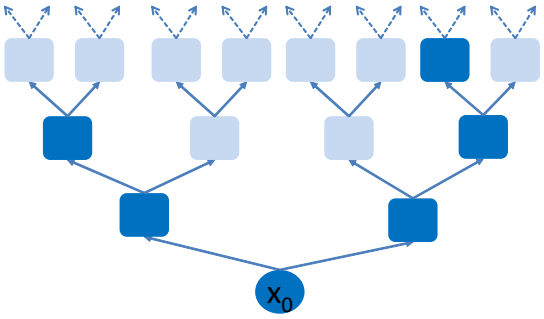


Figure 1:

player will become more selective in knowledge acquisition as his skillset expands.

Surprisingly, in learning environments where the player never completely stops learning, the player eventually becomes completely specialized. In the context of our model, this means that, over the long run, the player stops learning on all but one branch of the knowledge tree. The driving force behind such specialization is the increasing selectivity effect. When acquisition on one branch of the knowledge tree falls sufficiently far behind acquisition on the other branches, the player will cease subsequent acquisition on that branch because the opportunity cost of such acquisition exceeds the marginal value of any available skill on that branch. In our model, over the long run, acquisition on all but one branch of the tree inevitably falls behind in this fashion. The player is locked into knowledge acquisition on the remaining branch, resulting in specialization.

The specialization result assumes that skills are noncomplementary: exploitation of any piece of knowledge does not require knowledge from other branches. This is a reasonable approximation in many environments. For example, knowledge of one narrow field is often sufficient for academic research. On the other hand, knowledge is highly complementary in other environments. For example, for an entrepreneur, advanced technical knowledge are useless unless he also has the marketing knowhow to sell his product. To model the latter situation, we consider a learning environment that captures the notion of knowledge complementarity: to exploit a skill, the player has to know a similarly advanced skill on another branch of the knowledge tree. In this environment, unbalanced acquisition (where the player acquires skills on one branch that are substantially deeper than his acquisitions on the other branches) cannot be immediately exploited, and thus is inefficient. As a result, the player optimally adopts a balanced acquisition strategy.

I then present an example of a hybrid structure where some skills require complementary skills and other skills do not. The hybrid structure demonstrates that the long-run learning outcome may be history-dependent: depending on the initial draws of opportunities, the player may end

up with a specialized or a balanced skillset. Intuitively, this occurs because of positive feedback in the acquisition process. A small amount of initial specialization increases the returns to further specialization, while initially balanced skillsets are conducive to further balanced acquisition. As a result, the player eventually and inevitably falls either into a specialized learning strategy or a balanced learning strategy.

The rest of the paper is organized as follows. Section 1.1 discusses related literature. Section 2 presents the basic framework. Section 3 discusses determinants of the player's optimal acquisition decision. Section 4 commences discussion of skillset structure and presents the main specialization result. We then move to on extensions of the basic model. Section 5 introduces the notion of complementary skills and shows how this may induce the player to generalize. Section 5.1 then introduces the hybrid structure and discusses history dependence. Appendix A introduces correlations to the availability of skills. Appendix B discusses an alternative formulation for skill complementarity. Appendix C contains the proofs.

1.1 Related Literature

There is a substantial economic literature on path dependence in technology adoption in communities, highlighted by Arthur (1989). This literature highlights similar lock-in effects to those that the present paper analyzes, albeit in a different setting. Our model presents a distinct set of insights; for example, it highlights how path dependence in sequential learning results from the exploitation-acquisition tradeoff.

Our paper is most closely related to the literature on the dynamics of technological innovation. Aghion and Harris (1992) analyze a model where firms progress step-by-step along a technology ladder indexed by $\{1, 2, 3, ...\}$ at a rate which depends on their investment in research. Jovanovic and Nyarko (1996) consider a ladder model which explicitly incorporates optimal experimentation at each step of the ladder: learning about the optimal action at each step also generates information about the optimal action for succeeding steps. Jovanovic and Nyarko show that an agent who knows too much about the optimal action at an existing step may be reluctant to forgo this knowledge by progressing to the next step on the ladder, thus falling into a competency trap. Acemoglu's (2002) model of directed technical change emphasizes the tradeoff between research into different and possibly complementary factors. These papers incorporate elements of sequential learning into their analysis; the present paper builds on this literature by flexibly modeling the structure of knowledge and analyzing its effect on specialization (or the lack thereof) in the learning process.

The current paper is also related to the literature on the exploration-exploitation tradeoff in learning. March (1991) discusses, qualitatively, the tradeoff between exploration and exploitation in organizations. Doraszelski (2004) analyzes a dynamic model of learning which shares with the

present paper the feature that knowledge is accumulated. The present paper focuses on how sequential learning occurs in the presence of an exploration-exploitation tradeoff.

There are a number of papers that study specialization in knowledge acquisition, in a static setting. Lazear (2005) considers an agent who makes a one-time investment in each of two types of skills, then chooses either a production technology that utilizes only one of these skills or a production technology that generates output proportional to the minimum skill level across both skills. He equates the latter technology with being an entrepreneur. This is quite similar to the setting of section 5.1, where some skills require other complementary skills. Jones (2009) presents a model of innovation where agents make a one-time investment in knowledge and face a one-time choice over the location, breadth and depth of their acquired knowledge. The present model builds on these results by analyzing the dynamics of knowledge acquisition and demonstrating how specialization arises in the learning process.

2 The Basic Model

We start with a basic framework which allows us to describe an environment where skills are noncomplementary. In section 5, the model will be augmented to study settings where skills are complementary.

2.1 Knowledge Structure

The structure of knowledge is represented as a directed graph (X, Z), with vertices $x \in X$ of vertices and edges $(x, x') \in Z$. Specifically, assume that (X, Z) is a directed tree.$|X|$ may be finite or infinite. Each vertex x corresponds to a discrete skill. If $(x, x') \in Z$, then we write $x \to x'$, and we say that skill x' is adjacent to skill x. We also abuse notation and say that x is adjacent to a set $X' \subset X$ if there exists $x' \in X'$ such that $x' \to x$.

Definition 1. A directed tree (X, Z) with root x_0 is a directed graph where, for some vertex $x_0 \in X$ and every other vertex $x \neq x_0$, there is a unique path $x_0 \to ... \to x$ between x_0 and x.

One may interpret this set of assumptions as saying that if x' is adjacent to x, then x' is a more advanced skill than x.

A *path* is an infinite series $\{x_0 \to x_1 \to x_2 \to ...\}$ of adjacent skills. For any skill $x \neq x_0$, denote the depth (the minimum distance from the root x_0 to x, as measured by the number of edges between the root and x) of a given skill x by $d(x)$. If $d(x) > d(x')$, then we say that x *is deeper than* x'.

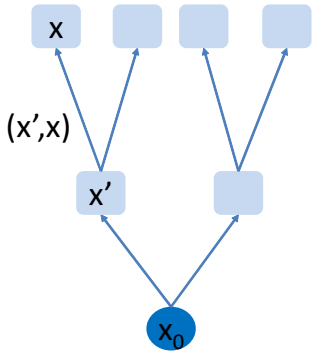


Figure 2: *An example of a tree structure of knowledge*

2.2 Actions, Payoffs, and Timing

There is a single player in a infinite-horizon, discrete-time setting with $t \in \{1, 2, 3, ...\}$.

The relevant domain of knowledge is the set X of skills. We denote the player's knowledge in period t as $\Psi_t \subset X$. We assume that the player is initially endowed with the root skill: $\Psi_1 = \{\underline{x}\}$. Let $\Lambda_t \subseteq X$ be the set of available skills in period t. To capture the premise that the player encounters random opportunities for exploitation and acquisition, we assume that for each skill $x \in X$,

$$\Pr[x \in \Pi_t] \equiv p \in (0, 1),$$

and that availability is independent across skills and periods.

In each period t, the player chooses an action

$$x_t \in (\Psi_t \cup \Lambda(\Psi_t)) \cap \Pi_t, \text{ where}$$
$$\Lambda(\Psi_t) \equiv \{x : x_t \longleftarrow x \text{ for some } x \in \Psi_t\}.$$

In other words, this action x_t corresponds to an available skill that is either in his skillset Ψ_t, or adjacent to a skill in his skillset. If $x_t \in \Psi_t$, then we say that the player exploits an existing skill. If $x_t \in \Lambda(\Psi_t)$, then $\Psi_{t+1} = \Psi_t \cup x_t$, and we say that the player acquires a new skill.

In each period t, the player's stage payoff is

$$u(x_t) = \begin{cases} \pi(x_t) : x_t \in \Psi_t, \\ -c : x_t \in \Lambda(\Psi_t) \end{cases}$$

so he receives an skill-dependent exploitation payoff for exploiting a skill, and incurs a cost $c > 0$

for acquiring a new skill.

The model we have presented so far may be interpreted as a stylized learning-by-doing environment. In this environment, when an opportunity to utilize a skill arises, the player can acquire the skill and then exploit it in the future.[1]

2.3 Strategies

We call $(X, Z, \pi(\cdot))$ the *learning environment*, and we assume that it is known to the player.

Let Φ_t be the set of available skills in period t. For each period t, a history h_t is the information set

$$h_t = \{\Phi_1, x_1, ..., \Phi_{t-1}, x_{t-1}, \Phi_t\},$$

available to the player prior to taking action x_t. A strategy $\varphi(\cdot)$ is a mapping from the set of all histories to actions $x_t \in X$. We consider optimal strategies φ^* that, in each period t, maximize the player's expected discounted payoff

$$E\left[\sum_{\tau=t}^{\infty} \delta^{\tau} u(\varphi(h_\tau)) \middle| h_t\right].$$

Under an optimal strategy φ^*, since the optimal continuation strategy at any history h_t depends only on the realization of available skills Φ_t in period t and the player's skillset Ψ_t, we denote

$$V^*(\Psi_t, \Phi_t) = E\left[\sum_{\tau=t}^{\infty} \delta^{\tau} u(\varphi(h_\tau)) \middle| h_t, \varphi^*\right], \text{ and}$$
$$V^*(\Psi_t) = E_{\varphi^*, \Phi_t}\left[V(\Psi_t, \Phi_t)\right],$$

so $V^*(\Psi_t)$ is the expected continuation payoff to the player, at the start of period t, given skillset Ψ_t and an optimal strategy φ^*. We also denote Ψ_t^* to be the skillset in period t, on the equilibrium path.

3 Optimal Acquisition

This section presents some preliminary analysis of the player's optimal acquisition choice.

Lemma 1. *In any period t and for any strategy φ, the skillset Ψ_t forms a subtree of X.*

[1] Learning is deterministic in the following sense. Learning costs are fixed and the player learns the skill perfectly in exactly one period; thereafter, he receives the same payoff whenever he exploits that situation. This is a simplifying assumption relative to usual learning-by-doing models where learning takes place by repeated experimentation.

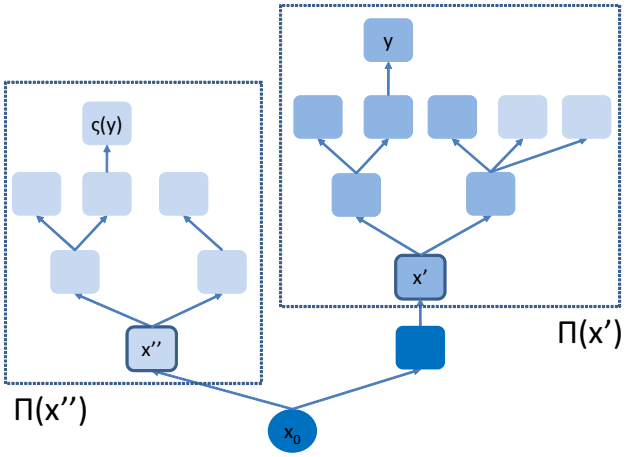


Figure 3: $\Pi(x')$ *is richer than* $\Pi(x'')$

This result is a mechanical consequence of the sequential learning assumption: because only skills that are adjacent to previously-acquired skills can be acquired, the player's skillset Ψ_t must always be connected and thus must form a tree.

Next, we present a intuitive result: when faced with a choice between acquiring two distinct skills, the player prefers to acquire the skill that (i) has more exploitation value, and (ii) gives access to skills with more exploitation value. To formalize this result, we first define

Definition 2. The *subtree* $\Pi(x)$ *of a skill* x is the set of skills x', including x, such that there exists a path $x \to ... \to x'$ from x to x'.

Definition 3. We say that $\varsigma : X_1 \to X_2$ is a *structural isomorphism* between X_1 and X_2 if ς is a bijection and $x', x'' \in X_1$, $x' \to x''$ if and only if $\varsigma(x') \to \varsigma(x'')$. That is, X_1, X_2 have identical structure (although payoffs may differ).

We say $x' \succeq x''$ if, for some $X' \subset \Pi(x')$, there exists a structural isomorphism $\varsigma : X' \to \Pi(x'')$ such that $\pi(\varsigma(x)) \leq \pi(x)$ for all $x \in X'$. That is, x' has a richer subtree than x'': loosely speaking, $\Pi(x')$ has (weakly) more branches than $\Pi(x'')$, and the skills in $\Pi(x')$ are weakly more profitable to exploit than those in $\Pi(x'')$. Figure 3 depicts an example where $x' \succeq x''$, without depicting $\pi(\cdot)$.

The following result formalizes our intuition that the player prefers to acquire the skill with the richest subtree.

Proposition 1. *Assume that* x' *and* x'' *are both adjacent to* Ψ. *If* $x' \succeq x''$, *then* $V(\Psi \cup x') \geq V(\Psi \cup x'')$.

This result is key to the rest of the analysis. Informally, it states that the player prefers to acquire the richest possible subtree because it opens up weakly more future opportunities for acquisition and exploitation, and because the opportunities it opens up are weakly more lucrative.

Another feature of optimal acquisition is what we call *increasing selectivity*, as described in the following proposition.

Proposition 2. *Assume that* Ψ *and* Ψ' *are skillsets with* $\Psi \subset \Psi'$*, and that* x *is adjacent to both* Ψ *and* Ψ'*. Then*

$$V^*\left(\Psi' \cup x\right) - V^*\left(\Psi'\right) \geq V^*\left(\Psi \cup x\right) - V^*\left(\Psi\right).$$

Proposition 2 states that the marginal value of acquiring a given skill is weakly decreasing in the player's existing skillset. This result has a straightforward explanation. The marginal value of acquiring skill x is, loosely speaking, the increase in future expected exploitation payoffs when x and skills in the subtree of x are added to the skillset. However, a player with a larger skillset is less likely to choose x to exploit in any subsequent period. A player with a larger skillset also likely has better alternative actions compared to exploiting x. As a result, x has a smaller marginal value to a player with a larger skillset. The main point of Proposition 2 is that as the player's skillset expands, he becomes more selective in his skill acquisition: the marginal value of acquiring x is smaller, and thus less likely to exceed the player's best alternative action, when the player has a larger skillset.

3.1 Path Trees

For the rest of the analysis, we will impose some fairly restrictive regularity conditions.

Definition 4. (X, Z) is called a path tree if all skills x (except the root) have degree 1.

In a path tree, each skill (except the root) has exactly one adjacent skill. X then consists of a finite number of paths emanating from the root x_0. Figure 4 depicts an example of a path tree.

Here is some notation for use when X is a path tree. In that setting, each skill x is uniquely identified by (i) the path that the skill lies on, and (ii) the depth $d(x)$ of the skill. This allows us to use $x_{j,d}$ to denote the skill that is on path $j \in \{1, ..., k\}$ and has depth d. We use d_t^j to denote the level of the skillset Ψ_t on path j, at the beginning of period t.

Path trees are relatively simple structures, but nevertheless demonstrate two crucial dimensions of a skill's position in the knowledge structure: the depth of a skill, which allows us to discuss the distance of the skill from the player's existing skillset, and the path that the skill is on, which allows us to make the point that skill acquisition on one branch of knowledge does not facilitate skill acquisition on other branches.

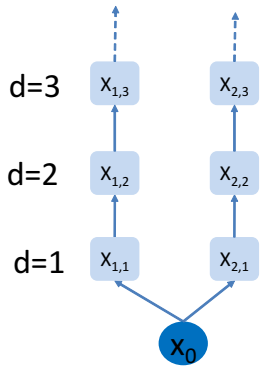


Figure 4: *A path tree with two paths*

We will assume for most of the results of the paper that (X, Z) is a path tree, and that the exploitation value of each skill i linearly increasing in depth.

Assumption 1. *(i) (X, Z) is a path tree. (ii) The exploitation value of each skill $x_{i,d}$ is a linear function of its depth: $\pi(x_{i,d}) = \alpha_i d$ where $\alpha_i > 0$ for all paths i.*

The restrictions to a path tree structure and linearly increasing exploitation value are maintained to preserve tractability. I conjecture that they can be significantly relaxed while preserving the results in this section. Nonetheless, they preserve some degree of flexibility in the model.

4 Specialization

This section investigates qualitative features of the player's skillset over the long run. The main result is that specialization is, under certain assumptions, the inevitable outcome. First, we define specialization.

Definition 5. Assume that $|\Psi^*_\infty| = \infty$. We say that the player specializes if there exists exactly one path $x_0 \to x_1 \to x_2 \to ...$ such that $\{x_0, x_1, x_2, ...\} \subset \Psi^*_\infty$.

This definition of specialization has an intuitive interpretation: in the long run, there is exactly one infinite path that is completely acquired by the player. The player may acquire skills on other paths, but skill acquisition on any other path stops eventually. Note that this definition of specialization is only meaningful if $|\Psi^*_\infty| = \infty$: a skillset of finite size cannot satisfy the condition.

We may now present Proposition 3, which states that specialization is inevitable in learning environments where knowledge structure is regular.

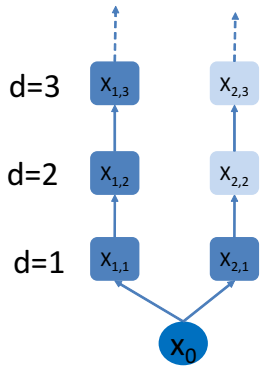


Figure 5: *Skillset Specialized on Path 1*

Proposition 3. *Suppose Assumption 1 holds and that $\alpha_i \equiv \alpha > 0$ for all paths i. Given α, there exists $\bar{c}$ such that for $c < \bar{c}$, the player specializes almost surely.*

At any history, the player does not necessarily know which path he will specialize on, as this may depend on the realizations of available skills in subsequent periods. In this sense, there is history dependence in the details of specialization, even though specialization as a qualitative outcome is inevitable. For an example (under alternative assumptions) where the structure of the player's skillset is history dependent in a qualitative sense, see Section 5.1.

5 Complementary Skills

The setting we have explored so far implies that skills are noncomplementary: the exploitation value of a skill does not depend on the player's knowledge of other skills. Such an analysis neglects that knowledge is often complementary. Some opportunities require multiple types of knowledge, each of which involves a distinct incremental acquisition path. For example, an economist should understand the theoretical as well as empirical literature in his field to perform useful research. Similarly, an entrepreneur has to be competent at both the technical and managerial aspects of running a business to efficiently exploit the business opportunities that he encounters.

To succinctly model complementarity, we define a specific restriction on the structure of (X, Z).

This section augments the setting of Section 2 by assuming that every skill x has an associated set $\chi(x)$ of complementary skills. A skill x may only be exploited in period t if it is available and if $x \cup \chi(x) \subset \Psi_t$: in other words, the player had previously acquired x as well as all skills complementary to x.

The learning environment is then defined by the objects $(X, Z, \pi(\cdot), \chi(\cdot))$. Note that the setting

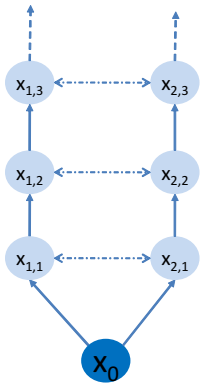


Figure 6: *Complementary knowledge structure*

from Section 2 emerges as a special case when we assume that $\chi(x) \equiv \varnothing$, in which case we can describe the learning environment as $(X, Z, \pi(\cdot))$, as before.

For the rest of this section, we maintain Assumption 1, so that (X, Z) is a path tree with two paths and exploitation payoffs are linearly increasing in depth (and equal across both paths). As previously noted, we may index each skill as $x_{j,d}$ where $j \in \{1, 2\}$ refers to the path the skill lies on and $d \geq 1$ is the depth of the skill.

To capture skill complementarity in this setting, we assume that each skill x has a single complementary skill,

$$\chi^{compl}(x_{i,d}) \equiv \{x_{-i,d}\}$$

where $-i = \{1, 2\} \setminus i$, on the other path This knowledge structure reflects the premise that skills on the two paths are complementary to each other: a skill of depth d cannot be exploited unless the player has acquired skills of depth d on both skill paths. Figure 6 depicts a complementary knowledge structure.

Intuitively, complementarity between skills should induce the player to adopt a balanced acquisition strategy. The next proposition formalizes this intuition.

Proposition 4. *Assume that Assumption 1 holds with $\alpha_i \equiv \alpha > 0$ for all paths i, and that $\chi = \chi^{compl}$. Then there exists $m < \infty$ such that almost surely,*

$$\lim_{t\to\infty} d_t^1 = \lim_{t\to\infty} d_t^2 = m.$$

In addition to highlighting the player's balanced acquisition, Proposition 4 also points out that the player almost always ceases acquisition once the levels of both paths reaches some value m.

To understand this, assume without loss of generality that $d_t^1 \geq d_t^2$ (so that the level of path 1 is

weakly greater than on path 2) in period t. Consider the value to the player of acquiring x_{1,d_t^1+1} on path 1. x_{1,d_t^1+1} can be exploited only after x_{2,d_t^1+1} is also acquired. Consider the player's actions in subsequent periods. Because the player's acquisition rate slows as his skillset grows large, for large t, $x_{d_t^1+1}$ will likely not be exploited for a long time after x_{1,d_t^1+1} is acquired. The expected present value of acquiring x_{1,d_t^1+1} is thus small, and is outweighed by the opportunity cost c when $|\Psi_t|$ is large. As a result, acquisition never takes place on the weakly deeper path (and thus eventually ceases on both paths, since an increase in the depth of the skillset always entails acquisition on the weakly deeper path).

The reader may find the result that knowledge acquisition always ceases eventually to be undesirable. In Appendix B, we present a setting with a weaker notion of complementarity where knowledge acquisition may continue indefinitely.

5.1 Hybrid Model

This section presents a knowledge structure with both stand-alone skills and complementary skills. Here, if the player focuses skill acquisition on only one path, he can exploit stand-alone skills, while if he acquires skills on both paths, he can exploit complementary skills.

Hybrid Knowledge Structure As before, we maintain assumption 1 so (X, Z) is a path tree and exploitation payoffs are linearly increasing in depth. We make the following additional assumptions. First, the paths differ in their need for complementary skills: skills on path 1 require a complementary skill of equal depth on path 2, while skills on path 2 can be exploited without :

$$\chi^{hybrid}(x_{1,d}) = \{x_{2,d}\},$$
$$\chi^{hybrid}(x_{2,d}) = \varnothing.$$

Also, the payoff function of each skill is linearly increasing[2] in depth, but differs between the two paths:

$$\pi(x_{1,d}) = \alpha d(x),$$
$$\pi(x_{2,d}) = \widetilde{\alpha} d(x).$$

with $\alpha > \widetilde{\alpha} > 0$. So, compared to skills on path 2, skills on path 1 require complementary skills, but have higher exploitation payoffs. This set of assumptions is depicted in Figure 7.

[2]We can generalize to the case of regularly increasing payoffs easily, but at some cost to expositional convenience.

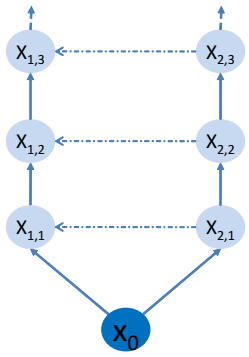


Figure 7: *Hybrid knowledge structure*

This set of assumptions captures an environment where a single player acquires and then subsequently exploits skills, possibly in combination with complementary skills. This environment is similar of the model of entrepreneurship in Lazear (1995). In that model, each agent is endowed with skills on two dimensions: $(x_1, x_2) \in \mathbb{R}^2$. The agent may then choose to become an employee specializing in one skill and obtain a payoff of $\max\{x_1, x_2\}$, or become an entrepreneur and obtain a payoff of $\lambda \min\{x_1, x_2\}$ with $\lambda > 1$. This assumption on payoffs is in the same spirit as the assumption in the present model that $\alpha > \widetilde{\alpha}$.

5.1.1 Specialization and Balance: Lock-in

We may show that certain conditions may trigger long-run specialization or long-run balance in the player's skill acquisition.

Lemma 2. *Assume that Assumption 1 holds, that (X, Z) is a path tree with two paths, that $\chi = \chi^{hybrid}$, and that $\alpha > \widetilde{\alpha}$. There exists y such that, at any history h_t, if*

$$\widetilde{\alpha} d_t^2 - \alpha d_t^1 \geq y,$$

then $d_\tau^1 = d_t^1$ for all $\tau \geq t$ in the continuation game of the optimal strategy φ^*.

Lemma 2 states that if, in any period, the exploitation payoff of the deepest accessible skill on path 2 far exceeds the exploitation payoff of the deepest accessible skill on path 1, the player will thereafter stop acquiring skills on path 1. In other words, the player thereafter exclusively specializes acquisition on path 2.

The reasoning behind Lemma 2, which posits the existence of a "specialization trap," is analogous to the reasoning behind Proposition 3. The condition $\widetilde{\alpha} d_t^2 - \alpha d_t^1 \geq y$ ensures that the level of path

2 is much higher than the level of path 1, and that the payoff of the deepest exploitable skill on path 2 far exceeds the payoff to the deepest exploitable skill on path 1. As in Proposition 3, this implies that the marginal present value of acquiring a skill on path 1 is so small that acquisition is not worthwhile.

Lemma 3. *Assume that Assumption 1 holds, that (X, Z) is a path tree with two paths, that $\chi = \chi^{hybrid}$, and that $\alpha > \widetilde{\alpha}$. There exist $y > 0$ and $y' > 0$ such that, at any period t, if*

$$\alpha d_t^1 - \widetilde{\alpha} d_t^2 \geq y$$

and

$$d_t^2 - d_t^1 > y',$$

then $d_\tau^2 - d_\tau^1 \leq d_t^2 - d_t^1$ for all $\tau \geq t$ in the continuation game of the optimal strategy φ^* and, almost surely, $\lim_{t\to\infty} d_\tau^1 = \lim_{t\to\infty} d_\tau^2 < \infty$.

Loosely speaking, Lemma 3 states that, when the exploitation payoff of the deepest acquired skill on path 1 far exceeds the exploitation payoff of the deepest acquired skill on path 2, the player will thereafter ensure that his skillset is balanced.

To see why this result holds, assume that the conditions in Lemma 3 are satisfied. Consider the value of acquiring x_{2,d_t^2+1} on path i. $\alpha d_t^1 - \widetilde{\alpha} d_t^2 \geq y$ ensures, for sufficiently large y, that x_{2,d_t^2+1} will rarely be chosen in subsequent periods. Of course, x_{2,d_t^2+1} is also valuable because it is necessary for exploiting x_{1,d_t^2+1}. However, because $d_t^2 - d_t^1 > y'$, x_{1,d_t^2+1} will not be accessible for at least the next y' periods. Thus the present value of acquiring on path 1 is low when y and y' are large, in which case the player optimally chooses not to acquire on path 1.

5.1.2 History Dependence

Lemmas 2 and 3 suggest that specialization and balance are possible outcomes. It turns out that they are the only two possible long-run outcomes, as proposition 5 shows.

Proposition 5. *Assume that Assumption 1 holds, that (X, Z) is a path tree with two paths, that $\chi = \chi^{hybrid}$, and that $\alpha > \widetilde{\alpha}$.*

- Almost surely, the player either specializes or has $\lim_{t\to\infty} d_t^1 = \lim_{t\to\infty} d_t^2 < \infty$.
- There exists $(\alpha, \widetilde{\alpha}, c)$ such that (i) the player specializes with positive probability and has $d_\infty^1 = d_\infty^2 < \infty$ with positive probability.

The implication of Proposition 5 is that the player inevitably falls into either a specialization routine or a balance routine. The intuition is as follows. For each period t, let $\eta_t = \alpha d_t^1 - \widetilde{\alpha} d_t^2$. When η_t is large and positive, the most lucrative accessible skills require complementary skills; when η_t is large and negative, the most lucrative accessible skills are stand-alone. Lemmas 2 and 3 show that there exist upper and lower thresholds for η_t beyond which specialization or balance become inevitable. For intermediate values, η_t follows a random walk. The key is to show that this random walk is not mean reverting, so that $|\eta_t|$ eventually hits one of the two thresholds.. Specifically, we may show that an acquisition on path i increases the marginal value of further acquisition on path i. As a result, the larger is η_t, the greater the likelihood (informally speaking) that the next increment is upward rather than downward. This then implies that the random walk will inevitably attain one of the thresholds. Which threshold is attained depends on the initial draws of available skills.

6 Conclusion

This paper builds a dynamic model of skill acquisition that highlights the role of knowledge structure in learning outcomes. It suggests that history dependence may play a role, not just in terms of the specific skills an individual acquires, but also in terms of the structure of the acquired skillset. This raises the hypothesis that specialists and generalists do not become so simply because of innate differences, either in endowed ability or preferences, but because of the opportunities they encounter early in life.

The analytical framework is very flexible in the forms of knowledge structure that it permits. This is both a feature and a bug, as flexibility makes empirical predictions difficult to pin down. To make progress, the present paper makes many of its points by discussing a series of specific examples. An improved framework would tie these examples together and permit more comprehensive comparative static analysis.

In Section 5, we introduced the possibility of complementarities between skills by assuming that some skills require complementary skills before they can be exploited. This captures the notion of complementarities in exploitation. Another possibility is that there are complementarities in acquisition: to acquire a given skill x, the player has to have previously acquired a set $\widehat{\chi}(x)$ of associated skills with $|\widehat{\chi}(x)| > 1$. This possibility is not discussed in the present paper and would require us to move away from the assumption that (X, Z) is a tree.

This paper also ignores the possibility of multi-player learning. Extending the analysis to multiple players would require thinking about how knowledge can be transferred across players, and how agents can skillset existing knowledge to cooperate in exploiting skills.

References

Aghion, P., Harris C., and Vickers J. (1997) "Competition and Growth with Step-by-Step Innovation; An Example." *European Economic Review* 41: 771-782.

Acemoglu, D. (2002) "Directed Technical Change." *Review of Economic Studies*, 69: 781-810.

Arthur, B. (1989) "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." *The Economic Journal*, 99(394): 116-131

Cohen, W. and Levinthal, D.A. (1990) "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35(1): 128-152

Doraszelski, U. (2004) "Innovations, Improvements, and the Optimal Adoption of New Technologies." *Journal of Economic Dynamics and Control* 28(7): 1461-1480

Jones, B. (2009) "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" *Review of Economic Studies* 76(1): 283-317

Jovanovic, B. and Nyarko, Y. (1996) "Learning by Doing and the Choice of Technology." *Econometrica* 64(6): 1299-1310

Lazear, E. (2005) "Entrepreneurship." *Journal of Labor Economics* 23(4): 649-681

March, J. (1991) "Exploration and Exploitation in Organizational Learning." *Organization Science* 2(1): 71–87.

Selove, M. (2009) "How Do Firms Become Different? A Dynamic Model." Working Paper.

Sobel, J. (2000) "Economists' Models of Learning." *Journal of Economic Theory*, 94: 241-261.

A Appendix: Correlated Availability

In the rest of the paper, we assumed that availability is independent across skills. In many environments, it may be more reasonable to assume that the availability of similar skills is positively correlated. Consider, for example, a law firm.[3] Some law firms specialize in corporate work, while others specialize in litigation. Many others take on both types of work. We may think of a law firm as operating within our learning environment, gaining expertise as it takes on successively more complex cases. However, the opportunities available to law firms tend to be correlated in a particular way. In normal times, more corporate work is available, while litigation work tends to be more available during economic downturns. That is, availability of skills on the same path is positively correlated.

We may modify our model to incorporate this correlation. We return to the setting of Section 2 where $\chi(\cdot) \equiv \varnothing$, so that there are no complementarities between skills. To simplify matters, consider a path tree with 2 paths. We may think of the skills required for corporate work and for litigation work as being on separate paths, as expertise does not easily transfer between the two types of work.

There is a series of i.i.d. binary random variables $\{\mu_1, \mu_2, ...\}$ with each $\mu_t \in \{1, 2\}$ and i.i.d. $P(\mu = 1) = \frac{1}{2}$. In each period t and for each skill $x_{j,d}$, the conditional probability that $x_{j,d} \in X_t$ is

$$P(x_{j,d} \in \Phi_t | \mu_t) = \begin{cases} p + \epsilon : j = \mu_t \\ p - \epsilon : -j = \mu_t \end{cases} \quad \text{for } \epsilon > 0.$$

Conditional on μ_t, each event $x_{j,d} \in \Phi_t$ is independently distributed.

Finally, assume that payoffs are constant, so $\pi(x) \equiv 1$. This assumption is not essential but clarifies the exposition.

This specification generates positive correlation in availability of skills on the same path, and negative correlation in availability of skills on different paths, as intended. At the same time, the unconditional probability that $x_{j,d} \in \Phi_t$ remains equal to p. We call this set of assumptions the *correlated-availability setting.*

The following proposition is the main result of this section.

Proposition 6. *Consider the correlated-availability setting. Assume that the player plays optimally and randomizes uniformly between any choices between which he is indifferent. Assume that* $\epsilon > 0$. *Then, under* φ^*, $E\left[\frac{\left(d_t^1 - d_t^2\right)^2}{d_t^1 + d_t^2}\right] < 1$ *at every period* t .

[3] To fit the present single-player model, think of the law firm as a monolithic entity.

Proposition 6 states that the player's skillset is, in distribution, more balanced under the correlated-availability setting than under the no-correlation setting. This is due to two effects, which both operate in the same direction. First, for given $|\Psi_t|$, a more balanced skillset is more likely to yield an available and accessible skill. Specifically, given $\left(d_t^1, d_t^2\right)$, the probability that at least one skill is available and accessible for exploitation in period t is

$$\begin{aligned}
&\Pr\left[\exists x \in \Phi_t \cap \Psi_t\right] \\
&= 1 - \left(\frac{1}{2}\left(1-(p+\epsilon)\right)^{d_t^1}\left(1-(p-\epsilon)\right)^{d_t^2} + \frac{1}{2}\left(1-(p+\epsilon)\right)^{d_t^2}\left(1-(p-\epsilon)\right)^{d_t^1}\right) \\
&= 1 - \left(\frac{1}{2}\left(\left(1-(p+\epsilon)\right)\left(1-(p-\epsilon)\right)\right)^{\frac{d_t^1+d_t^2}{2}}\left(\begin{array}{c}\left(1-(p+\epsilon)\right)^{\frac{\left|d_t^1-d_t^2\right|}{2}} + \left(1-(p-\epsilon)\right)^{-\frac{\left|d_t^1-d_t^2\right|}{2}} \\ + \left(1-(p+\epsilon)\right)^{-\frac{\left|d_t^1-d_t^2\right|}{2}} + \left(1-(p-\epsilon)\right)^{\frac{\left|d_t^1-d_t^2\right|}{2}}\end{array}\right)\right),
\end{aligned}$$

which is decreasing in $\left|d_t^1 - d_t^2\right|$. Thus balanced skillsets are more profitable. Given a choice, the player would rather acquire on the shallower path than on the deeper path as the former choice results in a more balance skillset.

A second effect appears when exploitation is more valuable than acquisition. In this case, the player will acquire a new skill only if no previously acquired skills are available for exploitation. The unconditional probability that the accessible unacquired skill on either path is available is the same for both paths. However, for the deeper path i, this occurs more often when at least one previously acquired skill is available for exploitation, because availability of the unacquired skill on path i is correlated with availability of previously acquired skills on path i (of which there are more than path j). As a result, the player is less likely to acquire on path i even if he has the opportunity to do so.

B Appendix: Weak Complementarity

In this section, we present a set of assumptions that capture a weaker notion of complementarity than the setting of Section 5. We show that, under these assumptions, knowledge acquisition may continue indefinitely.

As in Section 5, we assume that (X, Z) is a path tree with two paths. As previously stated, we may index each skill as $x_{j,d}$ where $j \in \{1, 2\}$ refers to the path the skill lies on and $d \geq 1$ is the depth of the skill. We also assume that payoffs are increasing at a weakly-faster-than-linear rate with depth: $\pi(x) \equiv \alpha d(x)^{\beta}$ and $\beta \geq 1$.

Each skill x has a single complementary skill

$$\chi^{WC}(x_{i,d}) \equiv \{x_{-i,d-1}\}$$

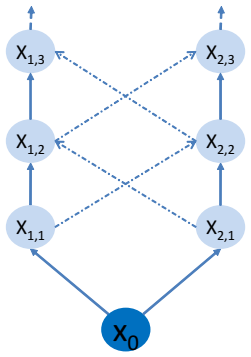


Figure 8: *Weak Complementarity*

where $-i = \{1, 2\} \setminus i$. Here, a skill of depth d cannot be exploited unless the player has acquired a skill of depth $d-1$ on the other path. We identify this set of restrictions on $(X, Z, \pi(\cdot), \chi(\cdot))$ as a *weakly complementary knowledge structure*, and depict it in Figure 8.

Under both strong and weak complementarity, skills are complements in the following sense: a skill with depth d cannot be exploited unless the player's level on the other path is also close to d. However, strong complementarity imposes a stricter condition on the required relative skillset levels. Under strong complementarity, a skill with depth d cannot be exploited unless the other path also has level d. On the other hand, under weak complementarity, a skill with depth d is useful for exploitation as long as the level of the other path is close to d (in our formulation, the depth has to be at least $d-1$).

Proposition 7. *Assume that Assumption 1 holds with $\alpha_i \equiv \alpha > 0$ for all paths i, and that $\chi = \chi^{WC}$. Then $\limsup_{t\to\infty} \left|d_t^1 - d_t^2\right| = 1$, and there exists $\underline{c}$ such that for $c < \underline{c}$, $|\Psi_\infty^*| = \infty$ almost surely.*

Proposition 7 states that skill acquisition is, over the long run, balanced under weak complementarity: for large $|\Psi_t|$, the difference in levels of the two paths never exceeds one. However, in contrast to strong complementarity, skill acquisition may continue indefinitely under weak complementarity. Under weak complementarity, as long as $d_t^1 \leq d_t^2 + 1$, x_{1,d_t^1+1} can be exploited immediately after it is acquired. The marginal value of acquiring a new skill thus does not vanish as $|\Psi|$ becomes large. This ensures that when opportunity cost c is small, skill acquisition never ceases completely, since at least one of $d_t^1 \leq d_t^2 + 1$ and $d_t^2 \leq d_t^1 + 1$ holds in every period t.

C Appendix: Proofs

Some notation:

Let Λ_t be the set of all skills $\{x : x \in \Psi_t \text{ or } x \leftarrow x' \in \Psi_t\}$ that can be exploited or acquired given the player's skillset in period t, and let Λ_t^a be the set of skills $\{x : x \leftarrow x' \in \Psi_t\}$ that the player may acquire in period t. We may sometimes write Λ_t (resp. Λ_t^a) as $\Lambda(\Psi_t)$ (resp. $\Lambda^a(\Psi_t)$) to emphasize that Λ_t is uniquely determined by Ψ_t.

Define $d(\Psi) = \arg\max\{d(x) : x \in \Psi\}$ to be the depth of the skillset Ψ. On a path tree, define $d^j(\Psi) = \arg\max_d \{x_{j,d} : x_{j,d} \in \Psi\}$ to be the depth of the deepest skill on path j in skillset Ψ. Equivalently, we say that $d^j(\Psi)$ *is the level of skillset* Ψ *on path* j. If the player acquires a skill that lies on path j in period t, we say that the player *acquires on* j.

Proposition 1 Assume that x' and x'' are both adjacent to Ψ. If $x' \succeq x''$, then $V(\Psi \cup x') \geq V(\Psi \cup x'')$.

Proof. First, some preliminaries. Recall that a history h_t takes the form

$$h_t = \{\Phi_1, x_1, ..., \Phi_{t-1}, x_{t-1}, \Phi_t\},$$

where Φ_τ is the set of available skills in period τ and x_τ is the choice of action in period τ. A strategy is then a mapping from the set of histories to the set of actions, with the additional condition that the action in period t must be both available and accessible:

$$x_t \in (\Psi(h_t) \cup \Lambda(\Psi(h_t))) \cap \Phi_t \text{ for all } t \text{ and all } h_t,$$

where $\Psi(h_t)$, as usual, equals $\cup_{\tau=1}^t x_\tau \cup \Psi$.

Let us call the realizations of available skills in the first t periods the exogenous history

$$g_t = \{\Phi_1, ..., \Phi_{t-1}, \Phi_t\}.$$

Then we can we restrict attention to strategies that condition the action in period t only on Ψ_t and Φ_t (and not on any randomizing device). In particular, we can denote the action taken at history g_t and under strategy φ as $a(g_t, \varphi)$, and the skillset at history g_t under strategy φ as $\Psi(g_t, \varphi)$.

Let G_t be the set of all possible t-period exogenous strategies g_t, and let G_∞ be the set of all untruncated histories

$$g_\infty = \{\Phi_1, \Phi_2, ...\}.$$

Note that each G_t admits a natural measure m, in the following sense. If we identify the nodes of (X, Z) as $\{x^1, x^2, ...\}$ then we can characterize each Φ_t in the form $\{s^1, s^2, ...\}$ where $s^k \in \{0, 1\}$ equals one if and only if skill x^k is available in period t. Then we can define

$$\mu\left(\widehat{G}_t\right) \equiv \Pr[g_t \in G];$$

for all $\widehat{G}_t \subset G_t$, so $\mu\left(\widehat{G}_t\right)$ is the probability that a random realization of $\{\Phi_1, \Phi_2, ..., \Phi_t\}$ lies in $\widehat{G}_t$. We then say that a map f_t is measure-preserving if $\mu\left(f_t\left(\widehat{G}_t\right)\right) \equiv \mu\left(\widehat{G}_t\right)$ for all $\widehat{G}_t \subset G_t$.

Next, note that

$$V_\varphi(\Psi) = \sum_{t=0}^{\infty} \int_{g \in G_t} \pi(\varphi(g), \Psi(g)) \mu(dg);$$

in words, the value function at a skillset for a given strategy equals the net present value of exploitation following that strategy.

Our assumption that $x' \succeq x''$ means that there exists a isomorphism $\varsigma : X' \to \Pi(x'')$ (that preserves the tree structure) such that $\pi(\varsigma(x)) \leq \pi(x)$ for all $x \in X'$ for some $X' \subset \Pi(x')$. Given ς, define the binary relation $\succcurlyeq$ acting on subsets of X as follows: $\Psi \succcurlyeq \Psi'$ if and only if there exists an bijection $\omega : \Psi' \to \Psi$ such that for each $x \in \Psi'$, either $\omega(x) = x$ or $\omega(x) = \varsigma^{-1}(x)$. In words, this means that we can construct $\Psi \succcurlyeq \Psi'$ from Ψ' by switching some subset X' of Ψ' with $\varsigma^{-1}(X')$. Note that, by definition, $\Psi \cup x' \succcurlyeq \Psi \cup x''$.

Proof Strategy Consider $x', x'' \in X$ with $x' \succ x''$. We will show that for any strategy $\varphi(\cdot)$, there exists another strategy $\widetilde{\varphi}(\cdot)$ such that $V_{\widetilde{\varphi}}(\Psi \cup x') < V_{\varphi}(\Psi \cup x'')$. This would prove our proposition, since it would immediately imply that the optimal strategy cannot have $a(\Psi, \Theta) = x''$.

We will construct $\widetilde{\varphi}(\cdot)$ and prove that $V_{\widetilde{\varphi}}(\Psi \cup x') > V_{\varphi}(\Psi \cup x')$ in the following way.

1. For each t, introduce a measure-preserving bijection f_t on G_∞, with the property that $\Psi(f_t(g_t)) \succcurlyeq \Psi(g_t)$ (weakly).
2. Contruct a mapping $\widehat{\varphi}$ from exogenous histories to actions, and define $\widetilde{\varphi}(g) \equiv \widehat{\varphi}(f_t(g))$.
3. Show that for any exogenous history $g \in G_t$, the stage payoff under $\widetilde{\varphi}$ at history $f_t(g)$ is weakly larger than (and for some g with positive measure, strictly larger than) the stage payoff under φ at history g:
$$u(\widetilde{\varphi}(f_t(g)), \Psi(g)) \geq u(\varphi(g), \Psi(g))$$
4. Then conclude that
$$\begin{aligned} V_{\widetilde{\varphi}}(\Psi) &= \sum_{t=0}^{\infty} \int_{h \in H_t} \pi(\widehat{\varphi}(f_t(g)), \Psi(f_t(g)))\, df_t(g) \\ &\geq \sum_{t=0}^{\infty} \int_{h \in H_t} \pi(\varphi(g), \Psi(g))\, df_t(g) \\ &= \sum_{t=0}^{\infty} \int_{h \in H_t} \pi(\varphi(g), \Psi(g))\, dg \text{ since } f_t \text{ is measure-preserving} \\ &= V_{\varphi}(\Psi). \end{aligned}$$

Details We now proceed with the construction of f_t and $\widehat{\varphi}$. We start with f_t. We will prove that f_t, as we will construct, is a measure-preserving bijection by induction: assume that f_{t-1} is a measure-preserving bijection.

Consider
$$g_t = \{\Phi_1, \Phi_2, ..., \Phi_t\}.$$
This induces actions $\{x_\tau\}_{\tau=1}^t$ where, defined recursively, $x_\tau \equiv \varphi(\Phi_1, x_1, ..., x_{\tau-1}, \Phi_\tau)$, and $\Psi_\tau = \cup_{\tau=1}^t x_\tau \cup \Psi$. Note that, by assumption, $x_1 = x''$.

We now construct the following:

- a finite series $\{\Psi'_\tau\}_{\tau=1}^t$ with such that (i) $\Psi'_\tau \equiv \Psi'_{\tau-1} \sqcup x'_\tau$ for some $x'_\tau \in X$, and (ii) $\Psi'_\tau \succcurlyeq \Psi_\tau$ for all $1 \leq \tau \leq t$.
- a measure-preserving bijection f_t on G_t such that $\Psi(f_t(g_t)) \equiv \Psi'_t$.
- A strategy (a mapping $\widetilde{\varphi}_t$ from G_t to X satisfying the constraint that only available and accessible skills are chosen) with the properties that (i) $\widetilde{\varphi}_t(f_t(g_t)) \in \Lambda(\Psi(f_t(g_t)))$ if and only if $\varphi_t(g_t) \in \Lambda(\Psi(g_t))$, (ii) $\pi(\widetilde{\varphi}(f_t(g)), \Psi(g)) \geq \pi(\varphi(g), \Psi(g))$.

We proceed by induction; assume that the properties hold for $1 \leq \tau \leq t-1$; so, in particular, $\Psi'_t \succcurlyeq \Psi_t$. Define, with $A(\Psi) \equiv \Psi \cup \Lambda(\Psi)$,

$$\chi(x) \equiv \begin{cases} x : x \in A(\Psi'_t) \cap A(\Psi_t) \wedge x \notin A(\Psi'_t) \cup A(\Psi_t) \\ \omega(x) : x \in A(\Psi_t) \setminus A(\Psi'_t) \\ \omega^{-1}(x) : x \in A(\Psi'_t) \setminus A(\Psi_t) \end{cases}, \text{ and}$$
$$\widetilde{\chi}(\Phi) \equiv \{\chi(x) : x \in \Phi\}.$$

and let

$$f_t(\Phi_1, ..., \Phi_{t-1}, \Phi_t) = f_{t-1}(\Phi_1, ..., \Phi_{t-1}) \sqcup \chi(\Phi_t),$$
$$\Psi'_t = \Psi'_{t-1} \sqcup \chi(x_t).$$

Before we construct $\widetilde{\varphi}_t$, note that $\chi(\cdot)$ is a bijection on X, which implies that $\widetilde{\chi}$ is a bijection on the set B of all possible Φ_t and thus (since $G_t = G_{t-1} \times B$, and since we assume that f_{t-1} is a bijection on G_{t-1}) that f_t is a bijection on G_t. Note, also, that $\chi(\cdot)$ is constructed to be measure-preserving.

Finally, define $\widetilde{\varphi}_t(f_t(\Phi_1, ..., \Phi_{t-1}, \Phi_t)) \equiv \chi(\varphi(\Phi_1, ..., \Phi_{t-1}, \Phi_t))$. Note the following properties of $\widetilde{\varphi}_t$. First, $\widetilde{\varphi}_t(f_t(g_t)) \in \Lambda(\Psi(f_t(g_t)))$ if and only if $\varphi_t(g_t) \in \Lambda(\Psi(g_t))$. Second, $\widetilde{\varphi}_t(f_t(g_t))$ is an acquisition (resp. exploitation) if and only if $\varphi(g_t)$ is an acquisition (exploitation). Third, whenever $\varphi_t(g_t)$ is an exploitation, $\pi(\widetilde{\varphi}_t(f_t(g_t))) \geq \pi(\varphi_t(g_t))$.

Since we have verified steps 1 through 4, the result holds.

□

Lemma 4. *Suppose Assumption 1 holds and that $\alpha_i \equiv \alpha > 0$ for all paths i. Then given α and c, there exists a constant Δ_{stop} such that if $d_t^i - d_t^j \geq \Delta_{stop}$ for any i, j and for some t, then $d_\tau^i = d_t^i$ for $\tau \geq t$.*

Proof. Denote $d^i(h_t)$ to be the level of path i at history h_t. Without loss of generality, consider a history h_t^0 where $d^1(h_t^0) - d^2(h_t^0) \geq \Delta_{stop}$. For notational convenience, let us denote $d^1 = d^1(h_t^0)$ and $d^2 = d^2(h_t^0)$; and assume that there are only two paths (the proof generalizes trivially). Consider a strategy φ^* such that $\varphi^*(h_t^0) = x_{2,d_t^2+1}$, so that the player acquires on path 2 at history h_t^0. We will construct an alternative strategy $\widetilde{\varphi}$ such that $V_{\widetilde{\varphi}}(h_t^0) > V_{\varphi^*}(h_t^0)$; continuation payoff at h_t^0 is higher for $\widetilde{\varphi}$ than for φ^*.

First, some notation. Given a sequence $\Phi_\infty = \{\Phi_1, \Phi_2, ...\}$ of realizations of available skills. a strategy φ^* generates a sequence of actions $\{a_1, a_2, ...\}$ where

$$a_t = \varphi(g_t).$$

We call a_t *the action induced in period t by $g_t = \{\Phi_1, \Phi_2, ..., \Phi_t\}$ under φ*. Also, let $P_{\varphi,h_t}(\zeta)$ be the probability in period τ that event ζ occurs, conditional on h_t occurring and the player subsequently playing φ. Also, let $\Psi_\tau(g_\tau, \varphi^*)$ be the skillset induced by φ^* at history g_t; whenever convenient, we will write $\Psi_\tau(g_\tau, \varphi^*) = \Psi_\tau$ and suppress the dependence on (g_τ, φ^*).

Now, consider the alternative strategy $\widetilde{\varphi}$ where, for any sequence $\{\Phi_{t+1}, \Phi_{t+2}, ..., \Phi_\tau\}$ of realizations of available skills subsequent to h_t^0, the player imitates φ^* except that it never chooses $a_\tau \in \Upsilon = \left\{x_{2,d_t^2+1}, x_{2,d_t^2+2}, ...\right\}$; whenever the induced action under φ^* at an exogenous history g_t is $a_\tau^* \in \left\{x_{2,d_t^2+1}, x_{2,d_t^2+2}, ...\right\}$, the player instead chooses under $\widetilde{\varphi}$ to exploit the skill with the highest payoff that is available and exploitable. That is,

$$\widetilde{\varphi}(g_t) = \begin{cases} \varphi^*(g_t) : \varphi^*(g_t) \notin \Upsilon \\ \arg\max \pi(\Psi_\tau(g_\tau, \varphi^*) \cap \Phi_\tau \setminus \Upsilon) : \varphi^*(g_t) \in \Upsilon \end{cases}$$

We seek to show that, for sufficiently large Δ_{stop}, $V_{\widetilde{\varphi}}\left(h_t^0\right) > V_{\varphi^*}\left(h_t^0\right)$.

The continuation value $V_{\widetilde{\varphi}}\left(h_t^0\right)$ equals the expected discounted value of all exploitative actions subsequent to h_t^0, given that the player subsequently plays $\widetilde{\varphi}$; we may write this as

$$V_{\widetilde{\varphi}}\left(h_t^0\right) = \left(\sum_{i=1,2} \sum_{d=1}^{\infty} \sum_{\tau=t}^{\infty} \delta^{\tau-t} \pi\left(x_{i,d}\right) P_{\widetilde{\varphi},h_t^0}\left(a_\tau = x_{i,d} \in \Psi_\tau\right) \right)$$

where $P_{\widetilde{\varphi},h_t}\left(a_\tau = x_{i,d} \in \Psi_\tau\right)$ is the probability that, under strategy $\widetilde{\varphi}$, and in the continutation equilibrium following h_t^0, (i) action a_τ is taken in period τ, and (ii) a_τ is exploitative rather than acquisitive.

Note that, by definition, $\widetilde{\varphi}$ behaves identically to φ^* except at those exogenous histories g_τ where φ^* prescribes an action in $\Upsilon = \left\{x_{2,d^2+1}, x_{2,d^2+2}, ...\right\}$; in that case, $\widetilde{\varphi}$ instead plays the highest-value action that is exploitable and available (i.e. in $\Psi_\tau \cap \Phi_\tau \setminus \Upsilon$). With this in mind, we can rewrite $V_{\widetilde{\varphi}}\left(h_t^0\right)$ as

$$\begin{aligned}
&V_{\widetilde{\varphi}}\left(h_t^0\right) \\
&= \sum_{d=1}^{\infty} \sum_{\tau=t}^{\infty} \delta^{\tau-t} \pi\left(x_{1,d}\right) P_{\varphi^*,h_t}\left(a_\tau = x_{1,d} \in \Psi_\tau\right) \\
&+ \sum_{d=1}^{d^2+1} \sum_{\tau=t}^{\infty} \delta^{\tau-t} \pi\left(x_{2,d}\right) P_{\varphi^*,h_t}\left(a_\tau = x_{2,d} \in \Psi_\tau\right) \\
&+ \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \delta^{\tau-t} \pi\left(x_{1,d}\right) P_{\varphi^*,h_t}\left(a_\tau = x_{1,d} \notin \Psi_\tau\right) \\
&+ \sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{x \in \Psi_\tau \cap \Phi_\tau \setminus \Upsilon} \delta^{\tau-t} \pi\left(x\right) P_{\varphi^*,h_t}\left(\begin{array}{c} a_\tau = x_{2,d} \vee \\ \pi\left(x\right) = \max \pi\left(\Psi_\tau \cap \Phi_\tau \setminus \Upsilon\right) \end{array} \right) \\
&- \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \delta^{\tau-t} c P_{\varphi^*,h_t}\left(a_\tau = x_{1,d} \notin \Psi_\tau\right).
\end{aligned}$$

where $\Upsilon = \left\{x_{2,d^2+1}, x_{2,d^2+2}, ...\right\}$. Thus

$$\begin{aligned}
&V_{\varphi}\left(h_t\right) - V_{\widetilde{\varphi}}\left(h_t\right) \\
&= \sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{x \in \Psi_\tau \cap \Phi_\tau \setminus \Upsilon} \delta^{\tau-t} \left(\pi\left(x_{2,d}\right) - \pi\left(x\right)\right) P_{\varphi^*,h_t}\left(\begin{array}{c} a_\tau = x_{2,d} \vee \\ \pi\left(x\right) = \max \pi\left(\Psi_\tau \cap \Phi_\tau \setminus \Upsilon\right) \end{array} \right).
\end{aligned}$$

We seek to show that this last expression is negative. By assumption,

$$P_{\varphi^*,h_t}\left(a_\tau = x_{2,d^2+1} \notin \Psi_\tau\right) = 1;$$

so it is sufficient to show that

$$c > \sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{x \in \Psi_\tau \cap \Phi_\tau \setminus \Upsilon} \delta^{\tau-t} \left(\pi\left(x_{2,d}\right) - \pi\left(x\right)\right) P_{\varphi^*,h_t}\left(\begin{array}{c} a_\tau = x_{2,d} \vee \\ \pi\left(x\right) = \max \pi\left(\Psi_\tau \cap \Phi_\tau \setminus \Upsilon\right) \end{array} \right).$$

Since $\pi(x_{i,d}) = \alpha d$, we may rewrite the inequality as

$$\begin{aligned} & c > \\ & \sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{m=1}^{d-1} \delta^{\tau-t} \alpha (d-m) P_{\varphi^*,h_t} \left(\begin{array}{c} a_\tau = x_{2,d} \vee \\ m = \max d \left(\Psi_\tau \cap \Phi_\tau \setminus \Upsilon \right) \end{array} \right) \\ & = \alpha \sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{m=1}^{d-1} \delta^{\tau-t} P_{\varphi^*,h_t} \left(\begin{array}{c} a_\tau = x_{2,d} \vee \\ m \geq \max d \left(\Psi_\tau \cap \Phi_\tau \setminus \Upsilon \right) \end{array} \right). \end{aligned}$$

Now, note the following conditions are necessary for

$$\begin{array}{l} a_\tau = x_{2,d} \vee \\ m \geq \max d \left(\Psi_\tau \cap \Phi_\tau \setminus \Upsilon \right) \end{array}$$

to hold given conditional on h_t occurring and the player subsequently playing φ^*:

1. Each of $\left\{ x_{1,m+1}, x_{1,m+2}, ..., x_{1,d^1} \right\}$ not to be in $\Psi_\tau \cap \Phi_\tau \setminus \Upsilon$. Note that

$$\left\{ x_{1,1}, ..., x_{1,d^1} \right\} \subset \Psi_\tau \setminus \Upsilon,$$

   so this occurs with probability $\leq \left\{ \begin{array}{c} (1-p)^{d^1-m} : m+1 \leq d^1 \\ 1 : m+1 > d^1 \end{array} \right.$.

2. $x_{2,d}$ be acquired before period τ. This occurs with probability $\leq 1\left\{ d - d^2 \leq \tau - t \right\}$.

Note that the two conditions are independent. Thus we can evaluate

$$\begin{aligned} & \alpha \sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{m=1}^{d-1} \delta^{\tau-t} P_{\varphi^*,h_t} \left(\begin{array}{c} a_\tau = x_{2,d} \vee \\ m \geq \max d \left(\Psi_\tau \cap \Phi_\tau \setminus \Upsilon \right) \end{array} \right) \\ & \geq \alpha \left(\sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \delta^{\tau-t} \cdot 1\left\{ d - d^2 \leq \tau - t \right\} \sum_{m=1}^{d-1} \left\{ \begin{array}{c} (1-p)^{d^1-m} : m+1 \leq d^1 \\ 1 : m+1 > d^1 \end{array} \right. \right) \\ & = \frac{\alpha}{1-\delta} \left(\sum_{d=d^2+1}^{d^1} \delta^{d-d^2} \sum_{m=1}^{d-1} (1-p)^{d^1-m} + \sum_{d=d^1+1}^{\infty} \delta^{d-d^2} \left(\sum_{m=1}^{d^1-1} (1-p)^{d^1-m} + \left(d - d^1 \right) \right) \right) \\ & = O\left(\left(\max\left\{ 1-p, \delta \right\} \right)^{d^1-d^2} \right) \geq O\left(\left(\max\left\{ 1-p, \delta \right\} \right)^{\Delta_{stop}} \right). \end{aligned}$$

Thus we simply have to choose Δ_{stop} sufficiently large so that $V_\varphi(h_t) - V_{\widetilde{\varphi}}(h_t) < c$, and our result is proven.

□

Lemma 5. *Suppose Assumption 1 holds and that $\alpha_i \equiv \alpha > 0$ for all paths i. Given α, there exists $\underline{c} > 0$ such that for all histories h_t where $i = \arg\max_j \left\{ d^j(h_t) \right\}$, $\delta V\left(\Psi(h_t) \cup x_{i,d^i(h_t)+1} - V(\Psi(h_t)) \right) > \underline{c}$.*

Proof. If acquired at h_t, $s_{i,d^i(h_t)+1}$ will be available in the next period with probability p. Note that $\pi\left(s_{i,d^i(h_t)+1} \right) - \pi(s) \geq \alpha$ for all $s \in \Psi(h_t)$. Also, with probability $(1-p)^k$, there will be no opportunities to acquire a new skill in the next period. Noting that both events are independent, there is

probability $\geq p(1-p)^k$ that $s_{i,d^i(h_t)+1}$, if acquired, is exploited in the next period. This implies that $V\left(\Psi(h_t) \cup x_{i,d^i(h_t)+1}\right) - V(\Psi(h_t)) \geq \alpha\delta p(1-p)^k$. The claim follows immediately. □

Proposition 3 Suppose Assumption 1 holds and that $\alpha_i \equiv \alpha > 0$ for all paths i. Given α, there exists $\overline{c}$ such that for $c < \overline{c}$, the player specializes almost surely.

Proof. It is sufficient to show that, for any pair of paths i and j, learning will eventually cease on one of the two paths. WLOG, let $\{i,j\} = \{1,2\}$. Let

$$\upsilon(t) = d_t^1 + d_t^2,$$

be the total number of skills learned on the two paths up till period t, and let

$$\upsilon^{\leftarrow}(l) = \inf\{t : \upsilon(t) = l\}$$

the earliest date at which d skills were learned.on the two paths, in total. Then $d^i_{\upsilon^{\leftarrow}(l)}$ equals the depth of the player's skillset in path j when the total number of skills acquired equals d. Note that $d^1_{\upsilon^{\leftarrow}(l)} + d^2_{\upsilon^{\leftarrow}(l)} \equiv l$. Finally, let

$$\eta_l = \max\left\{d^1_{\upsilon^{\leftarrow}(l)}, d^2_{\upsilon^{\leftarrow}(l)}\right\} - \min\left\{d^1_{\upsilon^{\leftarrow}(l)}, d^2_{\upsilon^{\leftarrow}(l)}\right\}$$

be the absolute difference in depths between the two paths when the total number of skills learned equals d. It will be useful to think of $\{\eta_0, \eta_1, \eta_2, ...\}$ as a sequence of random variables throughout this proof. Note that $\eta_0 = 0$, $\eta_1 = 1$.

Note that Proposition 1 implies that acquisition is more likely to occur on deeper skill paths than shallower skill paths. This implies

$$\begin{aligned}
(\eta_l = 0) &\Rightarrow (\eta_{l+1} = 1), \\
P(\eta_{l+1} = \eta_l + 1) &\geq \frac{1}{2}, \\
P(\eta_{l+1} = \eta_l - 1) &= 1 - P(\eta_{l+1} = \eta_l + 1) \leq \frac{1}{2}.
\end{aligned}$$

Now consider the alternative sequence of rvs $\{\eta'_0, \eta'_1, \eta'_2, ...\}$ defined by

$$\begin{aligned}
\eta'_0 &= 0, \\
P\left(\eta'_{l+1} = \eta'_l + 1\right) &= \frac{1}{2}, \\
P\left(\eta'_{l+1} = \eta'_l - 1\right) &= \frac{1}{2}.
\end{aligned}$$

Note that the distribution of η_l first-order stochastically dominates that of η'_l for every $l > 0$. Note, also, that $\{\eta'_0, \eta'_1, \eta'_2, ...\}$ is an unbiased integer random walk. It is a well known fact that

$$\begin{aligned}
&P(\sup\{\eta'_0, \eta'_1, \eta'_2, ...\} = \infty) = 1; \text{ so} \\
&P(\sup\{\eta_0, \eta_1, \eta_2, ...\} = \infty) = 1.
\end{aligned}$$

This implies that, almost surely, there exists τ such that $\left|d^1_\tau - d^2_\tau\right| > \eta$, for any η. Lemma 4 then implies that one of the two skill paths will eventually cease learning. It remains to show that at least one path will

continue learning indefinitely; but this is an immediate consequence of Lemma 5. □

Proposition 4 is a direct consequence of the following Lemma.

Lemma 6. *Assume that Assumption 1 holds with $\alpha_i \equiv \alpha > 0$ for all paths i, and that $\chi = \chi^{compl}$. Then, as $\min\left\{d^1, d^2\right\} \to \infty$, the value of acquiring an additional skill on the deeper branch goes to zero.*

Proof. Without loss of generality, consider a history h_t^0 where $d^1\left(h_t^0\right) - d^2\left(h_t^0\right) \geq 0$. Let us denote $d^1 = d^1\left(h_t^0\right)$ and $d^2 = d^2\left(h_t^0\right)$.

For the sake of argument, assume for now that $c = 0$. Consider the alternative strategy $\widetilde{\varphi}$ where, for any sequence $\{\Phi_{t+1}, \Phi_{t+2}, ..., \Phi_\tau\}$ of realizations of available skills subsequent to h_t^0, the player imitates φ^* except that it never chooses a skill with depth $\geq d^1 + 1$; whenever the induced action a_τ under φ^* has $d(a_\tau) \geq d^1 + 1$, the player instead chooses to exploit the available skill with the highest payoff. A slight modification of the argument from Lemma 4 shows that

$$V_{\varphi^*}\left(h_t^0\right) - V_{\widetilde{\varphi}}\left(h_t^0\right) < \bar{V}$$

where $\bar{V}$ is some constant that is independent of d^1; that is, the value of acquiring a skill with height $\geq d^1+1$ rather than doing nothing is less than $\bar{V}$.

Returning to the case $c > 0$, the argument above implies that, as $\min\left\{d^1, d^2\right\} \to \infty$, the probability in any given period that it is optimal to acquire a skill with height $\geq d^1 + 1$ goes to zero, because

$$\lim_{\min\{d^1,d^2\}\to\infty} \Pr\left[\exists x_{i,d} : d \leq \min\left\{d^1, d^2\right\} \text{ and } x_{i,d} \text{is available and } \pi(x_{i,d}) > \bar{V}\right] = 1.$$

Now, consider the value of acquiring the skill x_{1,d^1+1} at h_t^0. In order to exploit skills with depth $\geq d^1+1$, the player has to acquire both x_{1,d^1+1} and x_{2,d^1+1}. Since the probability in any given period subsequent to acquiring x_{1,d^1+1} that the player then acquires x_{2,d^1+1} is small for large $\min\left\{d^1, d^2\right\}$, it follows that the gain in continuation value from acquiring x_{1,d^1+1} goes to zero as $\min\left\{d^1, d^2\right\} \to \infty$. □

Proposition 4 Assume that Assumption 1 holds with $\alpha_i \equiv \alpha > 0$ for all paths i, and that $\chi = \chi^{compl}$. Then there exists $m < \infty$ such that almost surely,

$$\lim_{t\to\infty} d_t^1 = \lim_{t\to\infty} d_t^2 = m.$$

Proof. Lemma 6 shows that

$$\lim_{t\to\infty} \max\left\{d_t^1, d_t^2\right\} < \infty,$$

so it remains to show that $\lim_{t\to\infty} d_t^1 = \lim_{t\to\infty} d_t^2$ almost surely. To see this, assume towards a contradiction that the player stops acquisition at some history with $d_t^2 < d_t^1$. But then the player would have been strictly better off if he had not acquired x_{1,d_t^1}. It follows that the optimal strategy must entail the player acquiring subsequently on path 2. □

Lemma 7. *Assume that Assumption 1 holds with $\alpha_i \equiv \alpha > 0$ for all paths i, and that $\chi = \chi^{WC}$. Then* $\limsup_{t\to\infty} \left|d_t^1 - d_t^2\right| = 1$.

Proof. Similar to the proof of Lemma 6. □

Lemma 8. *Assume that Assumption 1 holds with $\alpha_i \equiv \alpha > 0$ for all paths i, and that $\chi = \chi^{WC}$. Assume that $d^i(h_t) \geq d^{-i}(h_t)$. Then there exists $\underline{c}$ such that for all histories h_t,*

$$\delta\left(V\left(\Psi(h_t) \cup x_{-i,\Delta_{-i}(h_t)+1}\right) - V\left(\Psi(h_t)\right)\right) > \underline{c}.$$

Proof. Similar to the proof of Lemma 5. □

Proposition 7 *Assume that Assumption 1 holds with $\alpha_i \equiv \alpha > 0$ for all paths i, and that $\chi = \chi^{WC}$. Then $\limsup_{t\to\infty} \left|d_t^1 - d_t^2\right| = 1$, and there exists $\underline{c}$ such that for $c < \underline{c}$, $\left|\Psi_\infty^*\right| = \infty$ almost surely.*

Proof. Follows from Lemmas 7 and 8. □

Lemma 2 Assume that Assumption 1 holds, that (X, Z) is a path tree with two paths, that $\chi = \chi^{hybrid}$, and that $\alpha > \widetilde{\alpha}$. There exists y such that, at any history h_t, if

$$\widetilde{\alpha} d_t^2 - \alpha d_t^1 \geq y,$$

then $d_\tau^1 = d_t^1$ for all $\tau \geq t$ in the continuation game of the optimal strategy φ^*.

Proof. The argument is similar to that of Lemma 4. Without loss of generality, consider a history h_t^0 where

$$\left(\alpha d^1\left(h_t^0\right) + \beta\right) - \left(\widetilde{\alpha} d^2\left(h_t^0\right) + \widetilde{\beta}\right) \geq \Delta_{stop}.$$

Let us denote $d^1 = d^1\left(h_t^0\right)$ and $d^2 = \Delta_2\left(h_t^0\right)$. Without loss of generality, assume for the sake of contradiction that $\varphi^*\left(h_t^0\right) = x_{2,d^2+1}$; that is, the player acquires on path 2 in period t. Now, consider the alternative strategy $\widetilde{\varphi}$ where, for any sequence $\{\Phi_{t+1}, \Phi_{t+2}, ..., \Phi_\tau\}$ of realizations of available skills subsequent to h_t^0, the player imitates φ^* except that it never chooses $a_\tau \in \left\{x_{2,d_t^2+1}, x_{2,d_t^2+2}, ...\right\}$ or $a_\tau = s$ where for some $x \in \chi(s)$, $x \in \left\{x_{2,d_t^2+1}, x_{2,d_t^2+2}, ...\right\}$; whenever the induced action a_τ under φ^* has $\chi(a_\tau) \cap \left\{x_{2,d_t^2+1}, x_{2,d_t^2+2}, ...\right\}$ nonempty, the player instead chooses to exploit the available skill with the highest payoff. We seek to show that, for sufficiently large Δ_{stop}, $V_{\widetilde{\varphi}}\left(h_t^0\right) > V_{\varphi^*}\left(h_t^0\right)$; continuation payoff at h_t^0 is higher for $\widetilde{\varphi}$ than for φ^*.

We may show, as in proposition 4, that

$$\begin{aligned}
&V_{\varphi^*}(h_t) - V_{\widetilde{\varphi}}(h_t) \qquad (1)\\
&= \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon} \delta^{\tau-t}\left(\pi\left(s_{2,d}^a\right) - \pi(s)\right) P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^a \vee \\ \pi(s) = \max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right)\\
&+ \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon} \delta^{\tau-t}\left(\pi\left(s_{2,d}^c\right) - \pi(s)\right) P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^c \vee \\ \pi(s) = \max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right)\\
&+ \sum_{d=d^2+2}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon} \delta^{\tau-t}\left(\pi\left(s_{1,d}^c\right) - \pi(s)\right) P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{1,d}^c \vee \\ \pi(s) = \max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right)\\
&- \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\delta^{\tau-t} c P_{\varphi^*,h_t^0}\left(a_\tau = x_{2,d}\right).
\end{aligned}$$

where

$$\begin{aligned}
\Upsilon = &\left\{s_{1,d^2+2}^c, s_{1,d^2+3}, ...\right\} \cup \left\{s_{2,d^2+1}^c, s_{2,d^2+2}, ...\right\}\\
&\cup \left\{s_{2,d^2+1}^a, s_{2,d^2+2}^a, ...\right\}.
\end{aligned}$$

Since $P_{\varphi^*,h_t^0}\left(a_t = x_{2,d^2+1}\right) = 1$, it is sufficient to show that

$$
\begin{aligned}
c > & \\
& \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon}\delta^{\tau-t}\left(\pi\left(s_{2,d}^a\right)-\pi\left(s\right)\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^a \vee \\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right) \\
& + \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon}\delta^{\tau-t}\left(\pi\left(s_{2,d}^c\right)-\pi\left(s\right)\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^c \vee \\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right) \\
& + \sum_{d=d^2+2}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon}\delta^{\tau-t}\left(\pi\left(s_{1,d}^c\right)-\pi\left(s\right)\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{1,d}^c \vee \\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right).
\end{aligned}
$$

Noting that $\pi\left(s_{i,d}^a\right) = \alpha d + \beta$ and $\pi\left(s_{i,d}^c\right) = \widetilde{\alpha} d + \widetilde{\beta}$, we now deal with the RHS term-by-term. The first term is

$$
\begin{aligned}
& \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon}\delta^{\tau-t}\left(\pi\left(s_{2,d}^a\right)-\pi\left(s\right)\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^a \vee \\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right) \\
& \leq \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon}\delta^{\tau-t}\alpha\left(d-m\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^a \vee \\ m=\max d\left(\Lambda_\tau^{S^a}\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right) \\
& = \alpha\sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon}\delta^{\tau-t}P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^a \vee \\ m\geq\max d\left(\Lambda_\tau^{S^a}\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right)
\end{aligned}
$$

and we can show that this term is $O\left(\left(\max\left\{1-p,\delta\right\}\right)^{\Delta_{stop}}\right)$ via an identical argument to the proof of Lemma 4.

Next, consider the second term, which is

$$
\begin{aligned}
& \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau\setminus\Upsilon}\delta^{\tau-t}\left(\pi\left(s_{2,d}^c\right)-\pi\left(s\right)\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^c \vee \\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau\setminus\Upsilon\right)\end{array}\right) \\
& \leq \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau\setminus\Upsilon}\delta^{\tau-t}\widetilde{\alpha}\left(d-m\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^c \vee \\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau\setminus\Upsilon\right)\vee \\ m=d\left(s\right)\end{array}\right) \\
& = \sum_{d=d^2+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau\setminus\Upsilon}\delta^{\tau-t}\widetilde{\alpha}P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^c \vee \\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau\setminus\Upsilon\right)\vee \\ m\geq d\left(s\right)\end{array}\right)
\end{aligned}
$$

Now, note the following conditions are necessary for

$$
a_\tau = s_{2,d}^c \vee \pi\left(s\right) = \max\pi\left(\Lambda_\tau^S\cap\Phi_\tau\setminus\Upsilon\right)\vee m\geq d\left(s\right)
$$

to hold conditional on h_t occurring and the player subsequently playing φ^*:

1. Each of $\left\{s_{1,m+1}^c, s_{1,m+2}^c, ..., s_{1,\min\{d^1,d\}}^c\right\}$ not to be $\Lambda_\tau^{S_C}\cap\Phi_\tau\setminus\Upsilon$; so this occurs with probability

$$\leq \begin{cases} (1-p)^{d-m} : d \leq d^1 \\ (1-p)^{d^1-m} : d > d^1 \geq m+1 \\ 1 : m+1 > d^1 \end{cases} \cdot$$

2. Each of $\left\{ s^a_{1,\widetilde{\Delta}(d)}, s^a_{1,\widetilde{\Delta}(d)+1}, ..., s^a_{1,d^1} \right\}$ not to be in $\Lambda^S_\tau \cap \Phi_\tau \setminus \Upsilon$, where $\widetilde{\Delta}(d)$ solves

$$\begin{aligned} \widetilde{\Delta}(d) &= \min_{d' \in \mathbb{N}} \left\{ d' : \alpha d' + \beta \geq \widetilde{\alpha} d + \widetilde{\beta} \right\} \\ &= \left\lceil \frac{\widetilde{\alpha}}{\alpha} d + \frac{\widetilde{\beta} - \beta}{\widetilde{\alpha}} \right\rceil. \end{aligned}$$

This occurs with probability

$$\leq \begin{cases} (1-p)^{d^1 - \widetilde{\Delta}(d)+1} : d^1 \geq \widetilde{\Delta}(d) \\ 1 : \widetilde{\Delta}(d) > d^1 \end{cases}.$$

Note that the definition of $\widetilde{\Delta}(d)$ implies $\widetilde{\Delta}(d) > d$.

3. $x_{2,d}$ be acquired before period τ. This occurs with probability $\leq 1\left\{ d - d^2 \leq \tau - t \right\}$.

Note that all three conditions are independent. Thus we can evaluate

$$\begin{aligned} &\sum_{d=d^2+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{s \in \Lambda^S_\tau \cap \Phi_\tau \setminus \Upsilon} \delta^{\tau - t} \widetilde{\alpha} P_{\varphi^*, h^0_t} \begin{pmatrix} a_\tau = s^c_{2,d} \vee \\ \pi(s) = \max \pi\left(\Lambda^S_\tau \cap \Phi_\tau \setminus \Upsilon \right) \vee \\ m \geq d(s) \end{pmatrix} \\ &\leq \frac{\alpha}{1-\delta} \left(\begin{array}{c} \sum_{d=d^2+1}^{\infty} \delta^{d-d^2} \times \\ \sum_{m=1}^{d-1} \left\{ \begin{array}{c} (1-p)^{d-m} : d \leq d^1 \\ (1-p)^{d^1-m} : d > d^1 \geq m+1 \\ 1 : m+1 > d^1 \end{array} \right. \left\{ \begin{array}{c} (1-p)^{d^1 - \widetilde{\Delta}(d)+1} : d^1 \geq \widetilde{\Delta}(d) \\ 1 : \widetilde{\Delta}(d) > d^1 \end{array} \right. \end{array} \right) \\ &\leq \frac{\alpha}{1-\delta} \left(\begin{array}{c} \sum_{d=d^2+1}^{\infty} \delta^{d-d^2} \times \\ \sum_{m=1}^{d-1} \left\{ \begin{array}{c} (1-p)^{d-m} : \widetilde{\Delta}(d) \leq d^1 \\ 1 : \widetilde{\Delta}(d) > d^1 \end{array} \right. \left\{ \begin{array}{c} (1-p)^{d^1 - \widetilde{\Delta}(d)+1} : d^1 \geq \widetilde{\Delta}(d) \\ 1 : \widetilde{\Delta}(d) > d^1 \end{array} \right. \end{array} \right) \\ &\leq \frac{\alpha}{1-\delta} \left(\sum_{d=d^2+1}^{\infty} \delta^{d-d^2} \times \sum_{m=1}^{d-1} \left\{ \begin{array}{c} (1-p)^{d^1 - \widetilde{\Delta}(d)+d-m+1} : \widetilde{\Delta}(d) \leq d^1 \\ 1 : \widetilde{\Delta}(d) > d^1 \end{array} \right. \right) \\ &= \frac{\alpha}{1-\delta} \left(\sum_{d=d^2+1}^{\widetilde{d}^1} \delta^{d-d^2} \sum_{m=1}^{d-1} (1-p)^{d^1 - \widetilde{\Delta}(d)+d-m+1} + \sum_{d=\widetilde{d}+1}^{\infty} \delta^{d-d^2} (d-1) \right). \end{aligned}$$

where $\widetilde{d}^1$ solves $\widetilde{\Delta}\left(\widetilde{d}^1\right) \leq d^1 \leq \widetilde{\Delta}\left(\widetilde{d}^1 + 1\right)$, or $\left\lceil \frac{\widetilde{\alpha}}{\alpha} \widetilde{d}^1 + \frac{\widetilde{\beta} - \beta}{\widetilde{\alpha}} \right\rceil \leq d^1 \leq \left\lceil \frac{\widetilde{\alpha}}{\alpha} \left(\widetilde{d}^1 + 1 \right) + \frac{\widetilde{\beta} - \beta}{\widetilde{\alpha}} \right\rceil$. Note that

$\left(\alpha d^1+\beta\right)-\left(\widetilde{\alpha}d^2+\widetilde{\beta}\right)>\Delta_{stop}$ implies $d^1-\left(\frac{\widetilde{\alpha}}{\alpha}\left(d_2+\frac{\Delta_{stop}}{\widetilde{\alpha}}\right)+\frac{\widetilde{\beta}-\beta}{\widetilde{\alpha}}\right)\geq 0$; so

$$d^1-\left(\frac{\widetilde{\alpha}}{\alpha}\left(d_2+\frac{\Delta_{stop}}{\widetilde{\alpha}}\right)+\frac{\widetilde{\beta}-\beta}{\widetilde{\alpha}}\right)\geq 0\geq d^1-\left\lceil\frac{\widetilde{\alpha}}{\alpha}\left(\widetilde{d}^1+1\right)+\frac{\widetilde{\beta}-\beta}{\widetilde{\alpha}}\right\rceil,$$

which implies $\widetilde{d}^1\geq d_2+\frac{\Delta_{stop}}{\widetilde{\alpha}}-1$. Also, combining $d^1-\left(\frac{\widetilde{\alpha}}{\alpha}\left(d_2+\frac{\Delta_{stop}}{\widetilde{\alpha}}\right)+\frac{\widetilde{\beta}-\beta}{\widetilde{\alpha}}\right)\geq 0$ and $\widetilde{\Delta}\left(d^2+\frac{\Delta_{stop}}{\widetilde{\alpha}}\right)=\left\lceil\frac{\widetilde{\alpha}}{\alpha}\left(d^2+\frac{\Delta_{stop}}{\widetilde{\alpha}}\right)+\frac{\widetilde{\beta}-\beta}{\widetilde{\alpha}}\right\rceil$ implies that $d^1-\widetilde{\Delta}\left(d\right)+d\geq d^2+\frac{\Delta_{stop}}{\widetilde{\alpha}}-1$ for $d\leq d^2+\frac{\Delta_{stop}}{\widetilde{\alpha}}$. We can now continue with

$$\begin{aligned}
&\frac{\alpha}{1-\delta}\left(\sum_{d=d^2+1}^{\widetilde{d}^1}\delta^{d-d^2}\sum_{m=1}^{d-1}\left(1-p\right)^{d^1-\widetilde{\Delta}(d)+d-m+1}+\sum_{d=\widetilde{d}^1+1}^{\infty}\delta^{d-d^2}\left(d-1\right)\right).\\
&\leq\frac{\alpha}{1-\delta}\left(\sum_{d=d^2+1}^{d^2+\frac{\Delta_{stop}}{\widetilde{\alpha}}-1}\delta^{d-d^2}\sum_{m=1}^{d-1}\left(1-p\right)^{d^2+\Delta_{stop}-m}+\sum_{d=d^2+\Delta_{stop}}^{\infty}\delta^{d-d^2}\left(d-1\right)\right)\\
&\leq\frac{\alpha}{1-\delta}\left(\delta\frac{\left(1-p\right)\left(\left(1-p\right)^{\frac{\Delta_{stop}}{\widetilde{\alpha}}}-\left(1-p\right)\delta^{\Delta_{stop}}\right)}{p\left(\left(1-p\right)-\delta\right)}+\frac{\delta^{\Delta_{stop}}\left(\left(1-\delta\right)\left(d^2+\frac{\Delta_{stop}}{\widetilde{\alpha}}\right)+\left(2\delta-1\right)\right)}{p\left(1-\delta\right)^2}\right)\\
&=O\left(\left(\max\left\{1-p,\delta\right\}\right)^{\frac{\Delta_{stop}}{\widetilde{\alpha}}}\right).
\end{aligned}$$

Almost identically, the third term

$$\begin{aligned}
&\sum_{d=d^2+2}^{\infty}\sum_{\tau=t}^{\infty}\sum_{s\in\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon}\delta^{\tau-t}\left(\pi\left(s_{1,d}^c\right)-\pi\left(s\right)\right)P_{\varphi^*,h_t^0}\left(\begin{array}{c}a_\tau=s_{1,d}^{c\vee}\\ \pi\left(s\right)=\max\pi\left(\Lambda_\tau^S\cap\Phi_\tau^S\setminus\Upsilon\right)\end{array}\right)\\
&=O\left(\left(\max\left\{1-p,\delta\right\}\right)^{\frac{\Delta_{stop}}{\widetilde{\alpha}}}\right).
\end{aligned}$$

Thus we simply have to choose Δ_{stop} sufficiently large so that $V_\varphi\left(h_t\right)-V_{\widetilde{\varphi}}\left(h_t\right)<c$, and our result is proven.

□

Lemma 3 Assume that Assumption 1 holds, that (X,Z) is a path tree with two paths, that $\chi=\chi^{hybrid}$, and that $\alpha>\widetilde{\alpha}$. There exist $y>0$ and $y'>0$ such that, at any period t, if

$$\alpha d_t^1-\widetilde{\alpha}d_t^2\geq y$$

and

$$d_t^2-d_t^1>y',$$

then $d_\tau^2-d_\tau^1\leq d_t^2-d_t^1$ for all $\tau\geq t$ in the continuation game of the optimal strategy φ^* and, almost surely, $\lim_{t\to\infty}d_\tau^1=\lim_{t\to\infty}d_\tau^2<\infty$.

Proof. Without loss of generality, consider a history h_t where

$$d^1(h_t) - d^2(h_t) \geq \Delta_{stop},$$

and

$$\alpha d^1(h_t) + \beta + \Delta_{stop} \leq \widetilde{\alpha} d^2(h_t) + \widetilde{\beta}.$$

Let us denote $d^1 = d^1(h_t)$ and $d^2 = d^2(h_t)$. Without loss of generality, assume for the sake of contradiction that $\varphi^*\left(h_t^0\right) = x_{1,d^1+1}$; that is, the player acquires on path 1 in period t. Now, consider the alternative strategy $\widetilde{\varphi}$ where, for any sequence $\{\Phi_{t+1}, \Phi_{t+2}, ..., \Phi_\tau\}$ of realizations of available skills subsequent to h_t^0, the player imitates φ^* except that it never chooses a skill with depth $\geq d^1 + 1$; whenever the induced action a_τ under φ^* has $d(a_\tau) \geq d^1 + 1$, the player instead chooses to exploit the available skill with the highest payoff. We seek to show that, for sufficiently large Δ_{stop}, $V_{\widetilde{\varphi}}\left(h_t^0\right) > V_{\varphi^*}\left(h_t^0\right)$; continuation payoff at h_t^0 is higher for $\widetilde{\varphi}$ than for φ^*.

We may show (noting that $P_{\varphi^*,h_t^0}\left(a_t = x_{1,d^1+1}\right) = 1$) that

$$\begin{aligned}
&V_{\varphi^*}(h_t) - V_{\widetilde{\varphi}}(h_t) \\
&= \sum_{i=1,2} \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{s \in \Lambda_\tau^S \cap \Phi_\tau^S \setminus \Upsilon} \delta^{\tau-t} \left(\pi\left(s_{i,d}^a\right) - \pi(s)\right) P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{i,d}^a \vee \\ \pi(s) = \max \pi\left(\Lambda_\tau^S \cap \Phi_\tau^S \setminus \Upsilon\right) \end{array}\right) \\
&+ \sum_{i=1,2} \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{s \in \Lambda_\tau^S \cap \Phi_\tau^S \setminus \Upsilon} \delta^{\tau-t} \left(\pi\left(s_{i,d}^c\right) - \pi(s)\right) P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{i,d}^c \vee \\ \pi(s) = \max \pi\left(\Lambda_\tau^S \cap \Phi_\tau^S \setminus \Upsilon\right) \end{array}\right) \\
&- \sum_{i=\{1,2\}} \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \delta^{\tau-t} c P_{\varphi^*,h_t^0}(a_\tau = x_{i,d}) \\
&\leq \sum_{i=\{1,2\}} \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{m=1}^{\infty} \delta^{\tau-t} \alpha P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^a \vee \\ \pi(s) = \max \pi\left(\Lambda_\tau^S \cap \Phi_\tau \setminus \Upsilon\right) \vee \\ m = d(s) \end{array}\right) \\
&+ \sum_{i=\{1,2\}} \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{m=1}^{\infty} \delta^{\tau-t} \widetilde{\alpha} P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^c \vee \\ \pi(s) = \max \pi\left(\Lambda_\tau^S \cap \Phi_\tau \setminus \Upsilon\right) \vee \\ m = d(s) \end{array}\right) - c
\end{aligned}$$

It is thus sufficient to show that

$$\begin{aligned}
c > &\sum_{i=\{1,2\}} \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{m=1}^{\infty} \delta^{\tau-t} \alpha P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^a \vee \\ \pi(s) = \max \pi\left(\Lambda_\tau^S \cap \Phi_\tau \setminus \Upsilon\right) \vee \\ m = d(s) \end{array}\right) \\
&+ \sum_{i=\{1,2\}} \sum_{d=d^1+1}^{\infty} \sum_{\tau=t}^{\infty} \sum_{m=1}^{\infty} \delta^{\tau-t} \widetilde{\alpha} P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s_{2,d}^c \vee \\ \pi(s) = \max \pi\left(\Lambda_\tau^S \cap \Phi_\tau \setminus \Upsilon\right) \vee \\ m = d(s) \end{array}\right).
\end{aligned}$$

We can show, via an identical argument to the proof of Lemma 6, that the second term of the RHS is

$$\sum_{i=\{1,2\}}\sum_{d=d^1+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{m=1}^{\infty}\delta^{\tau-t}\widetilde{\alpha}P_{\varphi^*,h_t^0}\left(\begin{array}{c}a_\tau=s^c_{2,d}\vee\\ \pi(s)=\max\pi\left(\Lambda^S_\tau\cap\Phi_\tau\setminus\Upsilon\right)\vee\\ m=d(s)\end{array}\right)$$
$$=O\left(\left(\max\{1-p,\delta\}\right)^{\Delta_\downarrow}\right).$$

Next, consider the first term, which is

$$\sum_{i=\{1,2\}}\sum_{d=d^1+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{m=1}^{\infty}\delta^{\tau-t}\alpha P_{\varphi^*,h_t^0}\left(\begin{array}{c}a_\tau=s^a_{2,d}\vee\\ \pi(s)=\max\pi\left(\Lambda^S_\tau\cap\Phi_\tau\setminus\Upsilon\right)\vee\\ m=d(s)\end{array}\right)$$

Now, note the following conditions are necessary for

$$a_\tau=s^a_{2,d}\vee\pi(s)=\max\pi\left(\Lambda^S_\tau\cap\Phi_\tau\setminus\Upsilon\right)\vee m=d(s)$$

to hold conditional on h_t occurring and the player subsequently playing φ^*:

1. Each of $\left\{s^a_{1,m+1},s^a_{1,m+2},...,s^a_{1,d^1}\right\}$ not to be in $\Lambda^S_\tau\cap\Phi_\tau\setminus\Upsilon$; so this occurs with probability $\leq$ $\begin{cases}(1-p)^{d^1-m}:d^1\geq m+1\\ 1:m+1>d^1\end{cases}$.
2. Each of $\left\{s^c_{1,\widetilde{\Delta}(d)},s^c_{1,\widetilde{\Delta}(d)+1},...,s^c_{1,d^2}\right\}$ not to be in $\Lambda^S_\tau\cap\Phi_\tau\setminus\Upsilon$, where $\widetilde{\Delta}(d)$ solves

$$\widetilde{\Delta}(d)=\min_{d'\in\mathbb{N}}\left\{d':\widetilde{\alpha}d'+\widetilde{\beta}\geq\alpha d+\beta\right\}=\left\lceil\frac{\alpha}{\widetilde{\alpha}}d+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right\rceil.$$

   This occurs with probability $\leq\begin{cases}(1-p)^{d^2-\widetilde{\Delta}(d)+1}:d^2\geq\widetilde{\Delta}(d)\\ 1:\widetilde{\Delta}(d)>d^2\end{cases}$.
3. $x_{i,d}$ be acquired before period τ. This occurs with probability $\leq 1\left\{d-d^1\leq\tau-t\right\}$.

Note that all three conditions are independent. Thus we can evaluate

$$\begin{aligned}
&\sum_{i=\{1,2\}}\sum_{d=d^1+1}^{\infty}\sum_{\tau=t}^{\infty}\sum_{m=1}^{\infty}\delta^{\tau-t}\alpha P_{\varphi^*,h_t^0}\left(\begin{array}{c} a_\tau = s^a_{2,d}\vee \\ \pi\left(s\right)=\max\pi\left(\Lambda^S_\tau\cap\Phi_\tau\setminus\Upsilon\right)\vee \\ m=d\left(s\right)\end{array}\right)\\
&\leq\frac{2\alpha}{1-\delta}\left(\begin{array}{cc}\sum_{d=d^1+1}^{\infty}\delta^{d-d^1}\times\left\{\begin{array}{c}\left(1-p\right)^{d^2-\widetilde{\Delta}(d)+1}:d^2\geq\widetilde{\Delta}\left(d\right)\\ 1:\widetilde{\Delta}\left(d\right)>d^2\end{array}\right. & \times\\ \sum_{m=1}^{d-1}\left\{\begin{array}{c}\left(1-p\right)^{d^1-m}:d^1\geq m+1\\ 1:m+1>d^1\end{array}\right. & \end{array}\right)\\
&\leq\frac{2\alpha}{1-\delta}\left(\sum_{d=d^1+1}^{\infty}\delta^{d-d^1}\times\left\{\begin{array}{c}\left(1-p\right)^{d^2-\widetilde{\Delta}(d)+1}:d^2\geq\widetilde{\Delta}\left(d\right)\\ 1:\widetilde{\Delta}\left(d\right)>d^2\end{array}\right.\times\sum_{m=1}^{d-1}\left\{\begin{array}{c}\left(1-p\right)^{d^1-m}:d^1\geq m+1\\ 1:m+1>d^1\end{array}\right.\right)\\
&=\frac{2\alpha}{1-\delta}\left(\sum_{d=d^1+1}^{\infty}\delta^{d-d^1}\times\left\{\begin{array}{c}\left(1-p\right)^{d^2-\widetilde{\Delta}(d)+1}:d^2\geq\widetilde{\Delta}\left(d\right)\\ 1:\widetilde{\Delta}\left(d\right)>d^2\end{array}\right.\times\left(\sum_{m=1}^{d^1-1}\left(1-p\right)^{d^1-m}+\left(d-d^1\right)\right)\right)\\
&=\frac{2\alpha}{1-\delta}\left(\sum_{d=d^1+1}^{\widetilde{d}^2}\delta^{d-d^1}\left(1-p\right)^{d^2-\widetilde{\Delta}(d)+1}+\sum_{d=\widetilde{d}+1}^{\infty}\delta^{d-d^1}\right)\left(\sum_{m=1}^{d^1-1}\left(1-p\right)^{d^1-m}+\left(d-d^1\right)\right)
\end{aligned}$$

where $\widetilde{d}^2$ solves $\widetilde{\Delta}\left(\widetilde{d}^2\right)\leq d^2\leq\widetilde{\Delta}\left(\widetilde{d}^2+1\right)$, or $\left\lceil\frac{\alpha}{\widetilde{\alpha}}\widetilde{d}^2+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right\rceil\leq d^2\leq\left\lceil\frac{\alpha}{\widetilde{\alpha}}\left(\widetilde{d}^2+1\right)+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right\rceil$. Note that $\alpha d^1+\beta+\Delta_{stop}\leq\widetilde{\alpha}d^2+\widetilde{\beta}$ implies $d^2-\left(\frac{\alpha}{\widetilde{\alpha}}\left(d^1+\frac{\Delta_{stop}}{\alpha}\right)+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right)\geq 0$; so

$$d^2-\left(\frac{\alpha}{\widetilde{\alpha}}\left(d^1+\frac{\Delta_{stop}}{\alpha}\right)+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right)\geq 0\geq d^2-\left\lceil\frac{\alpha}{\widetilde{\alpha}}\left(\widetilde{d}^2+1\right)+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right\rceil,$$

which implies $\widetilde{d}^2\geq d^1+\frac{\Delta_{stop}}{\alpha}-1-\frac{\widetilde{\alpha}}{\alpha}$. Also, combining $d^2-\left(\frac{\alpha}{\widetilde{\alpha}}d^1+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right)\geq\frac{\Delta_{stop}}{\widetilde{\alpha}}$ and $\widetilde{\Delta}\left(d^1\right)=\left\lceil\frac{\alpha}{\widetilde{\alpha}}d^1+\frac{\beta-\widetilde{\beta}}{\widetilde{\alpha}}\right\rceil$ implies that $d^2-\widetilde{\Delta}\left(d\right)+1\geq\frac{\alpha}{\widetilde{\alpha}}\left(d^1+\frac{\Delta_{stop}}{\alpha}-d\right)$ for $d\geq d^1$. We can thus write

$$\begin{aligned}
&\frac{2\alpha}{1-\delta}\left(\sum_{d=d^1+1}^{\widetilde{d}^2}\delta^{d-d^1}\left(1-p\right)^{d^2-\widetilde{\Delta}(d)+1}+\sum_{d=\widetilde{d}+1}^{\infty}\delta^{d-d^1}\right)\left(\sum_{m=1}^{d^1-1}\left(1-p\right)^{d^1-m}+\left(d-d^1\right)\right)\\
&\leq\frac{2\alpha}{1-\delta}\left(\sum_{d=d^1+1}^{d^1+\frac{\Delta_{stop}}{\alpha}-1-\frac{\widetilde{\alpha}}{\alpha}}\delta^{d-d^1}\left(1-p\right)^{\frac{\alpha}{\widetilde{\alpha}}\left(d^1+\frac{\Delta_{stop}}{\alpha}-d\right)}+\sum_{d=d^1+\Delta_{\downarrow}-\frac{\widetilde{\alpha}}{\alpha}}^{\infty}\delta^{d-d^1}\right)\left(\sum_{m=1}^{d^1-1}\left(1-p\right)^{d^1-m}+\left(d-d^1\right)\right)\\
&=\frac{2\alpha}{1-\delta}\left(\sum_{d=d^1+1}^{d^1+\frac{\Delta_{stop}}{\alpha}-1-\frac{\widetilde{\alpha}}{\alpha}}\delta^{d-d^1}\left(1-p\right)^{\frac{\alpha}{\widetilde{\alpha}}\left(d^1+\frac{\Delta_{stop}}{\alpha}-d\right)}+\sum_{d=d^1+\Delta_{\downarrow}-\frac{\widetilde{\alpha}}{\alpha}}^{\infty}\delta^{d-d^1}\right)\left(\sum_{m=1}^{d^1-1}\left(1-p\right)^{d^1-m}+\left(d-d^1\right)\right)\\
&=O\left(\left(\max\left\{\left(1-p\right)^{\frac{\alpha}{\widetilde{\alpha}}},\delta\right\}\right)^{\frac{\Delta_{stop}}{\alpha}}\right)
\end{aligned}$$

where the last step is arrived at after some tedious calculations. Thus we simply have to choose Δ_{stop}

sufficiently large so that $V_{\varphi^*}(h_t) - V_{\widetilde{\varphi}}(h_t) < c$ to complete the proof.

□

Lemma 9. *Consider the hybrid knowledge structure. Assume that*

$$-\alpha\Delta_{stop} < \alpha d_t^i - \widetilde{\alpha} d_t^{-i} + \left(\beta - \widetilde{\beta}\right) < \widetilde{\alpha}\Delta_{stop}.$$

There exists sufficiently large $\overline{d}$ *such that, for* $d_t^i + d_t^{-i} \geq \overline{d}$*, if*

$$V\left(d_t^i + 1, d_t^{-i}\right) - V\left(d_t^i, d_t^{-i}\right) > V\left(d_t^i, d_t^{-i} + 1\right) - V\left(d_t^i, d_t^{-i}\right),$$

then

$$V\left(d_t^i + k + 1, d_t^{-i}\right) - V\left(d_t^i + k, d_t^{-i}\right) > V\left(d_t^i + k, d_t^{-i} + 1\right) - V\left(d_t^i + k, d_t^{-i}\right)$$

for all $k \geq 1$.

Proof. We start by proving the following claims:

1. As $|\Psi_t| \to \infty$, $P(a_t \in S) \to 1$. Specifically, there exists a sequence $\{g_1, g_2, ...\}$ such that $\lim_{\Delta\to\infty} g_\Delta = 1$ and such that for any Δ, $P(a_t \in S) \geq g_\Delta$ if $|\Psi_t| \geq \Delta$.
2. As $|\Psi| \to \infty$, $V(\Psi) = \frac{1}{1-\delta} E[\max\{v(s) : \chi(s) \subset \Psi\}] + o(1)$. Specifically, there exists a sequence $\{g_1, g_2, ...\}$ such that $\lim_{\Delta\to\infty} g_\Delta = 0$ and such that for any Δ,

$$V(\Psi) - \frac{1}{1-\delta} E[\max\{v(s) : \chi(s) \subset \Psi\}] \leq g_\Delta$$

   if $|\Psi_t| \geq \Delta$.
3. As $|\Psi| \to \infty$, $V\left(d^i + 1, d^{-i}\right) - V\left(d^i, d^{-i}\right)$ is increasing in d^i for $d^i = \left\{0, 1, ...d^{-i} - 1\right\}$ and for $d^i = \left\{d^{-i} + 1, d^{-i} + 2, ...\right\}$.

We start with the first claim. We first show that $V\left(d^i + 1, d^{-i}\right) - V\left(d^i, d^{-i}\right)$ is uniformly bounded above for all $\left(d^i, d^{-i}\right)$. To see this, note that

$$\begin{aligned} &V\left(d^i + 1, d^{-i}\right) - V\left(d^i, d^{-i}\right) \\ &\leq \sum_{\tau=0}^{\infty} \delta^\tau \alpha\tau + 2\sum_{\tau=0}^{\infty} \delta^\tau \widetilde{\alpha}\tau \end{aligned}$$

The last expression is finite and independent of d^i and d^{-i}. Now, $a_t \in S$ whenever

$$V\left(d^i + 1, d^{-i}\right) - V\left(d^i, d^{-i}\right) \leq \max\{v(s) : s \in S_t \vee \chi(s) \subset \Psi\}.$$

But as $|\Psi_t| \to \infty$, $P(\max\{v(s) : s \in S_t \vee \chi(s) \subset \Psi\}) > \mu \to 1$ for any constant μ (with uniform convergence), which implies that $P(a_t \in S) \to 1$.

The second claim follows quickly from the first claim. As $|\Psi| \to \infty$,

$$\begin{aligned} V(\Psi_t) = P(a_t \in S) E[\max\{v(s) : s \in S_t \cap \Phi_\Psi\}] + \delta V(\Psi_t) \quad (2) \\ + P(a_t \in X) E[V(\Psi_t \cup a_t)]; \quad (3) \end{aligned}$$

but since $P(a_t \in S) \to 1$, which implies that $P(a_t \in X) \to 0$, and since $V\left(d^i + 1, d^{-i}\right) - V\left(d^i, d^{-i}\right)$ is uniformly bounded, we have that $V(\Psi_t) = E[\max\{v(s) : s \in S_t \cap \Phi_\Psi\}] + o(1)$.

We proceed to the third claim. Consider the case $d^i \geq d^{-i}$.

Let us order all skills in $\Phi_{d^i,d^{-i}}$ in order of decreasing payoff, so that

$$\Phi_{d^i,d^{-i}} = \left\{ \begin{array}{c} s'_{-l}\left(d^i,d^{-i}\right),...,s'_{-1}\left(d^i,d^{-i}\right), \\ s'_1\left(d^i,d^{-i}\right),...,s'_m\left(d^i,d^{-i}\right) \end{array} \right\}$$

and

$$\begin{aligned} &v\left(s'_{-l}\left(d^i,d^{-i}\right)\right) > ... > v\left(s'_m\left(d^i,d^{-i}\right)\right), \\ &v\left(s'_{-1}\left(d^i,d^{-i}\right)\right) > v\left(s^{sp}_{i,d^i+1}\right) > v\left(s'_1\left(d^i,d^{-i}\right)\right) \end{aligned}$$

(we may assume WLOG that ordering is strict.) Then for $k \geq 1$,

$$P\left(s_k = \arg\max\left\{v(s) : s \in S_t \cap \Phi_{d^i,d^{-i}}\right\}\right) = p(1-p)^{k+l-1};$$

which implies that

$$\begin{aligned} &V\left(d^i+1,d^{-i}\right) - V\left(d^i,d^{-i}\right) \\ &= \frac{1}{1-\delta}\left\{ \begin{array}{c} E\left[\max\left\{v(s) : s \in S_t \cap \Phi_{d^i+1,d^{-i}}\right\}\right] \\ -E\left[\max\left\{v(s) : s \in S_t \cap \Phi_{d^i,d^{-i}}\right\}\right] \end{array} \right\} + o(1) \\ &= \frac{p}{1-\delta}\left(\begin{array}{c} \sum_{k=-l}^{-1} \max\left\{\left(v\left(s^{sp}_{i,d^i+1}\right) - v\left(s'_k\left(d^i,d^{-i}\right)\right)\right),0\right\} p(1-p)^{k+l} \\ +\sum_{k=1}^{m} \max\left\{\left(v\left(s^{sp}_{i,d^i+1}\right) - v\left(s'_k\left(d^i,d^{-i}\right)\right)\right),0\right\} p(1-p)^{k+l-1} \\ +\sum_{k=m+1}^{\infty} v\left(s^{sp}_{i,d^i+1}\right) p(1-p)^{k+l-1} \end{array} \right) \\ &= \frac{p(1-p)^l}{1-\delta}\left(\begin{array}{c} \sum_{k=1}^{m}\left(v\left(s^{sp}_{i,d^i+1}\right) - v\left(s'_k\left(d^i,d^{-i}\right)\right)\right) p(1-p)^{k-1} \\ +\sum_{k=m+1}^{\infty} v\left(s^{sp}_{i,d^i+1}\right) p(1-p)^{k-1} \end{array} \right) + o(1). \end{aligned}$$

Note that if $v\left(s'_k\left(d^i,d^{-i}\right)\right) > v\left(s^{sp}_{i,d^i+1}\right)$, then $s'_k\left(d^i,d^{-i}\right) \in S^c$. Note, next, that

$$
\begin{aligned}
&V\left(d^i+2,d^{-i}\right) - V\left(d^i+1,d^{-i}\right) \\
&= \frac{1}{1-\delta}\left\{ \begin{array}{c} E\left[\max\left\{v(s) : s \in S_t \cap \Phi_{d^i+2,d^{-i}}\right\}\right] \\ -E\left[\max\left\{v(s) : s \in S_t \cap \Phi_{d^i+1,d^{-i}}\right\}\right] \end{array} \right\} + o(1) \\
&= \frac{p}{1-\delta}\left(\begin{array}{c} \sum_{k=-l}^{-1} \max\left\{\left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s'_k\left(d^i,d^{-i}\right)\right)\right),0\right\} p(1-p)^{k+l} \\ +\left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s^{sp}_{i,d^i+1}\right)\right) p(1-p)^{l} \\ +\sum_{k=1}^{m} \max\left\{\left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s'_k\left(d^i,d^{-i}\right)\right)\right),0\right\} p(1-p)^{k+l} \\ +\sum_{k=m+1}^{\infty} v\left(s^{sp}_{i,d^i+2}\right) p(1-p)^{k+l} \end{array} \right) + o(1) \\
&\geq \frac{p(1-p)^l}{1-\delta}\left(\begin{array}{c} \left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s^{sp}_{i,d^i+1}\right)\right) p \\ +\sum_{k=1}^{m}\left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s'_k\left(d^i,d^{-i}\right)\right)\right) p(1-p)^{k} \\ +\sum_{k=m+1}^{\infty} v\left(s^{sp}_{i,d^i+1}\right) p(1-p)^{k} \end{array} \right) + o(1) \\
&= \frac{p(1-p)^l}{1-\delta}\left(\begin{array}{c} \left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s^{sp}_{i,d^i+1}\right)\right) p \\ +\sum_{k=2}^{m}\left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s'_{k-1}\left(d^i,d^{-i}\right)\right)\right) p(1-p)^{k-1} \\ +\left(v\left(s^{sp}_{i,d^i+2}\right) - v\left(s'_{m-1}\left(d^i,d^{-i}\right)\right)\right) \\ +\sum_{k=m+2}^{\infty} v\left(s^{sp}_{i,d^i+1}\right) p(1-p)^{k-1} \end{array} \right) + o(1) \\
&> \frac{p(1-p)^l}{1-\delta}\left(\begin{array}{c} \left(v\left(s^{sp}_{i,d^i+1}\right) - v\left(s'_1\left(d^i,d^{-i}\right)\right)\right) p \\ +\sum_{k=2}^{m}\left(v\left(s^{sp}_{i,d^i+1}\right) - v\left(s'_{k+1}\left(d^i,d^{-i}\right)\right)\right) p(1-p)^{k-1} \\ +v\left(s^{sp}_{i,d^i+1}\right) \\ +\sum_{k=m+1}^{\infty} v\left(s^{sp}_{i,d^i+1}\right) p(1-p)^{k-1} \end{array} \right) + o(1) \\
&= \frac{p(1-p)^l}{1-\delta}\left(\begin{array}{c} \sum_{k=1}^{m}\left(v\left(s^{sp}_{i,d^i+1}\right) - v\left(s'_k\left(d^i,d^{-i}\right)\right)\right) p(1-p)^{k-1} \\ +\sum_{k=m+1}^{\infty} v\left(s^{sp}_{i,d^i+1}\right) p(1-p)^{k-1} \end{array} \right) + o(1)
\end{aligned}
$$

where the final inequality can be verified by a term-by-term comparison. This implies that the third claim holds for the case $d^i \geq d^{-i}$. We may use a similar argument to demonstrate the case $d^i \leq d^{-i}-1$. In the same vein, it is straightforward to show that $V\left(d^i_t, d^{-i}_t+1\right) - V\left(d^i_t, d^{-i}_t\right)$ is decreasing in d^i_t.

We may then infer that if

$$V\left(d^i_t+1, d^{-i}_t\right) - V\left(d^i_t, d^{-i}_t\right) > V\left(d^i_t, d^{-i}_t+1\right) - V\left(d^i_t, d^{-i}_t\right),$$

then

$$V\left(d^i_t+k+1, d^{-i}_t\right) - V\left(d^i_t+k, d^{-i}_t\right) > V\left(d^i_t+k, d^{-i}_t+1\right) - V\left(d^i_t+k, d^{-i}_t\right)$$

for all $k \geq 1$ such that $d^i+k \leq \left\{0,1,...d^{-i}-1\right\}$ or $d^i \geq d^{-i}$. Given large $|\Psi|$, $-\alpha\Delta_{stop} < \alpha d^i_t - \widetilde{\alpha} d^{-i}_t < \widetilde{\alpha}\Delta_{stop}$ implies either $d^i+k \leq \left\{0,1,...d^{-i}-1\right\}$ or $d^i \geq d^{-i}$; this proves the lemma.

□

Proposition 5

Proof. The strategy of proof is as follows. First, note that if either

$$\alpha d_t^i - \widetilde{\alpha} d_t^{-i} + \left(\beta - \widetilde{\beta}\right) \geq \widetilde{\alpha}\Delta_{stop} \text{ or}$$
$$\alpha d_t^i - \widetilde{\alpha} d_t^{-i} + \left(\beta - \widetilde{\beta}\right) \leq -\alpha\Delta_{stop},$$

at some history h, then Lemmas 2 and 3 imply that the player will thereafter either specialize or balance thereafter. We thus need to show that $\alpha d_t^i - \widetilde{\alpha} d_t^{-i}$ will almost surely hit either of these barriers. The way we do so is as follows. Consider h_t such that

$$-\alpha\Delta_{stop} < \alpha d_t^i - \widetilde{\alpha} d_t^{-i} + \left(\beta - \widetilde{\beta}\right) < \widetilde{\alpha}\Delta_{stop}.$$

Now, as in Proposition 3, let

$$\widehat{d}_{j,v} \equiv d_{j,\widetilde{v}^{\leftarrow(d)}}$$

be the depth of the player's skillset in path j when the total number of skills acquired equals d. So our claim that eventually (almost surely) either

$$\alpha d_t^i - \widetilde{\alpha} d_t^{-i} + \left(\beta - \widetilde{\beta}\right) \geq \widetilde{\alpha}\Delta_{stop} \text{ or}$$
$$\alpha d_t^i - \widetilde{\alpha} d_t^{-i} + \left(\beta - \widetilde{\beta}\right) \leq -\alpha\Delta_{stop},$$

is equivalent to the claim that, almost surely,

$$\lim_{v\to\infty} \alpha\widehat{d}_{i,v} - \widetilde{\alpha}\widehat{d}_{-i,v} + \left(\beta - \widetilde{\beta}\right) \geq \widetilde{\alpha}\Delta_{stop} \text{ or}$$
$$\lim_{v\to\infty} \alpha\widehat{d}_{i,v} - \widetilde{\alpha}\widehat{d}_{-i,v} + \left(\beta - \widetilde{\beta}\right) \leq -\alpha\Delta_{stop}.$$

To show this, we will show the following. Assume that at h_t,

$$-\alpha\Delta_{stop} < \alpha\widehat{d}_{i,v} - \widetilde{\alpha}\widehat{d}_{-i,v} + \left(\beta - \widetilde{\beta}\right) < \widetilde{\alpha}\Delta_{stop}.$$

This implies that by either (i) learning $\leq \left\lceil \frac{(\widetilde{\alpha}+\alpha)\Delta_{stop}}{\alpha} \right\rceil$ skills on path i while not learning on path $-i$, or (ii) learning $\leq \left\lceil \frac{(\widetilde{\alpha}+\alpha)\Delta_{stop}}{\widetilde{\alpha}} \right\rceil$ skills on path $-i$ while not learning on path i, one of the barriers will be reached. This observation will come in useful later. Next, assume WLOG that

$$P\left(\widehat{d}_{i,v+1} = d+1 \middle| \widehat{d}_{i,v} = d\right) \geq \frac{1}{2}$$

and that $\widehat{d}_{i,v} \neq \widehat{d}_{-i,v}$; so that given $\widehat{d}_{i,v}$, the next learned situation is weakly more likely to be on path i than path $-i$. Lemma 9 then states that there exists $m > 0$ such that if $|\Psi(h_t)| \geq m$, then the following holds:

$$P\left(\widehat{d}_{i,v+k} = d+k \middle| \widehat{d}_{i,v+k-1} = d+k-1\right) \geq \frac{1}{2}$$

for all $k \geq 1$. That is, if at h_t and under the optimal strategy, it is more likely that the player's next move is to learn on path i than on path $-i$, then as long as the player continues to learn only on path i, he will always be more likely to learn on path i than on path $-i$. If this result holds, then there is probability of at

least $1/2^{\left\lceil \frac{(\widetilde{\alpha}+\alpha)\Delta_{stop}}{\alpha} \right\rceil}$ that one of the two barrier conditions

$$\alpha \widehat{d}_{i,\upsilon} - \widetilde{\alpha} \widehat{d}_{-i,\upsilon} + \left(\beta - \widetilde{\beta}\right) \geq \widetilde{\alpha} \Delta_{stop} \text{ or}$$
$$\alpha \widehat{d}_{i,\upsilon} - \widetilde{\alpha} \widehat{d}_{-i,\upsilon} + \left(\beta - \widetilde{\beta}\right) \leq -\alpha \Delta_{stop}$$

is attained in the next $\left\lceil \frac{(\widetilde{\alpha}+\alpha)\Delta_{stop}}{\alpha} \right\rceil$ increments of υ. This implies that the probability that one of the barrier conditions is attained within the next $k \left\lceil \frac{(\widetilde{\alpha}+\alpha)\Delta_{stop}}{\alpha} \right\rceil$ steps is at least $1 - 1/2^{k\left\lceil \frac{(\widetilde{\alpha}+\alpha)\Delta_{stop}}{\alpha} \right\rceil}$, which converges to one as $k \to \infty$. This is sufficient to prove our result. □

Lemma 6

Proof. The argument is very similar to the proof of proposition 1, so details will be glossed over.

We start with a preliminary result. Consider $\left(d^1, d^2\right)$ and $\left(d^{1\prime}, d^{2\prime}\right)$ such that $d^1 + d^2 = d^{1\prime} + d^{2\prime}$ and $\left|d^1 - d^2\right| > \left|d^{1\prime} - d^{2\prime}\right|$. Then, given levels $\left(d^1, d^2\right)$, the probability that at least one acquired event is available is

$$1 - \left(\frac{1}{2}\left(1-(p+\epsilon)\right)^{d^1}\left(1-(p-\epsilon)\right)^{d^2} + \frac{1}{2}\left(1-(p+\epsilon)\right)^{d^2}\left(1-(p-\epsilon)\right)^{d^1}\right)$$
$$= 1 - \left(\frac{1}{2}\left(\left(1-(p+\epsilon)\right)\left(1-(p-\epsilon)\right)\right)^{\frac{d^1+d^2}{2}} \left(\begin{array}{c} \left(1-(p+\epsilon)\right)^{\frac{\left|d^1-d^2\right|}{2}} + \left(1-(p-\epsilon)\right)^{-\frac{\left|d^1-d^2\right|}{2}} \\ + \left(1-(p+\epsilon)\right)^{-\frac{\left|d^1-d^2\right|}{2}} + \left(1-(p-\epsilon)\right)^{\frac{\left|d^1-d^2\right|}{2}} \end{array}\right)\right)$$

which, fixing $d^1 + d^2$, is decreasing in $\left|d^1 - d^2\right|$.

Define the binary relation $\succcurlyeq$ acting on skillsets as follows: $\Psi \succcurlyeq \Psi'$ if and only if $d^1 + d^2 = d^{1\prime} + d^{2\prime}$ and $\left|d^1 - d^2\right| > \left|d^{1\prime} - d^{2\prime}\right|$. It is sufficient to prove that if $\Psi \succcurlyeq \Psi'$, then $V(\Psi) \geq V(\Psi')$. To show this, for any optimal continuation strategy φ^* starting from Ψ', we will construct a continuation strategy $\widehat{\varphi}$ starting from Ψ which generates higher V. To fix notation, we start our continuation strategies in period t, at history h_t^0.

The strategy $\widehat{\varphi}$ will be constructed with the following property.

For any period $\tau \geq t$, skillset $\Psi \subset X$ and strategy φ, let $\widetilde{\Psi}_\tau(\Psi, \varphi)$ and $\widetilde{a}_t(\Psi, \varphi)$ be, respectively, the random variables whose realizations are respectively the skillset and action in period τ, conditional on the player having skillset Ψ in period and subsequently playing continuation strategy φ. Then $\widetilde{\Psi}_\tau(\Psi \cup x', \widetilde{\varphi})$ dominates $\widetilde{\Psi}_\tau(\Psi \cup x'', \varphi^*)$ with respect to $\succcurlyeq$ for some remappings $\upsilon_{\widetilde{\varphi},\tau} : [0,1] \to 2^X$ and $\upsilon_{\varphi^*,\tau} : [0,1] \to 2^X$, and $u\left(\widetilde{a}_\tau(\Psi \cup x', \widetilde{\varphi})\right)$ dominates $u\left(\widetilde{a}_\tau(\Psi \cup x'', \varphi^*)\right)$ with respect to the usual $>$ for some remappings $\phi_{\varphi^*,\tau} : [0,1] \to X$ and $\phi_{\widetilde{\varphi},\tau} : [0,1] \to X$.

Assume for the sake of induction that $\widetilde{\Psi}_\tau(\Psi \cup x', \widetilde{\varphi})$ dominates $\widetilde{\Psi}_\tau(\Psi \cup x'', \varphi^*)$ with respect to $\succcurlyeq$ for some remappings $\upsilon_{\widetilde{\varphi},\tau} : [0,1] \to 2^X$ and $\upsilon_{\varphi^*,\tau} : [0,1] \to 2^X$ for period τ. Then, as in proposition 1, we may show that $\widetilde{\Psi}_{\tau+1}(\Psi \cup x', \widetilde{\varphi})$ dominates $\widetilde{\Psi}_{\tau+1}(\Psi \cup x'', \varphi^*)$ with respect to $\succcurlyeq$ for some remappings $\upsilon_{\widetilde{\varphi},\tau+1} : [0,1] \to 2^X$ and $\upsilon_{\varphi^*,\tau+1} : [0,1] \to 2^X$, and $u\left(\widetilde{a}_\tau(\Psi \cup x', \widetilde{\varphi})\right)$ dominates $u\left(\widetilde{a}_\tau(\Psi \cup x'', \varphi^*)\right)$ with respect to the usual $>$ for some remappings $\phi_{\varphi^*,\tau} : [0,1] \to X$ and $\phi_{\widetilde{\varphi},\tau} : [0,1] \to X$.

Such a construction is possible because (i) conditional on $\left(d^1, d^2\right)$, the event that at least one skill is available for acquisition and the event that at least one skill is available for exploitation are uncorrelated, (ii) the probability that at least one skill is available for exploitation decreases with $\left|d^1 - d^2\right|$, and (iii) the probability that at least one skill is available for acquisition is independent of $\left(d^1, d^2\right)$.

This completes the induction step and also completes the proof. □