THE PINCHAS SAPIR CENTER
FOR DEVELOPMENT
Tel Aviv University

המרכז לפיתוח ע"ש פנחס ספיר
ליד אוניברסיטת תל אביב
"עמותה רשומה"
580010221

THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

On the Optimal Scheduling of Attention

Kfir Eliaz, Daniel Fershtman, Alexander Frug

Discussion Paper No. 7-2022

אוניברסיטת תל-אביב, תל אביב יפו 6997801 | טל': 03-6409202 | פקס: 03-6405815
Tel Aviv University, Tel Aviv Yafo 6997801 Israel | Tel. 972-3-6409202 | Fax 972-3-6405815 | email: sapircen@tauex.tau.ac.il

On the Optimal Scheduling of Attention*

Kfir Eliaz† Daniel Fershtman‡ Alexander Frug§

July 18, 2022

Abstract

We consider a decision-maker sequentially choosing which task to attend to when payoffs depend on both the chosen and unchosen tasks. We show that when tasks are substitutes (complements) such that the flow payoffs are a sum (product) of the tasks' outputs, the optimal policy is an index policy, generalizing the independence of irrelevant alternatives (IIA) property known in the classic multi-armed bandit problem. We illustrate the usefulness of our model through several applications, including repeated bargaining, dynamic occupational choice, on-the-job training, R&D, and dynamic supervision.

Keywords: Allocation of attention, Scheduling, Time management, Multitasking, Task juggling, Multi-armed bandits

*Financial support from the Foerder Institute and the Sapir Center at Tel-Aviv University is greatfully acknoweldged.

†Eitan Berglas School of Economics, Tel-Aviv University and David Eccles School of Business, University of Utah. kfire@tauex.tau.ac.il.

‡Eitan Berglas School of Economics, Tel-Aviv University. danielfer@tauex.tau.ac.il.

§Department of Economics and Business, Universitat Pompeu Fabra and Barcelona Graduate School of Economics. alexander.frug@upf.edu.

1 Introduction

As individuals, we are constantly engaged in scheduling tasks and in deciding when to switch from one activity to another. Oftentimes, we are affected by the outcomes of tasks even in periods in which we do not attend to them. For instance, the success of a manager who allocates his time between different teams/departments under his control depends on the overall performance of all of them. While the manager is attending to one of his teams, his overall success continues to depend on the performance of all the other teams that are not currently receiving his attention. Similarly, due to time constraints, consumers with monthly payments to multiple services (internet, cell phone coverage, gyms, streaming, etc.) can typically exert effort in lowering the cost of only one service at a time, and hence need to choose which service to focus on at any given point in time. While the consumer is engaged in lowering the cost of one service, his overall expenses are still affected by those of the remaining ones.

Put differently, many attention allocation problems have the feature that in each period, the decision-maker receives a payoff from *each* available activity – not just the one he is attending to. The goal of this paper is to develop a methodology to analyze such decision problems and to illustrate its applicability.

We consider a decision-maker (DM) who faces a finite number of tasks/alternatives, and in each period must decide which task to choose. The per-period payoff from a given task depends on its state and on whether or not the DM attends to it in the current period. The state of each given task may change only during those periods in which the DM attends to the task, in which case its new state is drawn from a distribution that depends on the task's previous state. The DM's overall flow payoff is a function of all of the task-specific per-period payoffs. This captures environments where several activities generate a payoff even when left unattended, as well as environments in which the flow payoff is generated only from the chosen task, but where this payoff depends on the states of the unchosen tasks.

We focus on two polar cases: the case of *substitute* tasks, in which the DM's flow payoff equals the *sum* of outputs of all of the tasks (both chosen and unchosen), and the case of *complementary* tasks, in which the DM's flow payoff equals the *product* of all of the tasks' outputs. In both cases, the classic Gittins index solution is not applicable since the DM's flow payoff depends on the states of *all* tasks.[1] Nevertheless, we establish that the

[1]That is, such problems do not fall into the classic paradigm of the multi-armed bandit problem, in which the DM receives a payoff *only* from the arm he pulls *and only* when he pulls it.

DM's optimal strategy *is* characterized by an index policy: there exists a function that assigns to each task a score, *which depends only on the characteristics of that particular task* (and, in particular, is independent of any information about the other tasks), such that in each period, given the states of all tasks, it is optimal to pick the task with the highest current score.

Our analysis begins with the case of substitute tasks. We characterize the optimal policy, which is based on indices that account for the externality a task imposes on others. Using this characterization, we also show that—owing to the additive separability of the substitute case across tasks and across periods—the DM's problem can be reformulated in a way that admits the standard Gittins index solution. The idea is to think of the payoff from choosing a task as also including the expected discounted sum of payoffs in future periods, assuming that the task will not be selected again. The attention policy induced by the Gittins index of this reformulated decision problem coincides with the policy induced by our index.

We illustrate the applicability of our index policy for the case of substitute tasks through a number of examples.

On-the-job training. An instructor/supervisor faces a set of identical workers and must decide in each period which worker to guide. The workers are initially unproductive, but their productivity stochastically improves when given guidance. Applying our index policy, we characterize the optimal guidance strategy.

Repeated bargaining. Each period a firm has to assign a project to one of two contractors. The surplus generated by a contractor is a function of his productivity, which increases with the number of projects assigned to him. The division of this surplus between the firm and the chosen contractor is determined according to the Stole and Zwiebel (1996b) reduced-form bargaining solution, which takes into account the productivity of both contractors (i.e., the chosen and the unchosen). The firm then faces the following trade-off: remaining with the same contractor increases his productivity, but can also raise his share in the surplus. We show that the firm's decision problem can be formulated in a way that fits into our framework. Applying our solution yields that depending on the discount factor and the firm's (exogenous) bargaining power, the firm either switches between contractors each period, or it remains with the same contractor in all periods. We derive a condition characterizing which case is optimal, as a function of the parameters.

Career paths and mobility between sectors. We extend the classic literature on dynamic occupational choice (most notably, Jovanovic, 1979; Miller, 1984) by allowing the transfer of human capital investment across occupations. Motivated by the observation that the career paths of many academics and professionals (lawyers, economists, engineers) often involve switching between positions in the private and the public sector, we consider the problem of a DM who in each period must choose between each of the sectors. We assume that the productivity of a worker in a given period is the sum of the human capital he accumulated in the current sector plus a proportion of the human capital he accumulated in the other sector. Each sector is then characterized by a parameter that captures the fraction of accumulated human capital that is transferable to the other sector. Thus, moving to a sector with a lower current payoff may be interpreted as an investment in human capital that increases future payoffs. Focusing on a simple parametric model of human capital accumulation in each sector, we show the conditions under which the DM chooses to begin his career in the public sector, and those under which he instead chooses to start in the private sector, switching for a brief stint in the public sector after having accumulated enough human capital.

We then turn our attention to the case of complementary tasks. This case captures environments where a team of agents works on a joint project and, in order to complete the project, all agents must succeed in their task. This case also fits environments in which several factors of production must be developed, and the total production has the Cobb–Douglas form.[2]

The index policy for complementary tasks is significantly different from the one for substitute tasks. In particular, two indices, each of which induces a distinct ranking, must be considered. In order to determine which of the indices must be used to evaluate the choice of a given task, we first need to check whether there exists a (possibly stochastic) stopping time such that the expected discounted output at the stopping time is higher than the current output. We refer to tasks for which the answer is affirmative as "augmenting." An augmenting task is always chosen over a nonaugmenting one and, among tasks of the same status (augmenting or nonaugmenting), the selection is made according to their associated index.

Unlike the case of substitutes, it remains an open question whether there exists a way to reformulate the decision problem such that applying the Gittins index policy to the

[2] At first glance, one might think that a monotonic transformation (e.g., $log(\cdot)$) on the periodic payoff function might yield an equivalent decision problem in which the periodic payoffs are additive in the states. However, this intuition is incorrect, as the objective being maximized is the discounted sum of payoffs, and hence a monotonic transformation on the objective function is different than a monotonic transformation on the periodic payoff function. Indeed we show that the dynamics under the optimal policy are very different under the two cases of substitutes (additive periodic payoff function) and complements (multiplicative periodic payoff function).

reformulated problem induces the optimal attention policy.

To illustrate the case of complementary tasks, and how it differs from the case of substitutes, we analyze the following examples.

On-the-job training. A pair of agents work on a joint project that comprises two independent tasks. The project is completed if and only if each agent completes his task. Each agent is able to complete his task with certainty after two periods of guidance by the DM. While one worker improves his productivity after a single period of guidance, the second worker makes no intermediate improvement. We characterize the optimal training strategy and show that unlike in the case of substitute tasks, it depends on the discount factor.

Developing multiple complementary attributes. A firm is developing two complementary attributes of a single product (say, speed and accuracy), and needs to decide which attribute a team of workers should focus on each period, taking into account that each period there is a fixed probability that the final product will need to be introduced on the market (because of competitive pressure). While the level of one attribute increases every time it is worked on, the level of the other attribute requires additional periods of attention before it increases in value. Assuming the firm's profit equals the product of the two attributes' levels, we examine how the sequencing of attention is affected by the probability that the product needs to be offered on the market (which captures the level of competition).

Supervising agents with stochastic costs. There are two agents who jointly work on a project. Each agent is in charge of a task, and the project is completed successfully if and only if both agents successfully complete their respective task. A principal needs to decide each period which agent to supervise. When supervised, an agent completes his task in the project, but when left unsupervised, the agent works if and only if he perceives the cost of effort to be lower than some threshold. Cost perceptions are stochastic, and each agent starts with an initial cost threshold that can change with the number of times he was supervised. We derive the optimal supervision policy when one agent always shirks without supervision, while the cost threshold of the second agent increases with supervision.

The two cases we study are nested within a broader class of problems in which the DM maximizes a discounted expected *generalized mean* of payoffs from all tasks. We take a preliminary step towards characterizing the solution to this entire class by focusing on the two extremes: the *arithmetic* and *geometric* means. As demonstrated by our examples, these two cases encompass a wide variety of economically relevant environments to which our index policies can be applied.

The remainder of the paper is organized as follows. The next section reviews the related literature. Section 3 presents the model of substitute tasks. The optimal strategy, which takes the form of an index policy, is characterized in Subsection 3.1 and illustrated in Subsection 3.2. Section 4 presents the model of complementarities between tasks, derives the optimal policy in Subsection 4.1 and illustrates its working in Subsection 4.2. Concluding remarks are given in Section 5. All omitted proofs are in the Appendix.

2 Related literature

The problem of allocating time/attention between tasks has been previously studied in the literature under different frameworks. In a series of papers, Coviello et al. (2014, 2015) study the problem of a DM who faces a growing queue of tasks that arrive at an exogenous rate. In their 2014 paper, the authors characterize the production function, which relates the output rate to the effort rate (which governs completion time) and the activation rate (at which tasks are started). Their 2015 paper applies this production function to a dataset of judges' handling of court cases to estimate the effect of increased case load. Bray et al. (2016) model how a judge schedules cases as a classic multi-armed bandit problem, and argue that prioritizing the oldest hearing (case) is optimal when the case completion hazard rate function is decreasing (increasing). Using data on Italian judges, they estimate that a switch from prioritizing the oldest hearing to prioritizing the oldest case greatly decreased average case duration.

The present paper complements the work of these authors by considering a different framework where the set of tasks is given, and the DM decides which task to attend to in each period, taking into account how his payoffs depend on both chosen and unchosen tasks. Our main contribution is a characterization of the DM's optimal strategy when tasks are substitutes and complements.

In their classic paper, Radner and Rothschild (1975) study the problem of a DM who needs to allocate a unit of attention in each period among a given set of tasks. The output of an attended (unattended) task increases (decreases) as a function of the amount of effort allocated to it. Instead of deriving the optimal strategy, the authors compare the survival probability of several heuristics (the probability that the outputs of all tasks remain above some threshold).

Our model is naturally related to the multi-armed bandit framework, which has been widely applied in a variety of fields in economics.[3] Due to the dependence of payoffs on the states of all tasks, however, our model does not fall under this classic framework. This feature also distinguishes our analysis relative to the classic literature on "learning by experimentation" (e.g., Keller et al., 2005). The DM's optimal policy in our problem takes the form of an index policy, and in this sense generalizes the key IIA property that has contributed to the applicability of the classic multi-armed bandit framework.

While our model is motivated by environments where learning about alternatives can be linked to the flow payoffs they generate, it is related to a small literature studying the problem of a DM that acquires information about multiple attributes of an object before deciding between the object and an outside option. Since the value of the object depends on all its attributes—whether inspected or not—this decision problem resembles the one we study. However, it does not fit into our framework because the final payoff of the DM is the maximum of the object's expected value and the value of the outside option. In particular, it is not known whether the optimal strategy in this environment admits an index policy.

Notable examples in this literature include Klabjan et al. (2014), who study a DM inspecting a good's attributes before choosing between the object and an outside option. The DM's payoff from the object is a weighted average of the attributes' values, of which he is initially uninformed. He can inspect attributes at a cost, thereby learning their value, before making a decision. Eliaz and Frug (2018) study a related problem, where a seller decides which attributes of an asset to inspect before proposing a price to a buyer, who does not observe the outcome of the seller's inspections.

Several recent papers have studied the problem of a DM that gradually acquires costly information about a set of options before stopping and choosing one of them. These include Ke et al. (2016), Fudenberg et al. (2018), Che and Mierendorff (2019), Ke and Villas-Boas (2019), and Liang et al. (2021). A key difference between these works and ours is that the DM's payoff in our model is a function of the states of all alternatives, whether chosen or not. Additionally, in contrast to our framework, the optimal strategy in these works does not take the form of an index policy.[4]

Similar to the classic multi-armed bandit framework and the papers discussed above, in the present paper the set of alternatives among which the DM allocates attention is fixed ex ante. Fershtman and Pavan (2020) analyze a model where, in addition to

[3]See Bergemann and Valimaki (2008) for an excellent survey.

[4]Gossner et al. (2020) also a study a problem in a which a DM sequentially learns about options before choosing one of them. They assume an exogenous stopping rule, which implies that the optimal learning strategy follows an index policy.

exploring alternatives already in the DM's consideration set, the DM can choose to search for additional alternatives in response to information gathered about existing ones.

3 Attention scheduling with substitutes

There are n alternatives that require the attention of a DM. Each period, $t = 0, 1, ...\infty$, the DM can allocate attention to at most a single alternative. The DM can also choose not to allocate attention to any alternative. In each period t, if the DM focuses on alternative i, his payoff is

$$U_t(x_{1,t}, ..., x_{n,t}) = u_i(x_{i,t}) + \sum_{j \neq i} v_j(x_{j,t}),$$

where u_i and v_i are bounded functions, and $x_{i,t} \in X_i$ represents the period-t state of alternative i. Each X_i is an arbitrary state space. The function u_i represents the payoff from alternative i at the time the DM allocates attention to it, while v_i is the *passive* payoff in periods in which the DM allocates attention to another alternative. As we will see in the applications of Section 3.2, v_i can also capture an (additive) externality that unchosen alternatives impose on the chosen alternative in situations where an alternative generates a payoff only when chosen.

The state of an alternative changes only in a period in which the DM allocates attention to it. Specifically, given $x_{i,t}$, if the DM allocates attention to alternative i in period t, the alternative's next state $x_{i,t+1}$ is drawn from the distribution $F_i(\cdot|x_{i,t})$ defined on X_i, and the states of the other alternatives remain unchanged, $x_{j,t+1} = x_{j,t}$. For example, the change in an alternative's state may capture investment, learning about an unknown characteristic, learning-by-doing, or habit formation.

An *attention policy* Γ specifies, given the current state of all alternatives $(x_1, ..., x_n)$, which alternative (if any) receives attention.[5] The DM wishes to maximize the expected discounted stream of payoffs. An *optimal* attention policy is therefore a policy that maximizes

$$\mathbb{E}\left(\sum_{s=0}^{\infty} \delta^s U(x_{1,s}, ..., x_{n,s}) | x_{1,0}, ..., x_{n,0}\right).$$

[5] The possibility of not allocating attention to any alternative can be captured by introducing a fictitious alternative whose state remains constant and for which the functions u and v are constant at zero.

To ease exposition, we assume that there is no direct cost for allocating attention to an alternative other than the opportunity cost of allocating attention to a different alternative and discounting.

3.1 The optimal policy

We now characterize the DM's optimal attention policy. Consider an alternative i that is in state x_i in some unspecified period. With a slight abuse of notation, we denote by $x^{+\tau}$ the (possibly random) state of that alternative following τ periods of receiving attention. We use this notation so that we do not need to specify the particular time period in which the alternative was in the initial state x_i. Using this notation, for any state x_i of alternative i, given a (realization-dependent) stopping rule τ, define

$$a_i(x_i, \tau) \equiv \mathbb{E}\left(\delta^\tau v_i(x_i^{+\tau})|x_i\right) - v_i(x_i). \tag{1}$$

The first component, $\mathbb{E}\left(\delta^\tau v_i(x_i^{+\tau})|x_i\right)$, represents the expected discounted *passive* payoff of alternative i after τ periods of receiving attention, starting from state x_i. Thus, for a given stopping rule τ, the function $a_i(x_i, \tau)$ captures the expected discounted increase in the passive payoff of alternative i, starting at x_i and stopping according to the stopping rule τ.

Given the state $x_{j,t} \in X_j$ of alternative j in period t, define an index

$$I_j(x_{j,t}) \equiv \sup_\tau \left\{ \frac{(1-\delta)\mathbb{E}\left(\sum_{s=0}^{\tau-1} \delta^s u_j(x_{j,t+s})\right) + a_j(x_{j,t}, \tau)}{\mathbb{E}\left(1-\delta^\tau\right)} \right\}, \tag{2}$$

where the sup is taken over all (possibly stochastic) stopping rules. The index I_j is a function only of the state of alternative j, and is independent of any information about the other alternatives. Note that when $v \equiv 0$ for all the alternatives, the payoff in each period is a function of only the state of the alternative that receives attention, such that (2) reduces to the classic Gittins index of Gittins and Jones (1974).

Standard dynamic programming results imply that a stopping rule $\tau_j^*(x_j)$ attaining the supremum in (2) satisfies the following useful property: beginning at state x_j, $\tau_j^*(x_j)$ is the first time at which the index becomes smaller than $I_j(x_j)$. That is, $\tau_j^*(x_j)$ is equal to the first period $t > 0$ such that $I_j(x_j^{+t}(x_j)) \leq I_j(x_j)$, where $x_j^{+t}(x_j)$ denotes alternative j's (stochastic) state after t periods of attention, starting at state x_j (see, e.g., Mandelbaum, 1986).[6] This property will be useful for deriving the index in practice (and will be used in the applications below).

Claim 1. τ^* *is a stopping rule that attains the supremum in (2) if and only if it satisfies*

$$min\{t > 0 : I_j(x_j) \geq I_j(x_j^{+t}(x_j))\} \leq \tau_j^*(x_j) \leq min\{t > 0 : I_j(x_j) > I_j(x_j^{+t}(x_j))\}.$$

In other words, it is without loss to assume that $\tau_j^*(x_j)$ stops once a state is reached for which the index is equal to $I_j(x_j)$. Furthermore, once a state is reached for which the index is strictly below $I_j(x_j)$, $\tau_j^*(x_j)$ stops immediately.

To illustrate the intuition for this property, suppose that the state of alternative j evolves deterministically. Let $\hat{\tau}_j$ be a stopping rule that, given each state x_j, stops immediately when the index drops weakly below $I_j(x_j)$. That is, for all[7] x_j, $\hat{\tau}_j(x_j) = \min\{t > 0 : I_j(x_j) \geq I_j(x_j^{+t}(x_j))\}$. Denote $D_j(x_j, \tau) = \sum_{s=0}^{\tau-1} \delta^s$ and

$$N_j(x_j, \tau) = \sum_{s=0}^{\tau-1} \delta^s u_j(x_j^{+s}) + \delta^\tau v_j(x_j^{+\tau}) - v_j(x_j).$$

Suppose by way of contradiction that $I_j(x_j^{+\hat{\tau}}) < I_j(x_j)$, but that $\hat{\tau} < \tau^*$. That is, stopping under τ^* does not occur immediately after reaching a state for which I_j is strictly smaller than $I_j(x_j)$. Then

$$I_j(x_j) = \frac{N_j(x_j, \tau^*)}{D_j(x_j, \tau^*)} = \frac{N_j(x_j, \hat{\tau}) + \delta^{\hat{\tau}} N(x_j^{+\hat{\tau}}, \tau^* - \hat{\tau})}{D_j(x_j, \hat{\tau}) + \delta^{\hat{\tau}} D(x_j^{+\hat{\tau}}, \tau^* - \hat{\tau})} \tag{3}$$

and

$$\frac{N(x_j^{+\hat{\tau}}, \tau^* - \hat{\tau})}{D(x_j^{+\hat{\tau}}, \tau^* - \hat{\tau})} \leq I_j(x_j^{+\hat{\tau}}) < I_j(x_j), \tag{4}$$

where the first inequality follows from the definition of the index I_j. But (3) and (4) together imply that $\frac{N(x_j, \hat{\tau})}{D(x_j, \hat{\tau})} > I_j(x_j)$, which yields the desired contradiction. An analogous argument shows that τ^* cannot stop before the index drops weakly below I_j.

Given the index defined in (2), we define the following attention policy.

[6] Note that allowing for infinity as a possible value of the stopping time, the supremum in the definition of (2) is attained; that is, an optimal stopping rule exists (see Puterman, 2014).

[7] To ease the exposition, the notation will omit below the dependence of x_j^{+s} and of the stopping rules τ^* and $\hat{\tau}$ on the state x_j.

Definition 1. *Denote by* Γ^* *an attention policy that allocates attention in each period to the alternative with the highest index.*[8]

Under this attention policy, the decision in each period boils down to a simple comparison of independent indices.

Theorem 1. Γ^* *is an optimal attention policy in the model with substitute alternatives.*

The proof builds on an interchange argument due to Gittins and Jones (1974). It suffices to show that any policy π^0 that differs from Γ^* in period 0 and subsequently coincides with it attains a discounted expected payoff no greater than that of Γ^* To show this, starting with any such arbitrary policy π^0, we construct a sequence of modifications of π^0, $(\pi^1, \pi^2, ...)$, such that each modified policy π^k coincides with Γ^* for at least the first k periods and attains a weakly higher expected payoff than its predecessor, and furthermore, the expected discounted payoff under π^k converges to the expected discounted payoff under Γ^* as $k \to \infty$.

In the appendix we give a unified proof of the optimal attention policy for both the case of substitutes (Theorem 1) and complements (Theorem 2) since both theorems use the same methodology outlined in the previous paragraph. In some steps of the proof, the details differ depending on whether the alternatives are substitutes or complements. In these steps, we describe which argument to use for each of the two cases.

Even though the DM's flow payoff depends on the states of all alternatives—the one he is attending to and all the others—the above result establishes that in each period the DM optimally chooses the alternative with the highest index. This therefore generalizes the IIA property known in the classic bandit framework (that is, the DM prefers to attend to alternative a rather than alternative b in a particular period independently of the other alternatives). The indices capture both the expected stream of payoffs u associated with allocating attention to an alternative, but also the payoffs v that the other alternatives generate when they do not receive attention.

The fact that the optimal policy takes such a simple separable form is important for applications. It is useful both for deriving properties of the dynamics and comparative statics under the optimal policy and for computational purposes. Without such separability, the optimal policy, in principle, could still be computed using dynamic programming. However, this would require strong assumptions on the state variables in order to simplify

[8] In the case of ties between indices, any tie-breaking rule may be specified.

computation due to the "curse of dimensionality." Rearranging (2) yields the following alternative representation of our index:

$$I(x_{j,t}) = \sup_\tau \left\{ \underbrace{\left(\frac{\mathbb{E}\left(\sum_{s=0}^{\tau-1} \delta^s u_j(x_{j,t+s})\right)}{\mathbb{E}(\sum_{s=0}^{\tau-1} \delta^s)} \right)}_{\text{(a)}} - \underbrace{\left(\frac{v_j(x_{j,t}) - \mathbb{E}\left(\delta^\tau v_j(x_{j,t+\tau})\right)}{(1-\delta)\mathbb{E}(\sum_{s=0}^{\tau-1} \delta^s)} \right)}_{\text{(b)}} \right\}.$$

The indices therefore maximize the difference between two components: (a) the expected discounted payoff per unit of expected discounted time, and (b) the discounted continuation value of the expected change in the passive payoff per unit of expected discounted time. The first component is the one maximized by the well-known Gittins index. The second component is new, and reflects the fact that when an alternative does not receive attention, it continues to contribute to the overall payoff, as a function of its state.

The following example illustrates the workings of the optimal policy characterized in Theorem 1.

Example. The output produced by a worker often depends on whether he works alone or receives guidance from the principal. This guidance is likely to improve the worker's skills, thereby increasing the worker's future productivity. Consider the following problem in which a principal dynamically allocates his attention among workers, training them one at a time, with the goal of maximizing the total discounted expected output they generate.

Let a worker's state $x \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$ denote the worker's skill level. Assume that when a worker is in state x, he produces x if he receives attention and only x^2 if he works alone. Training a worker in state x may affect the worker's future productivity. Specifically, we assume that if a worker is trained while in state x, he moves to a new state $x' \sim U\{x, x+\frac{1}{3}, .., 1\}$. For simplicity, we assume that training a worker entails no direct costs for the principal.

Claim 2. *For any δ the principal will first train workers who are in state $\frac{1}{3}$. Then, if $\delta > \delta^*$ (≈ 0.62), he will train workers in state 0, followed by workers in state $\frac{2}{3}$. Otherwise, he will train workers in state $\frac{2}{3}$ before workers in state 0.*

Proof: Denote by $I(x, \tau)$ the expected discounted per-period value from the stopping rule τ at state x. The index in state x therefore satisfies $I(x) = \sup_\tau\{I(x, \tau)\}$.

In order to characterize the principal's optimal training policy, first note that there is no value in training a worker who is at the highest skill level, $x = 1$, since this would not affect either the worker's current or future productivity. Formally, it is easy to see that for any stopping rule τ, $I(1) = I(1, \tau) = 0$.

Next, by Claim 1, $I(\frac{2}{3}) = I(\frac{2}{3}, 1)$; i.e., the index is supported by the deterministic stopping rule $\tau = 1$ (that is, τ specifies stopping after a single period, regardless of the realization of the new state). It can easily be verified that $I(\frac{2}{3}) = \frac{4+\delta}{18(1-\delta)}$.

To derive $I(\frac{1}{3})$, we first note that, again by simple calculation, $I(\frac{1}{3}, 1) = \frac{6+5\delta}{27(1-\delta)}$. For all $\delta \in (0, 1)$, $I(\frac{1}{3}, 1) > I(\frac{2}{3})$. Therefore, $I(\frac{1}{3}) > \max\{I(\frac{2}{3}), I(1)\}$ and hence, again by Claim 1, $I(\frac{1}{3}) = I(\frac{1}{3}, 1)$.

We have therefore shown that, for all $\delta \in (0, 1)$,

$$I(\frac{1}{3}) > I(\frac{2}{3}) > I(1) = 0, \tag{5}$$

which means that among those workers who have obtained a positive skill level $x > 0$, priority is given to the least-skilled workers.

We now turn to state $x = 0$. As before, we begin by looking at the simple deterministic stopping rule $\tau = 1$, which yields $I(0, 1) = \frac{7\delta}{27(1-\delta)}$. Note, however, that for all $\delta \in (0, 1)$,

$$I(0, 1) < I(\frac{1}{3}). \tag{6}$$

Suppose that it were optimal for the principal to stop immediately upon reaching state $\frac{1}{3}$. By Claim 1, this would imply that $I(\frac{1}{3}) \leq I(0)$, and hence, by (5), $I(0) \geq I(x)$ for all $x > 0$. But by Claim 1, this would imply that it is optimal to stop after a single period of training, regardless of the realization of the new state. In this case, $I(0) = I(0, 1) \geq I(\frac{1}{3})$, which stands in contradiction to (6). Therefore, we can conclude that, under the optimal stopping rule, if training leads to skill level $\frac{1}{3}$, it is suboptimal to stop training the worker. By Claim 1, this also implies that $I(0) < I(\frac{1}{3})$.

Note that the principal clearly stops training upon reaching $x = 1$. In other words, unlike in $x \in \{\frac{1}{3}, \frac{2}{3}, 1\}$, the optimal stopping rule in state $x = 0$ is stochastic—whether the principal stops or not depends on the realization of the next state.

As we have shown that the optimal stopping rule does not stop in state $\frac{1}{3}$, and necessarily stops in state 1, it remains to check whether stopping in state $\frac{2}{3}$ is optimal (by Claim 1, stopping and continuing in state 0 are both optimal).

Denote by $\hat{\tau}$ the stopping rule by which the principal trains the worker until his skill strictly exceeds $\frac{1}{3}$, and then stops. By Claim 1, $\hat{\tau}$ is optimal if and only if

$$I(0,\hat{\tau}) \geq I(\frac{2}{3}). \tag{7}$$

Whether or not this inequality holds fully pins down the principal's optimal training strategy. By (5) and (6), for any $\delta \in (0,1)$ it is optimal for the principal to first attend to workers in state $\frac{1}{3}$. If (7) holds, the next to receive attention are completely unskilled workers ($x = 0$), and lastly workers in state $\frac{2}{3}$. If the inequality in (7) is reversed, workers in state $\frac{2}{3}$ will be attended to before those in state 0.

To solve for which values of δ (7) holds, we calculate

$$I(0,\hat{\tau}) = \frac{(1-\delta)\mathbb{E}(\sum_{s=0}^{\hat{\tau}-1}\delta^s x_s | x_0 = 0) + \mathbb{E}(\delta^{\hat{\tau}} x_{\hat{\tau}}^2 | x_0 = 0)}{\mathbb{E}(1-\delta^{\hat{\tau}})}.$$

First, observe that $x_{\hat{\tau}}$ is either $\frac{2}{3}$ or 1 with equal probability, regardless of whether it is reached from state 0 or from state $\frac{1}{3}$. Therefore, $\delta^{\hat{\tau}}$ and $x_{\hat{\tau}}^2$ are uncorrelated. It follows that $\mathbb{E}(x_{\hat{\tau}}^2) = \frac{1}{2}\cdot(\frac{2}{3})^2 + \frac{1}{2}\cdot 1^2 = \frac{13}{18}$, and hence $\mathbb{E}(\delta^{\hat{\tau}} x_{\hat{\tau}}^2 | x_0 = 0) = \frac{13}{18}\cdot\mathbb{E}(\delta^{\hat{\tau}} | x_0 = 0)$.

Second, note that

$$\mathbb{E}(\delta^{\hat{\tau}} | x_0 = 0) = \frac{1}{4}\delta\mathbb{E}(\delta^{\hat{\tau}} | x_0 = 0) + \frac{1}{4}\delta\mathbb{E}(\delta^{\hat{\tau}} | x_0 = \frac{1}{3}) + \frac{1}{2}\delta, \tag{8}$$

where the first, second, and third summands on the RHS of (8) correspond to different realizations of states $x_1 = 0$, $x_1 = \frac{1}{3}$, and $x_1 > \frac{1}{3}$, respectively, after one period of training. To find $\mathbb{E}(\delta^{\hat{\tau}} | x_0 = 0)$, we first derive $\mathbb{E}(\delta^{\hat{\tau}} | x_0 = \frac{1}{3})$. Similar to the logic of (8), we can write $\mathbb{E}(\delta^{\hat{\tau}} | x_0 = \frac{1}{3}) = \frac{1}{3}\delta\mathbb{E}(\delta^{\hat{\tau}} | x_0 = \frac{1}{3}) + \frac{2}{3}\delta$, which gives $\mathbb{E}(\delta^{\hat{\tau}} | x_0 = \frac{1}{3}) = \frac{2\delta}{3-\delta}$. Plugging this into (8) and rearranging terms yields $\mathbb{E}(\delta^{\hat{\tau}} | x_0 = 0) = \frac{6\delta}{\delta^2 - 7\delta + 12}$.

It is left to calculate $\mathbb{E}(\sum_{s=0}^{\hat{\tau}-1}\delta^s x_s | x_0 = 0)$. Writing the expression in a form similar to (8) and following the same steps as in the derivation of $\mathbb{E}(\delta^{\hat{\tau}} | x_0 = 0)$ gives $\mathbb{E}(\sum_{s=0}^{\hat{\tau}-1}\delta^s x_s | x_0 = 0) = \frac{\delta}{\delta^2 - 7\delta + 12}$. Substituting the derived terms into $I(0,\hat{\tau})$ we obtain

$$I(0,\hat{\tau}) = \frac{(16-3\delta)\delta}{3(12-\delta)(1-\delta)},$$

and hence (7) holds if and only if $\delta \geq \delta^* \equiv \frac{44-4\sqrt{70}}{17} \approx 0.62$. Note that for such values of δ, $I(0) = I(0,\hat{\tau})$. Analogous calculations can be performed to derive the index for lower values of δ; however, they are not required in order to fully understand the principal's training priority. □

We conclude this section by examining whether there exists a classic attention allocation problem (that is, one where payoffs depend only on the state of the alternative receiving attention) that yields the optimal dynamics that emerge in our setting. Denote the problem described in the previous section by $\mathcal{P}$, and consider the fictitious environment $\hat{\mathcal{P}}$ in which, in each period t, the DM's payoff is equal to

$$w_i(x_{i,t}) \equiv u_i(x_{i,t}) - v_i(x_{i,t}) + \frac{\delta}{1-\delta}\left(v_i(x_{i,t+1}) - v_i(x_{i,t})\right),$$

where i is the alternative that receives attention in period t. In particular, in this fictitious environment the DM's payoff does not depend on the states of other alternatives.

Denote by $\{x_j^\infty\}$ an entire sample path of x_j from its initial state onward, and denote by $\{x_j^\infty\}_{j=1}^n$ the collection of paths for all alternatives.

Proposition 1. *For any collection of realization paths $\{x_j^\infty\}_{j=1}^n$ and for any period t, the same alternative receives attention in $\mathcal{P}$ as in the fictitious problem $\hat{\mathcal{P}}$.*

The function w_i can be interpreted as follows. Suppose that when a DM attends to an alternative he gets the immediate payoff from that alternative and the discounted expected value of the change in the stream of payoffs, assuming that the alternative is not picked again. Thus, when the DM attends to an alternative, the payoff today from that alternative changes from v_i to u_i and, from the next period on, the per-period payoff changes from $v_i(x_t)$ to $v_i(x_{t+1})$. The reason we can transform the original problem into one where the classic Gittins index applies follows from the additive separability across alternatives and across periods. Thus, for the case of substitutes, the optimal policy may be obtained by showing indirectly that applying the Gittins index to the fictitious problem yields the optimal policy for the original problem. We opted to establish the optimal policy more directly via Theorem 1 in order to have comparable proofs for the two cases of substitutes and complements. The reason is that the above mentioned separability breaks down in the case of complementary alternatives (see Section 4). It remains an open question whether an analogue of Proposition 1 is true for the case of complements.

3.2 Applications

In this section, we apply our characterization of the DM's optimal policy to several environments.

3.2.1 Repeated bargaining

We now demonstrate how our framework may be helpful in analyzing the dilemma often faced by firms of whether to switch a supplier/contractor or remain with the current one. On the one hand, repeatedly using the same supplier enables the supplier to improve over time its process for producing the input needed by the firm. On the other hand, staying with the same supplier may improve the supplier's bargaining position over time, as his level of expertise increases. Firms may differ in their solutions to this dilemma, which raises the question of what factors favor one decision over the other. For instance, Helper and Levine (1992) noted that in the auto industry, contracts with suppliers are typically renegotiated each period, and while Japanese automakers tend to maintain long-term relationships with the same supplier, American automakers often switch between different suppliers.

Applying our framework to a simple stylized model of the firm's decision problem, we show that a firm either sticks with the same supplier or switches suppliers in every period. In particular, we show that if a firm finds it optimal to switch every period, then so will a more patient firm.

Consider a principal P and two identical agents, 1 and 2. In each period there is a project that the principal can assign to one of the agents. When an agent is assigned a project, the surplus he generates is a function of his current productivity. This productivity evolves over time, such that the more projects are assigned to an agent, the higher his productivity (possibly due to learning-by-doing). More specifically, we assume that the surplus from the k-th ($k \geq 1$) assigned project (the agent's productivity in state k) is equal to $\sum_{n=0}^{k-1} \theta^n$, where $\theta \in (0, 1)$.

The surplus from a project is divided between the principal and the agent to whom the project is assigned, and the division is determined through bargaining (the nature of which is described below). At the start of each period, the principal simultaneously bargains with each of the agents on the division of surplus in case the project is assigned to that agent. Given the bargaining outcome, the principal assigns the project to one of the agents and payoffs are realized. P seeks to maximize the discounted sum of his payoffs with the discount factor δ.

Following Stole and Zwiebel (1996b), we take a reduced-form approach to modeling the bargaining between P and the agents. That paper characterizes a profile of payoffs that is stable in the following sense: prior to production, no individual agent can benefit from renegotiating with the principal, and the principal cannot benefit from renegotiating with the other agent. In all such negotiations, the principal and the agents split the joint surplus from their relationship according to their respective (exogenously given) bargaining powers, and relative to their respective outside options. Following the approach of Stole and Zwiebel (1996b), we assume that agreements are non-binding in the sense that in case of disagreement with agent i, the principal can renegotiate with j, and players anticipate the possibility of such changes following disagreement. The outside option of each agent is normalized to zero. This would also be the outside option of the principal if there were only a single agent. But in the presence of two agents, the outside option for the principal when bargaining with i is the outcome of bargaining with the other agent, j, *in the absence of agent* i. Stole and Zwiebel (1996a) show that the stable payoff profile coincides with the unique subgame perfect equilibrium outcome of an extensive-form bargaining game.

We embed the characterization of stable payoff profiles due to Stole and Zwiebel (1996b) into our dynamic decision problem, such that if P assigns the project to agent i in period t, the payoffs to P and i are given by the corresponding stable payoff profile. Formally, let $\beta \in [\frac{1}{2}, 1]$ be P's bargaining power, so that each agent's bargaining power is $1-\beta$. To guarantee the validity of the bargaining solution (described below), we assume $\beta \geq \theta$.[9] Suppose that the project is assigned to agent i with productivity q_i. If there were no agent i, and P were to bargain only with j, the surplus to be divided would be q_j with outside options of zero for both P and j. This would yield a payoff of βq_j to P and $(1-\beta)q_j$ to agent j. It follows that in the solution to the bargaining between P and agent i with productivity q_i, the payoff to each side $h \in \{P, i\}$ is defined as follows:

$$\begin{aligned}&[h\text{'s outside option}] + [h\text{'s bargaining power}]\\ &\qquad \times (\text{surplus} - [P\text{'s outside option}] - [i\text{'s outside option}])\,.\end{aligned}$$

Hence, P's payoff is $\beta q_i + \beta\,(1-\beta)\,q_j$, while i's payoff is $(1-\beta)q_i - \beta\,(1-\beta)\,q_j$. Note that a necessary condition for this solution to be valid is that at any history, $(1-\beta)q_i - \beta\,(1-\beta)\,q_j \geq 0$. Otherwise, the bargaining solution is invalid. We can rewrite this condition as

$$\frac{q_i}{q_j} \geq 1-\beta. \tag{9}$$

When (9) is met, although the payoff in every period is obtained only from the selected supplier in that period, the functional form of the payoff allows us to apply Theorem 1.

[9] If $\beta < \theta$, then inequality (9) defined below will be violated in some cases.

Specifically, setting $u_i(q_i) = \beta q_i$ and $v_i(q_i) = \beta(1-\beta)\, q_i$ translates the problem into our general form. Intuitively, the "passive payoff" $v_j(\cdot)$ captures the externality of supplier j (which depends on j's productivity) in periods when i is chosen.

As long as P does not have all the bargaining power, he benefits from a relatively high outside option when bargaining with a supplier. On the one hand, this may motivate P to frequently switch between suppliers so as to maintain a high outside option with each. On the other hand, it may create an incentive to stick with one supplier for some time, enabling that supplier to increase his productivity, and then capitalize on this improvement by switching to the other supplier with an improved outside option. The next result shows that P either goes back and forth between the suppliers in each period or remains with the same one throughout, and derives a condition characterizing which policy is optimal as a function of the parameters.

Proposition 2. *It is optimal for P to assign the project to a different agent in each period if*

$$\delta \geq \frac{\beta}{1-\theta(1-\beta)}. \tag{10}$$

Otherwise, it is optimal to assign the project to the same agent in all periods.

Note that if P has full bargaining power ($\beta = 1$), it is optimal for him to choose a supplier and stick with him in all periods. However, when $\beta < 1$, the extent to which the principal can capitalize on the improvements in a supplier's productivity depends on his outside option, creating the incentive to switch between the suppliers. As can be seen in (10), the smaller P's bargaining power (β) or the slower the suppliers' improvement rate (θ), the stronger P's incentive to switch becomes, and therefore the less patient P must be in order for the strategy of continually switching between suppliers each period to become optimal.

The result suggests a possible channel through which greater bargaining power on the side of the firm (and/or a sufficiently fast learning-by-doing on the side of the suppliers) gives rise to the emergence of long-term exclusive relationships.

3.2.2 Career paths and mobility between sectors

The multi-armed bandit problem has been a natural framework for studying the dynamics of occupational choice. In an influential paper, Jovanovic (1979) uses a multi-armed bandit problem to study a model in which an individual sequentially chooses employment among multiple firms, and learns through specific experience how suited he is to a given job. The individual's optimal policy yields a decreasing hazard: the conditional probability of turnover falls as tenure increases.[10] Intuitively, the more experience the individual accumulates in a particular job, the more precise is his assessment of his competence in this job. Therefore, new information is less likely to affect this assessment and therefore less likely to cause the individual to leave his job. Miller (1984) enriches this model by introducing ex-ante heterogeneity in jobs, using the multi-armed bandit problem to characterize the dynamics of optimal job choice. Since the value of job-specific experience varies across jobs, jobs yielding riskier (but potentially higher) returns are experimented with earlier. Young, inexperienced workers therefore experiment more with such risky jobs.

In the above papers, an individual's expertise in a given job has no bearing on the returns from other jobs. Our framework enables us to extend the classic literature on dynamic occupational choice by allowing transfer of human capital across jobs. The extent to which accumulated human capital is transferable jobs is relevant for individuals' decisions to switch jobs between sectors and, in particular, between the private and public sectors. For instance, it is fairly common for academics and professionals (e.g., lawyers, accountants, economists, engineers) to switch from private firms to government departments/agencies (which oftentimes involves a salary cut) and then switch back to the private sector.

The following simple example illustrates how career paths that display these movements between sectors are captured by our framework. In each period, an individual can work in one of two sectors, A or B. In each period that he works in a sector, the individual accumulates human capital (measured in monetary units). A fraction of that human capital is transferable to the other sector. The individual's per-period payoff is equal to the accumulated human capital in the current sector plus the transferable portion of the human capital that he accumulated in the other sector.

The initial human capital in sector A is zero and each period of experience in that sector increases the human capital by $r > 0$. After a total of T periods of experience (not necessarily consecutive) in sector A, the total human capital reaches its maximal level of Tr. That is, denoting by $u_A(s)$ the total human capital accumulated in sector A after s periods of experience, we have $u_A(s) = (s+1)r$ in every state $s < T$, and $u_A(s) = Tr$ for

[10]The relationship between job-specific skills and turnover decisions has been central to the economics of labor mobility since the work of Becker (1962), Mincer (1962), and Oi (1962).

all $s \geq T$. Not all of the human capital accumulated in sector A is directly transferable to sector B. Specifically, if an individual with a total experience of s periods in sector A decides to switch to sector B, then only a portion α of his accumulated human capital is added to his accumulated human capital in the new sector. This transfer of human capital is modeled via the function v_A, such that $v_A(s) = \alpha s r$, where $\alpha \in (0,1)$ is the human capital that is transferred to sector B when the individual switches to that sector after accumulating $s \geq 0$ periods of experience in sector A.

To simplify the analysis, we assume that from the very first period of work in sector B, the human capital in that sector remains constant at $b > 0$, and that a portion β of it is transferable to sector A. Thus, $u_B(s) = b$ for every s, while $v_B(0) = 0$ and $v_B(s) = \beta b$ for $s > 0$ where $\beta \in (0,1)$.

Our objective is to illustrate that one important reason for switching sectors is to accumulate human capital that can be useful in another sector. For instance, a law graduate may enhance his future productivity in a private law firm by first starting out working in a public defender's office. An alternative career path may begin in a private law firm, followed by a move to the justice department, and then a return to a more senior position in a private law firm. To highlight the role of the transferable human capital, we assume that there is no uncertainty.[11]

To fully characterize the optimal career paths, we will use the following notation:

$$A_0 = \frac{r}{1-\delta} - \frac{rT\delta^T(1-\alpha)}{1-\delta^T} \ ; \ A_1 = (1-\alpha)\cdot Tr$$

$$B_0 = b \cdot \frac{1-\delta(1-\beta)}{1-\delta} \ ; \ B_1 = (1-\beta)\cdot b.$$

The following result provides necessary and sufficient conditions for each possible optimal career path, when the individual is sufficiently patient.

Proposition 3. *Assume that $\delta \geq \frac{(1-\alpha)T-1}{(1-\alpha)(T-1)}$. The optimal career paths are characterized as follows.*

(A) It is optimal to work only in A if and only if $A_1 \geq B_0$.

(B) It is optimal to work only in B if and only if $B_1 \geq A_0$.

(AB) It is optimal to start a career in A and then move to B and remain there if and only if $A_0 \geq B_0$ and $B_1 \geq A_1$.

[11] The model can be extended to allow for the combination of both learning about the quality of matches and transferable human capital.

(BA) It is optimal to start a career in B and then move to A and remain there if and only if $B_0 \geq A_0$ and $A_1 \geq B_1$.

(ABA) It is optimal to start a career in A, then move to B, and then return to A and remain there if and only if $A_0 \geq B_0 \geq A_1 \geq B_1$.

(BAB) It is optimal to start a career in B, then move to A, and then return to B and remain there if and only if $B_0 \geq A_0 \geq B_1 \geq A_1$.

Thus, our simple model admits several possible career paths. In particular, it accommodates career paths where the individual switches sector at most twice during his career. The contribution of the proposition is to give the precise conditions on the model's primitives that correspond to each possible path. The condition on the discount factor ensures that the index of sector A is minimal when the individual reaches the maximal human capital in that sector.

As a corollary, the above result also offers the following simple characterization of the individual's long-run occupation: in the long run, it is optimal for the individual to work in sector A if and only if

$$\frac{Tr}{b} \geq \frac{1-\beta}{1-\alpha}.$$

Hence, under the maintained assumption on δ, the sector where the individual will eventually work is determined by comparing the ratio of the long-term accumulated human capital in the two sectors with the ratio of the fraction of non-transferable human capital in the sectors.

One of the career paths described in the proposition consists of the individual switching to a stint in the public sector only after reaching a senior position in the private sector. This career path is consistent with the finding by Su and Bozeman (2009) that the probability of switching to the public sector is much higher for those who held a managerial position in their previous private sector job than for those who held professional and technical positions.

More generally, the framework allows us to extend the classic model of dynamic occupational choice (Jovanovic, 1979; Miller, 1984) to one that reflects the "skill-weights approach" introduced in Lazear (2009). This approach views skills as general, but assumes different jobs attach different weights to these skills. In such a model, the dynamics of occupational choice are driven by the combination of learning and the applicability of accumulated experience across jobs.

4 Attention scheduling with complementarities

In this section, we modify the model in Section 3 such that the DM's flow payoff is a *product* of the functions $u_i(\cdot)$ and $v_i(\cdot)$ of the alternatives, rather than their *sum*. That is, in each period t, if the DM chooses alternative i, his payoff is

$$U_t(x_{1,t}, ..., x_{n,t}) = u_i(x_{i,t}) \prod_{j \neq i} v_j(x_{j,t}),$$

where u_i and v_i are bounded positive functions, and $x_{i,t} \in X_i$ continues to represent the period-t state of alternative i.[12] Each X_i is an arbitrary state space.

This version of the model allows us to capture the problem of dynamically allocating attention among alternatives that are complements, rather than substitutes.[13] For example, the operation of an organization's separate divisions may involve various complementarities, such that if one of them is too far behind, the organization suffers as a whole. The case of complementary alternatives also fits situations in which a manager supervises a team that works on independent components of a single project, such that the project is completed successfully if and only if each component of it is completed successfully. In addition, this case can also address the problem of a firm producing a good according to a Cobb–Douglas production function

$$U_t = AL_t^{\beta} K_t^{\alpha}$$

that needs to decide in each period whether to invest in labor (L) or capital (K) with the goal of maximizing its expected discounted production.

4.1 The optimal policy

In contrast to the substitutes model of Section 3, characterizing the optimal policy for this problem requires making a distinction between the states of an alternative for which allocating attention to it is "augmenting," and ones in which it is not.

Definition 2. *Say that a state x_i is* augmenting *if there exists a stopping time τ such*

[12] As in the substitutes case, the possibility of not paying attention to any alternative can be captured by introducing a fictitious alternative whose state remains constant and for which the functions u and v are constant at 1.

[13] As discussed in the Introduction and illustrated in Section (4.2.1), the problem with complementarities cannot be converted into one with substitutes by taking a log transformation.

that $a_i(x_i, \tau) \equiv \mathbb{E}\left(\delta^\tau v_i(x_i^{+\tau})|x_i\right) - v_i(x_i) > 0$. *For each alternative* i, *denote by* $A_i \subseteq X_i$ *the set of states that are augmenting.*

In other words, a state of an alternative is augmenting if there exists a stopping time at which its expected discounted passive payoff increases. This property depends on both the process governing the evolution of the states of an alternative, as well as the alternative's current state.[14]

For any state $x_i \in A_i$ of alternative i, define the index

$$J_i(x_i) = \inf_\tau \left\{ \frac{\mathbb{E}\left(\sum_{s=0}^{\tau-1} \delta^s u_i(x_{i,s})|x_i\right)}{a_i(x_i, \tau)} \right\} \quad \text{s.t. } a_i(x_i, \tau) > 0. \tag{11}$$

Note that this index is independent of any information about all alternatives $j \neq i$, and that the denominator of the expression in the curly brackets is strictly positive when x_i is augmenting.

For any state $x_i \notin A_i$ of alternative i with $a_i(x_i, \tau) < 0$, define the index

$$J_i(x_i) = \sup_\tau \left\{ \frac{\mathbb{E}\left(\sum_{s=0}^{\tau-1} \delta^s u_i(x_{i,s})|x_i\right)}{-a_i(x_i, \tau)} \right\}, \tag{12}$$

and if $a_i(x_i, \tau) = 0$, define $J_i(x_i) = \infty$. Note that for states that are not augmenting, the denominator of (12) is non-negative.

A key difference between the indices (11) and (12) is that the optimization in (11) is constrained to stopping times τ for which $a_i(x_i, \tau) > 0$. The indices could be written more compactly as single expression to reflect this fact, but that would require allowing such an index to take both negative and positive values (while also making a distinction between the two cases of augmenting and nonaugmenting, since in the former the optimization is constrained). We therefore find it convenient to distinguish between the indices in both cases.

Definition 3. *Define the following order* $\succsim$, *according to which the indices of alternatives will be ranked.*[15] *For any alternatives* i, j *(including* $i = j$*):*

1. *If* $x_i \in A_i$ *and* $x_j \notin A_j$, *then* $J_i(x_i) \succsim J_j(x_j)$.

2. *If* $x_i \in A_i$ *and* $x_j \in A_j$, *then* $J_i(x_i) \succsim J_j(x_j)$ *if and only if* $J_i(x_i) \leq J_j(x_j)$.

[14] Note that since $v(\cdot)$ is bounded, an alternative cannot be augmenting in all states.

[15] Ties may be broken according to any prespecified tie-breaking rule.

3. If $x_i \notin A_i$ and $x_j \notin A_j$, then $J_i(x_i) \succsim J_j(x_j)$ if and only if $J_i(x_i) \geq J_j(x_j)$.

The distinction between augmenting and nonaugmenting states is not merely a technical one; it is at the heart of the tradeoff between the alternatives. If a state is augmenting, it strictly enhances the future benefits from allocating attention to the other alternatives—when we take into account discounting—and is therefore currently preferred to alternatives in states that are not augmenting.

Among alternatives in augmenting states, both the direct payoffs that such alternatives are expected to generate (captured by $\mathbb{E}\left(\sum_{s=0}^{\tau-1} \delta^s u_i(x_{i,s})|x_i\right)$) and the degree to which they are expected to enhance the payoffs from other alternatives in the future (captured by $a_i(x_i, \tau)$) must be weighed. In particular, the index (11) becomes more preferred (that is, its value decreases[16]) the greater is $a_i(x_i, \tau)$ and, perhaps surprisingly, less preferred (that is, its value increases) the greater is $\mathbb{E}\left(\sum_{s=0}^{\tau-1} \delta^s u_i(x_{i,s})|x_i\right)$. This reflects the fact that, among alternatives in augmenting states, the higher the direct payoffs an alternative generates, the more desirable it is to postpone allocating attention to it in order to allow such payoffs to first be enhanced by focusing attention on alternatives that are augmenting today but generate lower payoffs. Put differently, among alternatives in augmenting states, all things equal, it is desirable (in terms of the ordering of the indices) to *back-load* allocating attention to alternatives with higher direct payoffs.

By contrast, allocating attention to alternatives in nonaugmenting states does not enhance the flow payoffs from other alternatives—again, when we take into account discounting. This may be the case simply because $\mathbb{E}\left(v_i(x_i^{+\tau})|x_i\right) - v_i(x_i) < 0$ for any possible stopping time τ. Alternatively, even if an alternative is expected to enhance the future payoffs from allocating attention to the other alternatives—i.e., $\mathbb{E}\left(v_i(x_i^{+\tau})|x_i\right) - v_i(x_i) > 0$ for some τ—it may to do so to an extent that is not sufficient to outweigh the opportunity cost of not allocating attention to the other alternatives in the meantime, i.e., $\mathbb{E}\left(\delta^\tau v_i(x_i^{+\tau})|x_i\right) - v_i(x_i) < 0$. Accordingly, among alternatives in nonaugmenting states, the index (12) of an alternative is improving (i.e., its value increases) in the direct payoff it generates (captured by $\mathbb{E}\left(\sum_{s=0}^{\tau-1} \delta^s u_i(x_{i,s})|x_i\right)$) and deteriorating (i.e., its value decreases) in $-a_i(x_i, \tau) = v_i(x_i) - \mathbb{E}\left(\delta^\tau v_i(x_i^{+\tau})|x_i\right) > 0$. In contrast to the intuition for alternatives in augmenting states, among alternatives in nonaugmenting states, all things equal, it is desirable (in terms of the ordering of the indices) to *front-load* allocating attention to alternatives with high direct payoffs.

[16]Note the "inf" in the definition of (11).

Definition 4. *Denote by Λ^* the index policy induced by the order $\succsim$, that is, the policy that allocates attention in each period to the alternative with the "most preferred" index according to the order $\succsim$.*

As in Section 3.1, under the attention policy Λ^*, the decision of which alternative to choose boils down to a simple comparison of indices across pairs of alternatives, where each alternative's index is a function of its own state only, and independent of any information about other alternatives.

Theorem 2. *Λ^* is an optimal attention policy in the model with complementarities.*

To see how the indices (11)–(12) relate to the index (2), consider the index of an alternative i in a nonaugmenting state $x_{i,t}$. This index satisfies

$$J(x_{i,t}) \propto sup_\tau \left\{ \frac{\overbrace{\left(\frac{\mathbb{E}\left(\sum_{s=t}^{\tau-1} \delta^s u_i(x_{i,s})\right)}{\mathbb{E}(\sum_{s=t}^{\tau-1} \delta^s)} \right)}^{\text{(a)}}}{\underbrace{\left(\frac{v_i(x_{i,t}) - \mathbb{E}\left(\delta^\tau v_i(x_{i,t+\tau})\right)}{(1-\delta)\mathbb{E}(\sum_{s=t}^{\tau-1} \delta^s)} \right)}_{\text{(b)}}} \right\}.$$

The index therefore maximizes the *ratio* between the two components discussed in Section 3.1: (a) the expected discounted payoff per unit of expected discounted time, and (b) the net present value of the expected change in the passive payoff, again per unit of expected discounted time. Recall that the first component is the one maximized by the Gittins index, while the second reflects the fact that when an alternative does not receive attention, it continues to contribute to the overall payoff, as a function of its state. The index for the augmenting case can be rewritten analogously.

The case of complements is substantially different from the case of substitutes, as discussed above, and as illustrated in the applications below. Indeed, it remains an open question whether analogously to Proposition 1, there exists an auxiliary dynamic decision problem in which the DM gets a payoff only from the chosen task, and for which the classic Gittins index policy coincides with the optimal one we characterize.

We conclude this subsection with the following property of the optimal policy, which is useful for applications.

Proposition 4. *Assume that the state space of all alternatives is $\mathbb{N}$, with each state $s \in \mathbb{N}$ of an alternative representing the number of periods in which the alternative has received attention in the past. If v is concave then the alternative is augmenting in state s if and only if $a(s, 1) = \delta v(s + 1) - v(s) > 0$.*

4.2 Applications

4.2.1 On-the-job training with complementarities

Consider an environment in which two new workers require training by a principal. The workers' productivity is measured by the probability with which they successfully complete a task in a given period. Suppose that the workers must receive, in total, two periods of training from the principal before attaining full proficiency. The only difference between the workers is that, while the productivity of one remains constant before completing full training, the other already improves after a single period of training. Specifically, the following table describes the success probabilities of each of the workers in their periodic tasks, given their level of training, where $p \in (0, 1)$ and $q \in (p, \sqrt{p})$:

	A	B
0	p	p
1	q	p
2	1	1

Substitutes: Suppose first that the principal's payoff is equal to the aggregate output produced by both workers.

Claim 3. *In the case of substitutes, for all $\delta \in (0, 1)$, the optimal training schedule is to train worker A in the first two periods and then train worker B for two periods.*

To see this, note that

$$I_B(0) = \frac{\delta^2(1-p)}{1-\delta^2} < \max\left\{ I_A(0,1) = \frac{\delta(q-p)}{1-\delta}, I_A(0,2) = \frac{\delta^2(1-p)}{1-\delta^2} + \frac{\delta(q-p)}{1+\delta} \right\} = I_A(0),$$

so that, in period 1, the principal trains worker A. Moreover, for all $\delta \in (0, 1)$ and $q \in (p, \sqrt{p})$, $I_A(1) = \frac{\delta(1-q)}{1-\delta} > \frac{\delta^2(1-p)}{1-\delta^2} = I_B(0)$. Hence, the principal will train worker A also in period 2.

Complements: We now turn to consider the case where the principal receives a positive payoff (which we normalize to 1) only if both workers complete their tasks. If at least one of the workers fails to complete his task in a given period, the principal's payoff in that period is zero. It turns out that in this case the optimal training schedule depends on the principal's discount factor.

We begin by deriving the indices for untrained workers. We must first check whether there exists a stopping rule τ for which $a_i(0, \tau) > 0$. Since the transition between states in our example is deterministic, in order to determine if $i \in \{A, B\}$ in state 0 is augmenting, we only need to check whether $\max\{a_i(0,1), a_i(0,2)\} > 0$. First note that $a_B(0,1) = \delta p - p < 0$. Next, note that if $a_i(0,2) = \delta^2 - p < 0$, then $\delta < \sqrt{p}$, which, in turn, implies that $a_A(0,1) = \delta q - p < 0$ as, by assumption, $q < \sqrt{p}$. Hence, both workers are augmenting in state 0 if and only if $\delta^2 - p > 0$, i.e., $\delta > \sqrt{p}$.

Augmenting at zero ($\delta > \sqrt{p}$). Since only $a_B(0,2)$ is positive, the index for B is supported by the stopping rule $\tau = 2$:

$$J_B(0) = \frac{p + \delta p}{\delta^2 - p}.$$

For worker A, we must compare $\frac{\sum_{s=0}^{\tau-1} \delta^s x_{A,s}}{a_A(0,\tau)}$ for $\tau \in \{1, 2\}$. Since we are in the augmenting case, the index is given by the minimum of the two values, provided that it is positive. We therefore have

$$J_A(0) = \frac{p + \delta q}{\delta^2 - p}.$$

Since $q > p$, it follows immediately that $J_A(0) > J_B(0)$, and since we are in the augmenting case, it is optimal for the principal to train worker B first. For B's index in state 1, we only need to consider the stopping rule $\tau = 1$. Since $a_B(1,1) = \delta - p > \delta^2 - p = a_B(0,2) > 0$, B is augmenting in state 1, and hence

$$J_B(1) = \frac{p}{\delta - p}.$$

Finally, as $J_B(1) < J_B(0) < J_A(0)$, the principal will train B in period 2.

Nonaugmenting at zero ($\delta \leq \sqrt{p}$). In this case, the index of $i \in \{A, B\}$ in state 0 is given by (12) where the relevant stopping rules for comparison are $\tau \in \{1, 2\}$. It is thus straightforward to verify that

$$J_A(0) = \frac{p + \delta q}{p - \delta^2} \qquad \text{and} \qquad J_B(0) = \frac{p + \delta p}{p - \delta^2}.$$

Since $q > p$, clearly $J_A(0) > J_B(0)$. As we are in the nonaugmenting case, the principal will select worker A in period 1. To complete the analysis, we consider $J_A(1)$. Since $a_A(1,1) = \delta - q$, there are two cases:

$\delta \in (q, \sqrt{p})$: worker A in state 1 becomes augmenting. Since worker B is still in the nonaugmenting state 0, the principal will choose worker A in period 2.

$\delta \leq q$: worker A in state 1 is nonaugmenting. In this case, A's index in state 1 is $J_A(1) = \frac{q}{q-\delta}$. Since $J_A(1) > J_B(0)$ for all such δ, it is optimal to train A in the second period as well.

We have therefore shown the following.

Claim 4. *In the case of complements, the optimal training schedule depends on the principal's discount factor. If $\delta > \sqrt{p}$, the unique optimal policy is to first fully train worker B and then switch to worker A, whereas if $\delta < \sqrt{p}$, the unique optimal policy is to first fully train worker A and then switch to worker B.*

This example illustrates that the training schedule depends on whether the untrained workers are augmenting.[17] In the example, the total increase in the probability of success as a result of full training is identical (from p to 1 in two periods) for both workers. The only difference is that, for A, part of the increase is already attained after one period of training. The question is, when does the principal benefit the most from this intermediate increase (from p to q)?

In the case of substitutes, appropriating this extra gain as early as possible is optimal because of discounting. Hence, as seen in Claim 3, for all $\delta \in (0, 1)$ the principal trains A first. In the case of complements, the probability of B's success affects the expected benefit from training A in a multiplicative manner—in the same way the discount factor affects the (current) from future training of A. When an untrained worker B is augmenting, the effect of discounting (i.e., delaying the training of A) is weaker than that of increasing the success probability of B. Hence, to maximize the *current value* of the benefit from

[17] It is easy to see in this example why the case of complements cannot be solved by simply taking the log transformation of the periodic payoffs. To see this, note that if we had taken the log transformation of the periodic payoff, then for $\delta > \sqrt{p}$, a policy that first trains A for two consecutive periods followed by training of B for two periods is superior to the opposite policy, in contrast to what is obtained when the periodic payoff is the product of the probabilities.

the intermediate increase in A's probability of success (from p to q), the principal first trains B for two consecutive periods. When an untrained B is nonaugmenting, the effect of discounting dominates, and therefore, the result is similar to the case of substitutes.

4.2.2 Developing multiple complementary attributes

Oftentimes, decision-makers are faced with the problem of allocating time between the development of multiple complementary attributes. For example, a firm developing optimization software for routing or scheduling problems often needs to prioritize work on either the speed or the accuracy of its algorithms. Similarly, a video camera developer has to allocate time between increasing the image resolution and increasing the field of view. In addition, investment in different complementary skills requires a decision of when to focus on each skill. For instance, an army unit needs to decide when to focus on the physical fitness of its soldiers and when to focus on their sharpshooting skills.

In many of these examples, the time-allocation or sequencing decision is made under competitive pressure and uncertainty about when the product will need to come out, or when the combination of all the skills will need to be applied. Introduction of similar products by rival firms may lead a firm to release its product even when some of its features could still be further developed, and a threat by an enemy may require an army to deploy its unit in the middle of its training regimen.

In this subsection, we apply our framework to a simple example that illustrates the potential effect of this uncertainty on the optimal sequencing of investment in complementary features/skills that differ in their rates of improvement. While the literature has analyzed the impact of competition on the incentives to innovate, and on the intensity of innovation (see, e.g., Grossman and Shapiro, 1987; Harris and Vickers, 1987; Boone, 2001), the question of how competitive pressure—more specifically, the uncertainty that it induces—affects the development *process* (in particular, the scheduling of development across different features) itself is largely underexplored.[18]

Consider a firm that develops a product with two complementary attributes, X (speed) and Y (accuracy). In each period t it needs to decide whether to focus on developing X or Y, where developing a feature means that its value increases. We denote by $X(t)$ and $Y(t)$ the value of the attribute at time t (where a particular value is denoted by a lower

[18] A related problem studied in Poggi (2021) is how to allocate a given amount of resources between two projects when the amount of resources required to complete each project is unknown, and when the payoff from completing both projects is higher than the sum of the payoffs from completing only one project.

case letter, x or y). We assume that $X(t) \in \{1, 2, ...\}$, such that $X(t) = X(t-1) + 1$ if the firm invested in X at time $t-1$, and $X(t) = X(t-1)$ otherwise. By contrast, Y has only two possible values, 1 and $h > 1$. Initially, $Y(0) = 1$, and the value of Y remains $y = 1$ until the firm invests in the attribute for two periods (not necessarily consecutive), at which point its value increases to $y = h$, and remains h thereafter. This captures the idea that increasing accuracy requires more periods of investment, whereas speed can be increased incrementally each period.

The firm's profit from releasing the product to the market at time t is

$$U(t) = X(t)Y(t).$$

In each period, there is a probability $1 - \beta$ that the firm needs to release the product (ending its investment decision problem) because a rival firm has introduced a competing product. We therefore interpret β as the *intensity* of competition in the market. We focus on the case where

$$\beta < \frac{1}{\sqrt{h}}. \tag{13}$$

As we will see, this ensures that investment in accuracy is always nonaugmenting.

The firm's ex-ante expected profit is then equal to

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t (1-\beta) X(t) Y(t),$$

and its problem is to choose a policy—specifying which attribute to develop at each point in time—to maximize these expected profits.

Our objective is to understand the effect of the intensity of competition β on the sequencing of investment in the two attributes.

Proposition 5. *Consider two environments, one where the intensity of competition is β, and another with a higher intensity $\beta' > \beta$. Under the firm's optimal policy, investment in accuracy (Y) occurs weakly later in the first environment (i.e., with β) than in the second environment (i.e., with β').*

For example, if $h = 2$ and $\beta = 0.4$, the firm invests in X for three consecutive periods, and only then switches to Y, but if $\beta = 0.6$, it switches to Y already in the third period.

4.2.3 Supervising agents with stochastic costs of effort

The literature on moral hazard has focused on characterizing payment schemes that incentivize agents to exert effort. However, there are many environments in which workers get a fixed wage, and hence cannot be incentivized with monetary transfers that depend on their output (for instance, in the public sector). In these environments, a principal may need to supervise agents while they work on a task, in order to ensure that the task is completed successfully. If the principal is in charge of multiple agents, he must decide in each period which agent to supervise, taking into account the effect of repeated supervision on the agent's willingness to work when left unsupervised. For some agents, supervision can be constructive and helpful, while for others it may be perceived as an unwanted annoyance. The following simple example illustrates how our framework can be applied to characterize the principal's optimal policy in the presence of heterogeneous agents.

There are two agents who jointly work on a project. Each agent is in charge of a task, and the project is completed successfully if and only if both agents successfully complete their respective task. When an agent is supervised, he successfully completes his task with certainty. When left unsupervised, the agent faces a stochastic cost of effort in completing the task. That is, the agent's motivation for carrying out the task fluctuates (e.g., it may depend on his mood that day, which is affected by factors outside of his control), and his realized motivation determines his perceived cost of exerting effort. The agent exerts effort if and only if the realized cost does not exceed a threshold. Each agent i starts with an initial threshold c_i, which can change with the number of times in which he is supervised.

Assume that in each period an agent draws a cost from a uniform distribution on $[0, 1]$. Agent 1's cost threshold is zero, and remains constant regardless of the number of times he is supervised. Thus, agent 1 never works when left unsupervised. By contrast, agent 2's threshold as a function of the number of times he was supervised s is $v(s)$, where $v(\cdot)$ is increasing and concave. The interpretation is that supervision helps agent 2 to learn how to perform the task more efficiently, and hence with a lower perceived cost, but the returns to supervision are diminishing. The question we consider is: How should the principal optimally supervise the agents over time?

Proposition 6. *There exists a minimal integer T such that $a_2(T, 1) < 0$. Under the optimal policy, the principal supervises agent 2 for T consecutive periods and then switches* *to supervising agent 1 indefinitely.*

To illustrate the above result, let $v(s) = (\frac{1}{2})^{\frac{1}{s+1}}$. If $\delta > \sqrt{\frac{1}{2}}$, then $a_2(0,1) > 0$, and the principal will begin supervising agent 2, as he is augmenting. He will continue to do so until a state in which agent 2 is nonaugmenting. Agent 2 is nonaugmenting at s if $a(s,1) \leq 0$, i.e., if $\left(\frac{1}{2}\right)^{\frac{1}{(s+1)(s+2)}} \geq \delta$. If $\delta \leq \sqrt{\frac{1}{2}}$, then agent 2 is also nonaugmenting and it is optimal to supervise agent 1 only. For example, if $\delta = 0.99$, then the principal supervises agent 2 for six periods, after which he switches to supervise agent 1 in all periods. Hence, from period 7 onward, the expected per-period output is $\left(\frac{1}{2}\right)^{\frac{1}{7}} \approx 0.9$. Thus, if the principal is sufficiently patient, he will first "invest" in lowering the cost of agent 2, even at the expense of no output, and only then switch to supervising agent 1.

5 Concluding remarks

This paper characterizes the optimal solution to a new class of dynamic decision problems that encompass a broad variety of environments. In these problems, a DM chooses in each period which task to attend to, given a periodic payoff that is affected by the output of both the chosen and unchosen tasks. Despite the externality that unchosen tasks have on the periodic payoff, we show that the optimal strategy is characterized by an index policy whereby each task (in each state) is assigned a score that is independent of the other tasks, and the choice of task consists of comparing their indices. The fact that the optimal policy takes this "separable" form is important for applications: it is useful for deriving key properties of the dynamics and comparative statics under the optimal policy, as well as for computational purposes (avoiding the "curse of dimensionality").

Our characterization of the optimal policy may potentially open the door for the analysis of new decision problems beyond the classic multi-armed bandit paradigm. The characterization also suggests that perhaps a more general class of problems may admit a tractable solution: one where the DM's periodic payoff is the generalized mean of payoffs/outputs of all tasks. Hopefully, our solution to the case of the arithmetic and geometric means s useful hints for addressing this general case.

References

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy 70*(5, Part 2), 9–49.

Bergemann, D. and J. Valimaki (2008). Bandit problems. In: The *New Palgrave Dictionary of Economics*. Ed. by Steven N. Durlauf and Lawrence E. Blume. Basingstoke, UK: Palgrave Macmillan.

Boone, J. (2001). Intensity of competition and the incentive to innovate. *International Journal of Industrial Organization 19*(5), 705–726.

Bray, R. L., D. Coviello, A. Ichino, and N. Persico (2016). Multitasking, multiarmed bandits, and the Italian judiciary. *Manufacturing & Service Operations Management 18*(4), 545–558.

Che, Y.-K. and K. Mierendorff (2019). Optimal dynamic allocation of attention. *American Economic Review 109*(8), 2993–3029.

Coviello, D., A. Ichino, and N. Persico (2014). Time allocation and task juggling. *American Economic Review 104*(2), 609–623.

Coviello, D., A. Ichino, and N. Persico (2015). The inefficiency of worker time use. *Journal of the European Economic Association 13*(5), 906–947.

Eliaz, K. and A. Frug (2018). Bilateral trade with strategic gradual learning. *Games and Economic Behavior 107*, 380–395.

Fershtman, D. and A. Pavan (2020). Searching for arms: Experimentation with endogenous consideration sets. *Working paper*.

Fudenberg, D., P. Strack, and T. Strzalecki (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review 108*(12), 3651–3684.

Gittins, J. and D. Jones (1974). A dynamic allocation index for the sequential design of experiments. In J. Gani (Ed.). *Progress in Statistics*, pp. 241–266. Amsterdam: North-Holland.

Gossner, O., J. Steiner, and C. Stewart (2020). Attention please! *Econometrica*, forthcoming.

Grossman, G. M. and C. Shapiro (1987). Dynamic R & D competition. *Economic Journal 97*(386), 372–387.

Harris, C. and J. Vickers (1987). Racing with uncertainty. *Review of Economic Studies 54*(1), 1–21.

Helper, S. and D. I. Levine (1992). Long-term supplier relations and product-market structure. *Journal of Law, Economics, & Organization 8*(3), 561–581.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy 87*(5), 972–990.

Ke, T. T., Z.-J. M. Shen, and J. M. Villas-Boas (2016). Search for information on multiple products. *Management Science 62*(12), 3576–3603.

Ke, T. T. and J. M. Villas-Boas (2019). Optimal learning before choice. *Journal of Economic Theory 180*, 383–437.

Keller, G., S. Rady, and M. Cripps (2005). Strategic experimentation with exponential bandits. *Econometrica 73*(1), 39–68.

Klabjan, D., W. Olszewski, and A. Wolinsky (2014). Attributes. *Games and Economic Behavior 88*, 190–206.

Lazear, E. P. (2009). Firm-specific human capital: A skill-weights approach. *Journal of Political Economy 117*(5), 914–940.

Liang, A., X. Mu, and V. Syrgkanis (2021). Dynamically aggregating diverse information. *Working paper*.

Mandelbaum, A. (1986). Discrete multi-armed bandits and multi-parameter processes. *Probability Theory and Related Fields 71*(1), 129–147.

Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy 92*(6), 1086–1120.

Mincer, J. (1962). On-the-job training: Costs, returns, and some implications. *Journal of Political Economy 70*(5, Part 2), 50–79.

Oi, W. Y. (1962). Labor as a quasi-fixed factor. *Journal of Political Economy 70*(6), 538–555.

Poggi, F. (2021). The timing of complementary innovations. *Working paper*.

Puterman, M. L. (2014). *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons.

Radner, R. and M. Rothschild (1975). On the allocation of effort. *Journal of Economic Theory 10*(3), 358–376.

Stole, L. A. and J. Zwiebel (1996a). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies 63*(3), 375–410.

Stole, L. A. and J. Zwiebel (1996b). Organizational design and technology choice under intrafirm bargaining. *American Economic Review 86*(1), 195–222.

Su, X. and B. Bozeman (2009). Dynamics of sector switching: Hazard models predicting changes from private sector jobs to public and nonprofit sector jobs. *Public Administration Review 69*(6), 1106–1114.

Appendix

Proofs of Theorems 1 and 2. The proofs of both theorems follow the same steps. Whenever the details of the arguments depend on whether the alternatives are substitutes or complements, we give a separate argument for each case.

For any alternative k, let $\mathcal{I}_k$ denote the index I_k in the case of substitutes and J_k in the case of complements. Let $\triangleright$ denote the binary relation $\geq$ in the case of substitutes and $\succsim$ in the case of complements. Finally, let $\mathcal{P}^*$ denote the index policy Γ^* in the case of substitutes, and Λ^* in the case of complements.

Let π^0 be an attention policy that allocates attention to some alternative i in period 0 and then proceeds according to the attention policy $\mathcal{P}^*$ from period 1 onward. In order to prove the optimality of $\mathcal{P}^*$, it is enough to show that the expected discounted payoff under the policy π^0 is no greater than the expected discounted payoff under $\mathcal{P}^*$, given any initial state $(x_{1,0}, ..., x_{n,0})$ of the DM.[19]

Let $(x_{1,0}, ..., x_{n,0})$ be the initial state of the DM's problem. Consider the attention policy π^0. If π^0 allocates attention in period 0 to the same alternative $\mathcal{P}^*$ would have chosen, the two policies coincide in all periods. Therefore, suppose that π^0 allocates attention to alternative i in period 0, while $\mathcal{P}^*$ would have allocated attention to alternative $j \neq i$ in period 0. Note that this means that $\mathcal{I}_j(x_{j,0}) \triangleright \mathcal{I}_i(x_{i,0})$. Also note that despite the fact that π^0 proceeds according to $\mathcal{P}^*$ from period 1 onward, π^0 need not allocate attention to j in period 1, since the state of alternative i may change as a result of the attention it received in period 0.

[19]This follows from standard results in the literature on Markov decision processes.

Define $\tau_k^*(x_k) = \min\{t > 0 : \mathcal{I}_k(x_k) \triangleright \mathcal{I}_k(x_k^{+t}(x_k))\}$, where $x_k^{+t}(x_k)$ denotes alternative i's (stochastic) state after t periods of attention, starting in state x_k. In other words, beginning in state x_k, $\tau_k^*(x_k)$ is the first time at which the index $\mathcal{I}_k$ becomes weakly worse than $\mathcal{I}_k(x_k)$ according to $\triangleright$.[20]

Denote by σ_1 the stochastic time at which an alternative other than i receives attention under π^0. Without loss of optimality, we can assume that this will be alternative j, and as j has not been chosen yet, its state in period σ_1 is equal to that of period 0. Let $\tau_j^*(x_{j,0})$ be the optimal stopping time in the definition of the index of j given state $x_{j,0}$. Setting $\sigma_2 = \tau_j^*(x_{j,0})$, π^0 will therefore allocate attention to alternative j from period σ_1 until (at least) period $\sigma_1 + \sigma_2 - 1$. At time $\sigma_1 + \sigma_2$, the index of alternative i will be $\mathcal{I}_i(x_i^{+\sigma_1}(x_{i,0}))$, the index of alternative j will be $\mathcal{I}_j(x_j^{+\sigma_2}(x_{j,0}))$, and the index of all other alternatives will be $\mathcal{I}_k(x_{k,0})$.

The final step of the proof will rely on the following Lemma. The details of its proof depend on whether the alternatives are substitutes or complements. In order not to disrupt the flow of the proof of Theorems 1 and 2, we give the separate proof of each case after the final step of the proof of Theorems 1 and 2.

Lemma 1. *The expected payoff under π^1 is weakly greater than that under π^0.*

If π^1 coincides with $\mathcal{P}^*$ during the periods $\sigma_2, ..., \sigma_2 + \sigma_1 - 1$, then π^1 and $\mathcal{P}^*$ are identical and the proof is complete. Otherwise, we can modify π^1 to a new attention policy π^2, repeating the argument in the preceding paragraphs.[21] We can proceed inductively and construct a sequence of policies $(\pi^0, \pi^1, \pi^2, ...)$, such that: (i) given the initial state $(x_{1,0}, ..., x_{n,0})$, π^{s+1} yields an expected discounted payoff no smaller than π^s, and (ii) the expected discounted payoff under π^s converges to the expected discounted payoff under $\mathcal{P}^*$ as $s \to \infty$ (to see this, note that π^s coincides with $\mathcal{P}^*$ for at least the first s periods).

[20] In the case of substitutes, $\tau_k^*(x_k)$ attains the supremum in (2) by Claim 1. Similarly, in the case of complements, it can be shown that $\tau_k^*(x_k)$ attains the infimum of (11) or the supremum of (12).

[21] In particular, consider the vector of states of all of the alternatives, starting from $(x_{1,0}, ..., x_{n,0})$ and having followed $\mathcal{P}^*$ in periods $0, ..., \sigma_2 - 1$. Suppose $\mathcal{P}^*$ would proceed to choose alternative $k \neq i$ at this stage (it may or may not be the case that $k = j$). Let $\tau_k^*(x_{k,\sigma_2})$ be the optimal stopping time in the definition of the index of k given state x_{k,σ_2}. The attention policy π^1 therefore pays attention to alternative j during periods $0, ..., \sigma_2 - 1$, then to alternative i during $\sigma_2, ..., \sigma_2 + \sigma_1 - 1$, and then to alternative k during (at least) $\sigma_2 + \sigma_1 + \tau_k^*(x_{k,\sigma_2}) - 1$. Denoting $\sigma_3 = \sigma_2 + \tau_k^*(x_{k,\sigma_2})$, define π^2 as follows. First, π^2 follows $\mathcal{P}^*$ during periods $0, ..., \sigma_3 - 1$, then it chooses alternative i during $\sigma_3, ..., \sigma_3 + \sigma_1 - 1$, and from then on it proceeds according to $\mathcal{P}^*$. Following precisely the same argument as above, π^2 yields an expected discounted payoff no smaller than π^1.

It follows that the expected discounted payoff under π^0 is no greater than under $\mathcal{P}^*$, which completes the proof. ■

Proof of Lemma 1. The policies π^0 and π^1 coincide from period $\sigma_1 + \sigma_2$ onward. Therefore we focus on periods $0, ..., \sigma_1 + \sigma_2 - 1$.

The case of substitutes. Under π^1, the expected discounted payoff during periods $0, ..., \sigma_1 + \sigma_2 - 1$ is equal to

$$\mathbb{E}\left(\sum_{k \neq i,j} \delta^{\sigma_1+\sigma_2-1} v_k(x_{k,0})\right) + \mathbb{E}\left(\sum_{t=0}^{\sigma_2-1} \delta^t u_j(x_{j,t})\right) + \mathbb{E}\left(\sum_{t=0}^{\sigma_2-1} \delta^t v_i(x_{i,0})\right)$$
$$+ \mathbb{E}\left(\delta^{\sigma_2} \sum_{t=0}^{\sigma_1-1} \delta^t v_j(x_{j,\sigma_2})\right) + \mathbb{E}\left(\delta^{\sigma_2} \sum_{t=0}^{\sigma_1-1} \delta^t u_i(x_{i,t})\right).$$

Similarly, the expected payoff during periods $0, ..., \sigma_1 + \sigma_2 - 1$ under π^0 is equal to

$$\mathbb{E}\left(\sum_{k \neq i,j} \delta^{\sigma_1+\sigma_2-1} v_k(x_{k,0})\right) + \mathbb{E}\left(\sum_{t=0}^{\sigma_1-1} \delta^t u_i(x_{i,t})\right) + \mathbb{E}\left(\sum_{t=0}^{\sigma_1-1} \delta^t v_j(x_{j,0})\right)$$
$$+ \mathbb{E}\left(\delta^{\sigma_1} \sum_{t=0}^{\sigma_2-1} \delta^t v_i(x_{i,\sigma_1})\right) + \mathbb{E}\left(\delta^{\sigma_1} \sum_{t=0}^{\sigma_2-1} \delta^t u_j(x_{j,t})\right).$$

Subtracting the two and rearranging, we have

$$\mathbb{E}\left(\sum_{t=0}^{\sigma_2-1} \delta^t u_j(x_{j,t})\right)(1 - \mathbb{E}(\delta^{\sigma_1})) + \frac{v_i(x_{i,0})\mathbb{E}(1-\delta^{\sigma_2})}{1-\delta}$$
$$+ \mathbb{E}(\delta^{\sigma_2} v_j(x_{j,\sigma_2}))\, \mathbb{E}\left(\frac{1-\delta^{\sigma_1}}{1-\delta}\right) - \mathbb{E}\left(\sum_{t=0}^{\sigma_1-1} \delta^t u_i(x_{i,t})\right)(1 - \mathbb{E}(\delta^{\sigma_2}))$$
$$- \frac{v_j(x_{j,0})\mathbb{E}(1-\delta^{\sigma_1})}{1-\delta} - \mathbb{E}(\delta^{\sigma_1} v_i(x_{i,\sigma_1}))\, \mathbb{E}\left(\frac{1-\delta^{\sigma_2}}{1-\delta}\right).$$

To see that this expression is non-negative note that multiplying by $\frac{1-\delta}{\mathbb{E}(1-\delta^{\sigma_1})\mathbb{E}(1-\delta^{\sigma_2})}$ and rearranging, the difference can be written as

$$\frac{(1-\delta)\mathbb{E}\left(\sum_{t=0}^{\sigma_2-1} \delta^t u_j(x_{j,t})\right) + \mathbb{E}(\delta^{\sigma_2} v_j(x_{j,\sigma_2})) - v_j(x_{j,0})}{\mathbb{E}(1-\delta^{\sigma_2})}$$
$$- \left(\frac{(1-\delta)\mathbb{E}\left(\sum_{t=0}^{\sigma_1-1} \delta^t u_i(x_{i,t})\right) + \mathbb{E}(\delta^{\sigma_1} v_i(x_{i,\sigma_1})) - v_i(x_{i,0})}{\mathbb{E}(1-\delta^{\sigma_1})}\right).$$

This difference is non-negative since the first summand equals to $I_j(x_{j,0})$ (as σ_2 is an optimal stopping time) and the second summand is at most $I_i(x_{i,0})$ (as σ_1 is some stopping time), and $I_j(x_{j,0}) \geq I_i(x_{i,0})$. This completes the proof for the case of substitutes.

The case of complements. The expected payoff during periods $0, ..., \sigma_1 + \sigma_2 - 1$ under π^1 is equal to

$$\prod_{k \neq i,j} v_k(x_{k,0}) \left\{ v_i(x_{i,0}) \mathbb{E} \left(\sum_{t=0}^{\sigma_2 - 1} \delta^t u_j(x_{j,t}) \right) + \mathbb{E} \left(\delta^{\sigma_2} v_j(x_{j,\sigma_2}) \right) \mathbb{E} \left(\sum_{t=0}^{\sigma_1 - 1} \delta^t u_i(x_{i,t}) \right) \right\}. \tag{14}$$

Similarly, under π^0, the expected payoff during these periods is equal to

$$\prod_{k \neq i,j} v_k(x_{k,0}) \left\{ v_j(x_{j,0}) \mathbb{E} \left(\sum_{t=0}^{\sigma_1 - 1} \delta^t u_i(x_{i,t}) \right) + \mathbb{E} \left(\delta^{\sigma_1} v_i(x_{i,\sigma_1}) \right) \mathbb{E} \left(\sum_{t=0}^{\sigma_2 - 1} \delta^t u_j(x_{j,t}) \right) \right\}. \tag{15}$$

Denote by $\Delta(\pi^1, \pi^0)$ the difference between the expected discounted payoff under π^1 and its counterpart under π^0. Subtracting (15) from (14) and rearranging, we have that $\Delta(\pi^1, \pi^0)$ is equal to

$$\begin{aligned} \prod_{k \neq i,j} v_k(x_{k,0}) \Bigg\{ & \mathbb{E} \left(\sum_{t=0}^{\sigma_2 - 1} \delta^t u_j(x_{j,t}) \right) \left(v_i(x_{i,0}) - \mathbb{E} \left(\delta^{\sigma_1} v_i(x_{i,\sigma_1}) \right) \right) \\ & - \mathbb{E} \left(\sum_{t=0}^{\sigma_1 - 1} \delta^t u_i(x_{i,t}) | x_{i,0} \right) \left(v_j(x_{j,0}) - \mathbb{E} \left(\delta^{\sigma_2} v_j(x_{j,\sigma_2}) \right) \right) \Bigg\}. \end{aligned} \tag{16}$$

We now verify that $\Delta(\pi^1, \pi^0) \geq 0$. Recall that $J_j(x_{j,0}) \succsim J_i(x_{i,0})$. There are three possible cases.

Case 1. Suppose that $x_{j,0}$ is an augmenting state and $x_{i,0}$ is nonaugmenting. Then, by the definition of $\sigma_2 = \tau_j^*(x_{j,0})$, $v_j(x_{j,0}) - \mathbb{E} \left(\delta^{\sigma_2} v_j(x_{j,\sigma_2}) \right) < 0$, and by the definition of $J_i(x_{i,0})$, $v_i(x_{i,0}) - \mathbb{E} \left(\delta^{\sigma_1} v_i(x_{i,\sigma_1}) \right) \geq 0$. This guarantees that $\Delta(\pi^1, \pi^0) \geq 0$.

Case 2. Suppose that $x_{j,0}$ and $x_{i,0}$ are both augmenting. Then $J_j(x_j) \succsim J_i(x_i)$ implies that $J_i(x_i) \geq J_j(x_j)$. Furthermore, $v_j(x_{j,0}) - \mathbb{E} \left(\delta^{\sigma_2} v_j(x_{j,\sigma_2}) \right) < 0$ by the definition of σ_2. Suppose first that $v_i(x_{i,0}) - \mathbb{E} \left(\delta^{\sigma_1} v_i(x_{i,\sigma_1}) \right) < 0$. Then

$$\frac{\mathbb{E} \left(\sum_{t=0}^{\sigma_1 - 1} \delta^t u_i(x_{i,t}) \right)}{\mathbb{E} \left(\delta^{\sigma_1} v_i(x_{i,\sigma_1}) \right) - v_i(x_{i,0})} \geq J_i(x_{i,0}) \geq J_j(x_{j,0}) = \frac{\mathbb{E} \left(\sum_{t=0}^{\sigma_2 - 1} \delta^t u_j(x_{j,t}) \right)}{\mathbb{E} \left(\delta^{\sigma_2} v_j(x_{j,\sigma_2}) \right) - v_j(x_{j,0})}.$$

The first inequality follows from (11), while the equality follows from the fact that $\sigma_2 = \tau_j^*(x_{j,0})$ is the optimal stopping time in the definition of the index $J_j(x_{j,0})$. Rearranging and multiplying both sides by $\prod_{k\neq i,j} v_k(x_{k,0})$, we have that $\Delta(\pi^1,\pi^0) \geq 0$.

Now suppose that $v_i(x_{i,0}) - \mathbb{E}\left(\delta^{\sigma_1} v_i(x_{i,\sigma_1})\right) > 0$. Then

$$\frac{\mathbb{E}\left(\sum_{t=0}^{\sigma_1-1}\delta^t u_i(x_{i,t})\right)}{\mathbb{E}\left(\delta^{\sigma_1} v_i(x_{i,\sigma_1})\right) - v_i(x_{i,0})} \leq \frac{\mathbb{E}\left(\sum_{t=0}^{\sigma_2-1}\delta^t u_j(x_{j,t})\right)}{\mathbb{E}\left(\delta^{\sigma_2} v_j(x_{j,\sigma_2})\right) - v_j(x_{j,0})},$$

and once again $\Delta(\pi^1,\pi^0) \geq 0$.

Finally, if $v_i(x_{i,0}) - \mathbb{E}\left(\delta^{\sigma_1} v_i(x_{i,\sigma_1})\right) = 0$ then clearly $\Delta(\pi^1,\pi^0) \geq 0$.

Case 3. Suppose that $x_{j,0}$ and $x_{i,0}$ are both nonaugmenting. Then $J_j(x_j) \succsim J_i(x_i)$ implies that $J_i(x_i) \leq J_j(x_j)$. Furthermore, $v_i(x_{i,0}) - \mathbb{E}\left(\delta^{\sigma_1} v_i(x_{i,\sigma_1})\right) \geq 0$ and $v_j(x_{j,0}) - \mathbb{E}\left(\delta^{\sigma_2} v_j(x_{j,\sigma_2})\right) \geq 0$, and by (12),

$$\frac{\mathbb{E}\left(\sum_{t=0}^{\sigma_1-1}\delta^t u_i(x_{i,t})\right)}{v_i(x_{i,0}) - \mathbb{E}\left(\delta^{\sigma_1} v_i(x_{i,\sigma_1})\right)} \leq J_i(x_{i,0}) \leq J_j(x_{j,0}) = \frac{\mathbb{E}\left(\sum_{t=0}^{\sigma_2-1}\delta^t u_j(x_{j,t})\right)}{v_j(x_{j,0}) - \mathbb{E}\left(\delta^{\sigma_2} v_j(x_{j,\sigma_2})\right)}.$$

Rearranging and multiplying by $\prod_{k\neq i,j} v_k(x_{k,0})$, we have that $\Delta(\pi^1,\pi^0) \geq 0$, which completes the case of complements. ■

Proof of Proposition 1. Since the fictitious problem is a standard multi-armed bandit problem, its optimal policy is simply the well-known Gittins index policy, which chooses in each period t the alternative with the greatest Gittins index

$$GI_j(x_{j,t}) = \sup_\tau \left\{ \frac{\mathbb{E}\left(\sum_{s=0}^{\tau-1}\delta^s w_j(x_{j,t+s})\right)}{\mathbb{E}(\sum_{s=0}^{\tau-1}\delta^s)} \right\}.$$

By the definition of w_j,

$$\begin{aligned}
&GI_j(x_{j,t}) \\
&= \sup_\tau \frac{\mathbb{E}\left(\sum_{s=0}^{\tau-1}\delta^s\left(u_j(x_{j,t+s}) - v_j(x_{j,t+s}) + \frac{\delta}{1-\delta}\left(v_j(x_{j,t+s+1}) - v_j(x_{j,t+s})\right)\right)\right)}{\mathbb{E}\left(\sum_{s=0}^{\tau-1}\delta^s\right)} \\
&= \sup_\tau \frac{\mathbb{E}\left((1-\delta)\sum_{s=0}^{\tau-1}\delta^s u_j(x_{j,t+s}) + \sum_{s=0}^{\tau-1}\delta^s\left(-v_j(x_{j,t+s}) + \delta v_j(x_{j,t+s+1})\right)\right)}{\mathbb{E}\left(1-\delta^\tau\right)} \\
&= \sup_\tau \frac{\mathbb{E}\left((1-\delta)\sum_{s=0}^{\tau-1}\delta^s u_j(x_{j,t+s}) - \sum_{s=0}^{\tau-1}\delta^s v_j(x_{j,t+s}) + \sum_{s=1}^{\tau}\delta^s v_j(x_{j,t+s})\right)}{\mathbb{E}\left(1-\delta^\tau\right)}
\end{aligned}$$

$$= \sup_\tau \frac{\mathbb{E}\left((1-\delta)\sum_{s=0}^{\tau-1}\delta^s u_j(x_{j,t+s}) - v_j(x_{j,t}) + \delta^\tau v_j(x_{j,t+\tau})\right)}{\mathbb{E}\,(1-\delta^\tau)}$$

$$= I_j(x_{j,t}).$$

Hence, the Gittins indices GI_j in the fictitious environment coincide with the indices I_j.

Therefore, since by Theorem 1 the attention policy Γ^* is optimal for the problem $\mathcal{P}$, and since the policy based on the indices GI_j is optimal for the fictitious problem $\hat{\mathcal{P}}$, this proves the result. ■

Proof of Proposition 2. We first assume that inequality (9) is satisfied, and then verify that this is indeed the case in the solution. Let $x_{i,t}$, the state of agent i in period t, be defined as the number of times agent i was assigned the task up to and including period t. Then

$$q_i(x_{i,t}) = \sum_{n=0}^{x_{i,t}-1} \theta^n.$$

Let $I_i(k)$ denote the index of agent i who is currently in state k and denote

$$f(\tau|k) \equiv \frac{(1-\delta)\left[\sum_{t=0}^{\tau-1}\delta^t\beta q_i(k+t)\right] + \left[\delta^\tau\beta\,(1-\beta)\,q_i(k+\tau) - \beta\,(1-\beta)\,q_i(k)\right]}{1-\delta^\tau}.$$

Then by our definition of u_i and v_i,

$$I_i(k) = \sup_\tau\,[f(\tau|k)] = f(\tau^*|k).$$

Some tedious algebra establishes that for every $\tau > 1$,

$$\begin{aligned} & f(\tau|k) - f(\tau-1|k) \\ & = \beta\theta^k\delta^{\tau-1}\,(\beta-\delta+\theta\delta-\theta\beta\delta)\left[\frac{(1-\theta\delta)(1-\theta^{\tau-1}) - \delta(1-\theta)(1-\theta^{\tau-1}\delta^{\tau-1})}{(1-\theta)\,(1-\delta\theta)\,(1-\delta^\tau)\,(1-\delta^{\tau-1})}\right]. \end{aligned}$$

Note that the term in square brackets on the RHS is positive if and only if

$$\sum_{n=0}^{\tau-2}\theta^n > \delta\sum_{n=0}^{\tau-2}\theta^n\delta^n.$$

Hence, it is positive for all $\theta, \delta > 0$. It follows that $f(\tau|k) - f(\tau-1|k) < 0$ for all $\tau > 1$ if and only if $\beta - \delta + \theta\delta - \theta\beta\delta > 0$, which holds if and only if (10) holds. Hence, $\tau^* = 1$ if and only if this condition holds; otherwise, $\tau^* = \infty$.

Suppose that (10) holds so that $\tau^* = 1$. It is optimal to switch agents each period if and only if

$$I(x_k) - I(x_{k+1}) = -\theta^k \frac{\beta}{1-\delta}\left(\beta - \delta + \theta\delta - \theta\beta\delta\right) \geq 0$$

for every state k. Since (10) holds, this inequality is satisfied.

If (10) is violated, then $\tau^* = \infty$, and it is optimal to remain with one agent for all periods.

It remains to verify that (9) is satisfied. Assume first that (10) holds. Then every period P switches an agent. This means that in any given period, there is some $k > 1$ such that $\max\{q_1, q_2\} = \sum_{n=0}^{k} \theta^n$ while $\min\{q_1, q_2\} = \sum_{n=0}^{k-1} \theta^n$. Hence, (9) is satisfied if and only if

$$\frac{\sum_{n=0}^{k-1} \theta^n}{\sum_{n=0}^{k} \theta^n} \geq 1 - \beta,$$

which is satisfied if and only if

$$\frac{\beta}{1-\beta} > \frac{\theta^k}{\sum_{n=0}^{k-1} \theta^n}.$$

Since $\beta \geq \frac{1}{2}$, the LHS is at least one, while the RHS is smaller than one.

Assume next that (10) is violated. Then it is optimal for P to remain with the same agent every period. Hence, (9) is satisfied if

$$\frac{1}{\lim_{k\to\infty} \sum_{n=0}^{k-1} \theta^n} \geq 1 - \beta$$

or, equivalently, if $1 - \theta \geq 1 - \beta$. ∎

Proof of Proposition 3. The proof proceeds in two steps. First we derive the indices for the two sectors, and then we apply them to characterize the optimal career path.

Step 1: First we show that the indices for the sector A are as follows:

$$I_A(s) = \begin{cases} \frac{r}{1-\delta} + rs(1-\alpha) - \frac{(1-\alpha)r(T-s)\delta^{(T-s)}}{1-\delta^{(T-s)}} & , \quad s < T \\ (1-\alpha)Tr & , \quad s \geq T, \end{cases}$$

while the indices for sector B are as follows:

$$I_B(s) = \begin{cases} b\left[\frac{1-(1-\beta)\delta}{1-\delta}\right] & , \quad s = 0 \\ (1-\beta)b & , \quad s > 0. \end{cases}$$

The index for sector B is relatively simple to derive. For $s > 0$,

$$I_B(s) = \sup_{\tau} \frac{(1-\delta)\sum_{k=0}^{\tau-1}\delta^k b + \delta^\tau \beta b - \beta b}{1-\delta^\tau} = (1-\beta)b,$$

while for $s = 0$,

$$I_B(0) = \sup_{\tau} \frac{(1-\delta)\sum_{k=0}^{\tau-1}\delta^k b + \delta^\tau \beta b}{1-\delta^\tau} = \sup_{\tau}\left[b + \frac{\delta^\tau}{1-\delta^\tau}\beta b\right] = b\left[\frac{1-(1-\beta)\delta}{1-\delta}\right].$$

We next turn to sector A. For $s \geq T$ the index is simple to derive:

$$I_A(s) = \sup_{\tau} \frac{(1-\delta)\sum_{k=0}^{\tau-1}\delta^k Tr + \delta^\tau \alpha Tr - \alpha Tr}{1-\delta^\tau} = (1-\alpha)Tr.$$

To derive the index for $s < T$, note that, for $s + \tau \leq T$,

$$I_A(s,\tau) = \frac{(1-\delta)\sum_{k=0}^{\tau-1}\delta^k (s+k+1)r + \delta^\tau \alpha(s+\tau)r - \alpha sr}{1-\delta^\tau},$$

and the RHS of the above expression reduces to $\frac{r}{1-\delta} + rs(1-\alpha) - \frac{(1-\alpha)r\tau\delta^\tau}{1-\delta^\tau}$, which is increasing in τ. Hence, since $I_A(s)$ is constant for $s \geq T$, it follows that for $s < T$,

$$I_A(s) \in \{I_A(s, T-s), I_A(s,\infty)\}.$$

The assumption that $\delta \geq \frac{(1-\alpha)T-1}{(1-\alpha)(T-1)}$ implies that $I_A(0,1) \geq I_A(T)$. Thus, $I_A(s) = I_A(s, T-s)$, which completes Step 1.

Step 2: First note that by Step 1, $A_0 = I_A(0) = I_A(0,T)$, which implies that if the individual starts working in A, he will remain there for T consecutive periods. In addition, Step 1 implies that $A_1 = I_A(T)$, $B_0 = I_B(0)$, and $B_1 = I_B(1)$. Finally, $B_0 > B_1$ and by the assumption on δ, $A_0 > A_1$. ■

Proof of Proposition 4. Assume that $a(s,\tau+1) = \delta^{\tau+1}v(s+\tau+1) - v(s) > 0$. Since

$$\begin{aligned} &\delta^{\tau+1}v(s+\tau+1) - v(s) \\ &= \delta^\tau\left[\delta v(s+\tau+1) - v(s+\tau)\right] + \delta^{\tau-1}\left[\delta v(s+\tau) - v(s+\tau-1)\right] + .. + \left[\delta v(s+1) - v(s)\right] \end{aligned}$$

$$< \delta^\tau \left[\delta v(s+1) - v\left(s\right)\right] + \delta^{\tau-1}\left[\delta v(s+1) - v\left(s\right)\right] + .. + \left[\delta v\left(s+1\right) - v\left(s\right)\right]$$
$$= \left(\delta^\tau + ... + 1\right)\left[\delta v\left(s+1\right) - v\left(s\right)\right],$$

we have that $\delta v\left(s+1\right) - v\left(s\right) > 0$. ■

Proof of Proposition 5. Since Y requires two periods of investment in order to increase its value from 1 to h, it is nonaugmenting if and only if $\beta^2 h - 1 \leq 0$. By (13), this inequality necessarily holds. Denote by 0 the state of attribute Y before it has received any periods of attention.

We first show that $J_Y(0) = J_Y(0, \tau = 2)$. That is, we show that the optimal stopping time in the definition of the index $J_Y(0)$ specifies investment in the attribute Y for exactly two periods. To see this, note that $J_Y(0, \tau = 2) > J_Y(0, \tau)$ for all $\tau > 2$ if and only if

$$\frac{1+\beta}{1-\beta^2 h} > \frac{1+\beta+\beta^2 h \frac{1-\beta^{\tau-2}}{1-\beta}}{1-\beta^\tau h},$$

which reduces to $h > 1$, and therefore clearly holds. It is also easy to verify that $J_Y(0, \tau = 2) > J_Y(0, \tau = 1)$. Hence,

$$J_Y(0) = \frac{1+\beta}{1-\beta^2 h}.$$

We next turn to the attribute X. Clearly, we can identify the state of the attribute X by its value, x. We now derive $J_X(x)$. By Proposition 4, X is nonaugmenting in state x if and only if $a(x,1) = \beta(x+1) - x \leq 0$. Thus, x is nonaugmenting if and only if

$$\beta \leq \frac{x}{x+1}. \tag{17}$$

Hence, for any given β, for a sufficiently large value x, X is nonaugmenting. If $X(t) = x$ and is augmenting in that state, then at t, under the optimal policy, the firm will necessarily develop X, since Y is nonaugmenting. It will continue to develop X until it reaches a state in which X is nonaugmenting, at which time the firm chooses between X and Y according to the higher index (as both attributes are nonaugmenting). Let us then derive $J_X(x)$ for states in which it is nonaugmenting.

Let x be a state for which X is nonaugmenting. We claim that $J_X(x,1) > J_X(x,\tau)$

for all $\tau > 1$. To see this, note that

$$J_X(x,\tau) = \frac{\sum_{t=0}^{\tau-1} \beta^t (x+t)}{x - \beta^\tau (x+\tau)} = \frac{x \cdot \frac{1-\beta^\tau}{1-\beta} + \frac{(\tau-1)\beta^{\tau+1} - \tau\beta^\tau + \beta}{(1-\beta)^2}}{x - \beta^\tau (x+\tau)},$$

where

$$\frac{\partial}{\partial \tau} \left(\frac{s \cdot \frac{1-\beta^\tau}{1-\beta} + \frac{(\tau-1)\beta^{\tau+1} - \tau\beta^\tau + \beta}{(1-\beta)^2}}{s - \beta^\tau (s+\tau)} \right) = \frac{\beta^{\tau+1}}{(1-\beta)^2} \cdot \frac{\tau \ln \beta - \beta^\tau + 1}{(\beta^\tau \tau - s + s\beta^\tau)^2}.$$

Since

$$\frac{\partial}{\partial \tau} (\tau \ln \beta - \beta^\tau + 1) = -(\ln \beta)(\beta^\tau - 1) < 0$$

and since $\ln \beta - \beta + 1 < 0$, it follows that $J_X(x,\tau)$ is decreasing in τ. Therefore, we have shown that

$$J_X(x) = \frac{x}{x - \beta(x+1)}.$$

To complete the proof, we show that if for x and β such that X is nonaugmenting, $J_X(x) < J_Y(0)$, then this inequality continues to hold for $\beta' > \beta$.[22]

First, note that if $h = 1 + (1+\beta)/x\beta$, then $J_X(x) = J_Y(0)$. If we now increase β, both sides of the equation increase, and the difference between the changes in $J_Y(0)$ and $J_X(x)$ is equal to

$$\begin{aligned} \frac{\partial}{\partial \beta} \left(\frac{1+\beta}{1-\beta^2 h} \right) - \frac{\partial}{\partial \beta} \left(\frac{x}{x - \beta(x+1)} \right) &= \frac{\beta h (\beta+2) + 1}{(1-\beta^2 h)^2} - \frac{x(x+1)}{(\beta - x + x\beta)^2} \\ &= \frac{x}{(\beta+1)(\beta - x + x\beta)^2} \\ &> 0, \end{aligned}$$

where the last equality follows from substituting $h = 1 + (1+\beta)/x\beta$. If we now increase h to above $1 + (1+\beta)/x\beta$, then $\frac{\partial}{\partial \beta}\left(\frac{x}{x-\beta(x+1)}\right)$ remains unchanged, but

$$\frac{\partial}{\partial x} \frac{\partial}{\partial \beta} \left(\frac{1+\beta}{1-\beta^2 x} \right) = \frac{\partial}{\partial x} \left(\frac{\beta x (\beta+2) + 1}{(x\beta^2 - 1)^2} \right) = \frac{\beta (x\beta^3 + 2x\beta^2 + 3\beta + 2)}{-(x\beta^2 - 1)^3} > 0.$$

Therefore, if whenever $J_X(x) = J_Y(0)$ increasing β leads to a greater rise in $J_Y(0)$ than in $J_X(x)$, then when $J_X(x) < J_Y(0)$, the increase in $J_Y(0)$ is even greater. This means that if the firm switches from X to Y at some time t for some value of β, it will not

[22] Note that if x and β are such that X is augmenting, then X is also augmenting under x and $\beta' > \beta$. That is, X becomes nonaugmenting under β sooner than it does under $\beta' > \beta$.

make the switch at a later time for $\beta' > \beta$. ■

Proof of Proposition 6. Let the state s of an agent denote the number of times he was supervised. Since $a_1(s, \tau) = 0$ for all (s, τ), in any state s, agent 1 is nonaugmenting and has an infinite index. Thus, whenever agent 2 is in an augmenting state it is optimal to supervise him; otherwise, it is optimal to supervise agent 1. Since $v(s)$ is concave, it follows from Proposition 4 that for all s, $a_2(s, \tau) > 0$ if and only if $a_2(s, 1) > 0$. Hence, agent 2 is augmenting in state s if and only if $\frac{v(s+1)}{v(s)} > \frac{1}{\delta}$, and he cannot be augmenting in all states since this would imply that $lim_{s \to \infty} v(s) > 1$. ■