DISCUSSION PAPER SERIES

No. 9460

RELATIONAL KNOWLEDGE TRANSFERS

Luis Garicano and Luis Rayo

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9460.asp**

ISSN 0265-8003

RELATIONAL KNOWLEDGE TRANSFERS

Luis Garicano, London School of Economics and CEPR
Luis Rayo, London School of Economics

Discussion Paper No. 9460
May 2013

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INDUSTRIAL ORGANIZATION**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Luis Garicano and Luis Rayo

CEPR Discussion Paper No. 9460

May 2013

ABSTRACT

Relational Knowledge Transfers*

An expert must train a novice. The novice initially has no cash, so he can only pay the expert with the accumulated surplus from his production. At any time, the novice can leave the relationship with his acquired knowledge and produce on his own. The sole reason he does not is the prospect of learning in future periods. The profit-maximizing relationship is structured as an apprenticeship, in which all production generated during training is used to compensate the expert. Knowledge transfer takes a simple form. In the first period, the expert gifts the novice a positive level of knowledge, which is independent of the players' discount rate. After that, the novice's total value of knowledge grows at the players' discount rate until all knowledge has been transferred. The inefficiencies that arise from this contract are caused by the expert's artificially slowing down the rate of knowledge transfer rather than by her reducing the total amount of knowledge eventually transferred. We show that these inefficiencies are larger the more patient the players are. Finally, we study the impact of knowledge externalities across players.

JEL Classification: c73, j24 and l14
Keywords: general human capital, knowledge, relational contracts and skills

Luis Garicano
Center for Economic Performance
London School for Economics (LSE)
Houghton Street
London WC2A 2AE

Email: l.garicano@lse.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=145086

Luis Rayo
Department of Management
London School of Economics
Houghton Street
London WC2A 2AE

Email: l.rayo@lse.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=163971

*We thank Steve Tadelis and participants at the 2013 meeting of the NBER Working Group on Organizational Economics for valuable suggestions.

Submitted 23 April 2013

Relational Knowledge Transfers*

Luis Garicano   Luis Rayo
London School of Economics

April 23, 2013

Abstract

An expert must train a novice. The novice initially has no cash, so he can only pay the expert with the accumulated surplus from his production. At any time, the novice can leave the relationship with his acquired knowledge and produce on his own. The sole reason he does not is the prospect of learning in future periods. The profit-maximizing relationship is structured as an apprenticeship, in which all production generated during training is used to compensate the expert. Knowledge transfer takes a simple form. In the first period, the expert gifts the novice a positive level of knowledge, which is independent of the players' discount rate. After that, the novice's total value of knowledge grows at the players' discount rate until all knowledge has been transferred. The inefficiencies that arise from this contract are caused by the expert's artificially slowing down the rate of knowledge transfer rather than by her reducing the total amount of knowledge eventually transferred. We show that these inefficiencies are *larger* the more patient the players are. Finally, we study the impact of knowledge externalities across players.

I Introduction

Teaching involves a transfer of human capital, and thus, unlike in other market transactions, it does not create a collaterizable asset. Moreover since the human capital of the student ('novice') is low at the start, he often does not have the means to pay in cash for the knowledge transfer. In this case, the absence of collateral means the rate of knowledge transfer is constrained not just by the learning technology, but also by the need to ensure the apprentice has the means to pay. Unlike in the first-best allocation, where knowledge is transferred as fast as technologically feasible, without contractual enforcement the rate of knowledge transfer is the result of a trade-off: the faster it is, the sooner valuable output can be generated; but if it is too fast, the novice will leave without compensating the teacher. For example, in a consulting firm, novices, usually called "associates," do not leave as long as they are learning enough and have a valuable enough promise at the end,

*We thank Steve Tadelis and participants at the 2013 meeting of the NBER Working Group on Organizational Economics for valuable suggestions.

the partnership. Or, in a international joint venture in a developing country, the local partner could potentially ignore the agreements in the IJV and establish his own operations, but will not do so as long as the future value of the remaining technology transfer is high enough to justify the continuation of the relationship. In these cases, the rate of knowledge transfer by the expert is constrained by the need to ensure a high enough continuation value.

The object of this paper is to study, in the absence of formal contracting, the form and duration of the relationship between expert and novice, as well as the determinants of the rate of knowledge transfer and its implications for efficiency.

To do this, we set up a simple model of the relationship between an expert and a novice. The expert has a stock of knowledge which he may use to teach the novice. Players can make spot swaps of knowledge for cash, but cannot commit to future promises. There are no technological limits to the rate at which knowledge can be transferred. Thus, in the unconstrained first-best allocation, all knowledge transfer occurs in the first period. Any slower rate of knowledge transfer is attributable to contractual difficulties.

Initially, the novice has no knowledge and no cash. At this stage, standard market contracting is worthless: the expert cannot transfer his knowledge in exchange for a payment, as knowledge is not contractible and the novice cannot commit to paying back in the future. Instead, the players can engage in a relational contract, where the relationship itself sustains knowledge transfers by providing an expectation of future payments. Such payments are sustained in turn by the expectation of further knowledge transfers from the expert.

A relational contract may in principle take many forms. In one extreme, which we call an apprenticeship, the novice captures no rents for the duration of his training. In the other extreme, all value created throughout the training is captured by the novice. We refer to intermediate cases, in which rents are shared between novice and expert, as "implicit partnerships."

To build intuition, we first solve a model in which knowledge transfers may occur in at most two periods. In the profit maximizing contract, there are two spot transfers of knowledge, one in each period. In the first period the spot transfer is for free, as the novice has no cash; the expert expects to get paid in the second period. At the start of the second period there are two relevant contracting constraints. The swap of knowledge for cash must be incentive compatible, that is the novice must prefer to pay, and stay in the relationship in order to receive the second part of the knowledge transfer, rather than quit with the knowledge he already received in the first period and produce on his own. Second, the contract must call for a cash transfer that is feasible given the money the novice has earned after the first period, using the knowledge acquired during that period. Moreover, the contract is structured as an apprenticeship, where the novice keeps no rents until knowledge is fully transferred.

This two-period model raises two questions. First, if the expert is not restricted to two transfer periods, is an apprenticeship still the optimal contractual form? Second, if so, what is the optimal duration of an apprenticeship?

We show that an apprenticeship is indeed the optimal contractual form. Intuition is as follows. Assume, for the sake of simplicity, that there is no discounting. Consider an implicit partnership instead of an apprenticeship, in which the novice accumulates some cash balances on the path of play. The expert can do better using the following alternative contract, which delivers the same payoff to the novice and leaves all his incentive constraints unaffected. Reduce the novice's current cash balance by one dollar and fully compensate the novice as follows. First, increase his current knowledge by a small amount, while keeping all future levels of knowledge constant (in other words, accelerate knowledge transfer), and let him keep the immediate incremental output generated. Second, one period in the future transfer back to the novice one dollar *minus* the incremental output already received by the novice. By accelerating knowledge transfer, additional surplus is created. As a result, the net effect for the expert is a higher cash flow at no additional cost.

Next, we examine the optimal rate of knowledge transfer. Firstly, we show that there is an interior solution. The duration of training cannot be infinite, since the apprentice is giving up increasingly large amounts of output per period to get additional knowledge, and thus in exchange he must be receiving, to maintain incentive compatibility, increasing knowledge transfers over time. Since total knowledge is finite, a full transfer of knowledge must occur in finite time. On the other hand, since the novice initially has no cash, training must last longer than a single period.

Secondly, the rate of knowledge transfer decreases (and the contract duration increases) in the discount factor (namely, the players' degree of patience). The loss of a longer contract duration for the expert is that less knowledge, and therefore less cash, is transferred between players each period; the gain is that there is one additional period in which the expert receives a money transfer from the apprentice. When players are more patient the loss shrinks, as the apprentice places a higher value on the promised future payoff and therefore knowledge can be transferred more quickly. Moreover, the gain grows as the expert places a higher value on the transfer taking place in the additional period.

The above result implies that the optimal contract becomes *less* efficient as players become more patient. The opposite is true in the social planner's constrained-efficient contract. The difference between the planner's problem and the expert one is that the planner only needs to extract enough output to ensure that the participation constraint of the expert is met. Since knowledge can be transferred more quickly when the novice is more patient (more cash can be credibly extracted), enough output to compensate the expert can be extracted more quickly.

We also extend the model to introduce knowledge externalities across players. An example of positive externalities is the case of team production, where the knowledge of the novice aids the expert's own production – for example, the novice asks fewer questions as his knowledge grows, leaving more time for the expert to help other novices (see Garicano, 2000, for a model where managers help workers along these lines). An example of negative externalities is the case of market rivalry, where the expert and novice compete for a fixed set of customers. We show that the smaller the externality the slower the rate of knowledge transfer.

The human capital acquisition literature, since Becker's (1994(1963)) classic analysis, shows firms will not pay for general human capital acquisition – if they were to do so they could not recoup the investment, as workers could work for another firm for a better wage. But efficiency will follow since workers have the right incentives to pay for their own human capital acquisition, for example by working for a lower wage. A large literature has tried to explain under these circumstances firms incentives to train in general skills by relying on market imperfections which allow firms to recoup the investment in training. This imperfections may be: imperfect competition for workers (e.g. Stevens, 1994, Acemoglu, 1997, and Acemoglu and Piscke, 1999); asymmetric information about training (e.g. Katz and Zimmerman, 1990, Chang and Wang, 1995, and Acemoglu and Pishke, 1998); or matching frictions (Burdett and Smith, 1996, and Loewenstein and Spletzer, 1998). By contrast, in our analysis there is no asymmetric information or lack of competition that allows the firm to extract returns from general training. It is simply the timing of training, with bigger promises of training coming along, that allows general training to be provided.[1]

A different literature studies the complementary problem of firms' credible promises to reward workers' investments in specific human capital. Prendergast (1993) suggests promotions may be a solution if promotion leads to an efficient assignment of jobs to workers only if the workers actually did train. Similarly, as Kahn and Huberman (1988) and Waldman (1990) show, an up-or-out rule can also lead to credible promises, even if the promoted worker has similar productivity in all jobs. Instead, in our paper careers endogenously involve increasingly large and increasingly valuable knowledge transfers promised for the next period to ensure the worker does not quit.

Malcomson et al. (2003) is closer to our work, as it has general training being the result of a relational contract. However, they assume the timing of training is exogenously determined to take place at the start of the training contract, whereas the key issue in our analysis is to determine the amount of training per period and the overall duration of training. In Malcomson et al., the reason workers do not quit is because of adverse selection – only the worker's current employer can see the level of training received by the worker. Finally, they assume firms can commit to a wage level post training, whereas in our model there is no commitment on either side.

In addition, our work is related to the literature on principal-agent models with relational contracts, in which, akin to our model, self-enforcing rewards motivate the agent (e.g. Bull, 1987, Spear and Srivastava, 1987, MacLeod and Malcomson, 1989, 1998, Baker, Gibbons and Murphy, 1994, Pearce and Stacchetti, 1998, and Levin, 2003). This literature focuses on eliciting a costly, productive effort from the agent while treating the agent's skill level as stationary and exogenous. In

[1] Alternatively, in the learning by doing models (following e.g. Heckman (1971),Weiss (1972),Rosen (1972), Killingsworth (1982), Shaw (1989) skill accumulation is a byproduct of work. Unlike in the learning by doing models, in our work the principal can control the rate of learning— being involved in production does not automatically determine the learning rate.

contrast, we treat the agent's skill as persistent and endogenous while assuming away effort costs.[2]

A related literature in finance studies dynamic contracting in environments in which the principal funds a project and the agent can privately divert cash flows at the expense of investors. Incentives are sustained through threats of contract termination, such as withholding future funding and seizing assets (e.g. DeMarzo and Sannikov, 2006, Biais et al., 2007, DeMarzo and Fishman, 2007). In these models the agent cannot expropriate a productive asset (or, say, contribute diverted cash flows to finance a new project), whereas in our model the agent can leave the relationship at any time with her accumulated knowledge and enjoy its full net present value.

Also related is Hörner and Skrzypacz (2010), which shows that in an environment without property rights a seller of information benefits from gradual revelation. In this paper, gradual revelation arises from the value of information being privately known by the seller. In our model, in contrast, the value of information is known to all and gradual transmission arises instead from the buyer being liquidity-constrained.

The remainder of the paper is structured as follows. Section II sets up the baseline model. Sections III and IV present our core analysis and findings. Section V discusses constrained-optimal contracts. Section VI studies four extensions of the baseline model. First, it considers the case in which the novice is endowed, up front, with a positive level of cash. Second, it considers the case of externalities between the expert and the novice – substitutabilities (like in competitive settings) and complementarities (like in team or hierarchy settings). Third, it considers the case in which knowledge has a degree of relationship specificity. Fourth, it considers competition in a model with multiple experts and novices in which experts are restricted to training at most one novice each. Section VII applies our model to three problems: training in professional service firms, the regulation of apprenticeships, and knowledge transfers in joint ventures. Finally, section VIII concludes.

II Baseline Model

There are two risk-neutral players: an expert (E) and a novice (N). Players interact over infinite periods $t = 1, 2, ...$ and discount future payoffs using a common discount factor $\delta \in (0, 1)$. The expert initially possesses one unit of general-purpose knowledge. This knowledge is divisible and can be transferred from the expert to the novice at any speed desired by the expert. Let x_t denote the fraction of knowledge transferred during period t and let X_t denote the novice's total knowledge, inclusive of x_t, in period t:

$$X_t = x_t + X_{t-1},$$

with $X_0 = 0$.

[2] A related literature in finance studies dynamic contracting in environments in which the agent can privately diverting cash flows at the expense of investors. Incentives are sustained through threats of contract termination, such as withholding futiure fundind and seizing assets (e.g.). In contrast,

The novice produces output $f(X_t)$ during period t, with f increasing and $f(0) = 0$.[3] This output belongs to the novice. After production, the novice splits his output between consumption $c_t \geq 0$ and savings. The novice has a bank account that pays gross interest $1+r = \frac{1}{\delta}$ between periods. Let B_t denote the balance of this bank account at the end of period t. The novice has no initial capital ($B_0 = 0$) and no borrowing ability ($B_t \geq 0$). In section VI we extend our model to the case in which $B_0 > 0$.

Each period has two stages:[4]

1. *Transaction stage.* Players make simultaneous proposals regarding a knowledge transfer x_t and a money transfer m_t from the novice to the expert. Formally, a proposal for player $i = E, N$ is a pair (x_t^i, m_t^i). A proposal is feasible if $x_t^i \leq 1 - X_{t-1}$ (namely, the expert has a sufficient stock of remaining knowledge) and $m_t^i \leq \frac{1}{\delta} B_{t-1}$ (namely, the novice has sufficient capital). If $(x_t^E, m_t^E) = (x_t^N, m_t^N)$ an agreement is reached and the corresponding knowledge and money transfers take place. Otherwise, no agreement is reached and $x_t = m_t = 0$.

2. *Production/consumption stage.* Output $f(X_t)$ is realized and the novice selects c_t. A consumption level is feasible if $0 \leq c_t \leq f(X_t) + \frac{1}{\delta} B_{t-1} - m_t$. The novice's savings at the end of period t satisfy
$$B_t = f(X_t) - c_t + \frac{1}{\delta} B_{t-1} - m_t.$$

All choices and output levels are publicly observable, but non-verifiable. The only formal contracts that can be written between players are the spot contracts described in the transaction stage.

For the time being, we assume that the expert faces zero costs when transferring knowledge to the novice. This assumption rules out, in particular, transaction costs associated to knowledge transfers as well as externalities (pecuniary or otherwise) experienced by the expert as the novice acquires knowledge. In section VI we extend the model to allow for a class of externalities.

From the standpoint of the start of period t, the payoffs for expert and novice are, respectively,

$$\begin{aligned} \Pi_t &= \sum_{\tau=t}^{\infty} \delta^{\tau-t} m_\tau, \text{ and} \\ V_t &= \sum_{\tau=t}^{\infty} \delta^{\tau-t} c_\tau = \frac{1}{\delta} B_{t-1} + \sum_{\tau=t}^{\infty} \delta^{\tau-t} \left[f(X_\tau) - m_\tau \right]. \end{aligned}$$

A *relational contract* (or more briefly, a *contract*) prescribes, on the path of play, a triple $\langle X_t, m_t, c_t \rangle$ for every period t and, upon any deviation from that path, a suspension of all further transactions between the players. For notational simplicity we denote a contract by its prescribed actions on the path of play: $\mathcal{C} = \langle X_t, m_t, c_t \rangle_{t=1}^{\infty}$. We say that a contract is feasible if it meets the

[3] See footnote 6 for the case in which $f(0) > 0$.

[4] The order in which these two stages take place is immaterial. The critical assumption is that spot trades of money and knowledge are feasible.

basic requirements described above and, in addition, it constitutes a subgame-perfect equilibrium of the dynamic game.

Given a feasible contract $\mathcal{C}$, denote the associated equilibrium values of Π_t, V_t, and B_t, respectively, $\Pi_t(\mathcal{C})$, $V_t(\mathcal{C})$, and $B_t(\mathcal{C})$. Feasibility of $\mathcal{C}$ requires that the following three constraints are met. First, the expert obtains a non-negative payoff from every t period onwards:

$$\Pi_t(\mathcal{C}) \geq 0 \text{ for all } t. \tag{Π_t}$$

Second, the novice has sufficient liquidity to make each transfer m_t:

$$m_t \leq \frac{1}{\delta} B_{t-1}(\mathcal{C}) \text{ for all } t. \tag{L_t}$$

Third, before the transaction stage of every period, the novice prefers to carry on with the prescribed actions over walking away with his current stock of knowledge and his current bank account:[5]

$$\frac{1}{1-\delta} f(X_{t-1}) + \frac{1}{\delta} B_{t-1}(\mathcal{C}) \leq V_t(\mathcal{C}) \text{ for all } t. \tag{IC_t}$$

Definition 1

*a. Period t is a **training period** of contract $\mathcal{C}$ if $X_{t-1} < X_{\sup}(\mathcal{C})$, where $X_{\sup}(\mathcal{C}) = \lim_{k\to\infty} X_k$.*

*b. The **training phase** of $\mathcal{C}$ consists of all training periods of $\mathcal{C}$.*

Definition 2 *Contract $\mathcal{C}$ is an **apprenticeship** if $m_t = \frac{1}{\delta} B_{t-1}(\mathcal{C})$ during every training period in $\mathcal{C}$.*

In what follows we assume that the discount factor is "generic" in the following sense: there is no $n \in \{1, 2, ...\}$ such that $\delta = \frac{n-1}{n}$. This assumption simplifies the analysis by ruling out (knife-edge) cases in which there are multiple optimal contracts for the expert.

III Preliminaries: short-duration contracts

To build intuition, we begin by restricting attention to simple contracts in which (1) the expert transfers all knowledge to the novice either during the first period or during the first two periods and (2) all money transfers are nonnegative (money flows only from the novice to the expert).

A. One training period

[5] A simple observation is in order. Given the general-purpose nature of knowledge, it is immaterial for the model whether the residual claimant of output is the novice (as we have assumed) or the expert. In the former case, the novice's key deviation consists in walking away from the relationship during the transaction phase of a given period t before making any money transfer to the expert. In the latter case, the novice's key deviation is identical except for the fact that it is initiated during the production/consumption phase of period $t-1$, with the novice producing output $f(X_{t-1})$ on his own rather than for the expert. In present value, both deviations lead to the same payoff.

Consider a contract in which all knowledge is transferred during the transaction stage of period 1, namely, $X_1 = 1$. In this case, the novice produces $f(1)$ during the production/consumption stage of period 1 and also produces $f(1)$ during the production/consumption stage of each period after that. Since the novice enters period 1 with zero capital, the expert receives a zero money transfer in period 1. In each subsequent period, the novice has no reason to transfer money to the expert (as he has no further knowledge to gain in return) and so the expert receives a zero money transfer as well. As a result, the overall payoff for the expert is zero and the overall payoff for the novice is the full present value of output $\frac{1}{1-\delta}f(1)$. This contract is ideal for the novice and also maximizes total surplus, but is undesirable for the expert.

B. Two training periods

Consider now a contract in which all knowledge is transferred during the first two periods, namely, $0 < X_1 < X_2 = 1$. In this case, the novice produces $f(X_1)$ during period 1 and produces $f(1)$ during each period after that. The only period in which the expert may hope to receive a positive money transfer is period 2, in exchange for knowledge transfer $1 - X_1$.

The expert faces two relevant constraints. First, the second-period money transfer cannot exceed the novice's bank balance at the beginning of period 2:

$$m_2 \leq \frac{1}{\delta}\left[f(X_1) - c_1\right], \tag{i}$$

where $f(X_1) - c_1$ are the novice's savings at the end of period 1. Second, from the standpoint of the beginning of period 2, the novice must prefer to continue with the relationship (which involves paying m_2 for knowledge $1 - X_1$ and enjoying the present discounted value of output $f(1)$) over leaving the relationship (which involves keeping her full bank balance and enjoying the present discounted value of the lower output $f(X_1)$):

$$\frac{1}{\delta}\left[f(X_1) - c_1\right] - m_2 + \frac{1}{1-\delta}f(1) \geq \frac{1}{\delta}\left[f(X_1) - c_1\right] + \frac{1}{1-\delta}f(X_1),$$

which simplifies to

$$m_2 \leq \frac{1}{1-\delta}\left[f(1) - f(X_1)\right]. \tag{ii}$$

The profit-maximizing contract solves the simple problem

$$\max_{X_1, c_1} \; \delta m_2$$
$$s.t. \; (i) \text{ and } (ii).$$

Notice that, without loss, the expert can set $c_1 = 0$, which maximally relaxes constraint (i) without affecting any other aspect of the problem.

Consequently, the expert's problem further simplifies to

$$\max_{X_1} \ \delta m_2$$
$$s.t. \ m_t \leq \min\left\{\frac{1}{\delta}f(X_1), \frac{1}{1-\delta}\left[f(1)-f(X_1)\right]\right\},$$

where the first term in the min operator is the R.H.S of (i) (for $c_1 = 0$) and the second term is the R.H.S of (ii).

Since the first term in the min operator is increasing in X_1 and the second term is decreasing in X_1, the solution involves selecting X_1 such that both terms are equal. Letting X_1^* denote the optimal value of X_1, we obtain $f(X_1^*) = \delta f(1)$, which in turn delivers an optimized payoff $\delta m_2 = \delta f(1)$ for the expert.

Intuitively, raising X_1 increases the novice's bank account, from which he can pay the expert for the remaining knowledge $(1 - X_1)$, but raising X_1 also increases the novice's incentive to walk away before the contract has been completed. The optimal X_1^* is just high enough so that the novice is willing to pay the entire first-period output $f(X_1^*)$ (plus interest) to the expert in period 2 in exchange for all remaining knowledge.

The above discussion provides several clues that are valuable when solving for the unrestricted optimal contract. First, training takes at least two periods to complete. Second, until training is complete, the novice enjoys zero consumption and uses all her output to pay the expert for additional knowledge. Third, until training is complete, the novice's liquidity and incentive constraints both bind.

IV Optimal contract

This section solves for the expert's optimal contract. First, we present three preliminary lemmas describing necessary features of any optimal contract. Second, we present a characterization theorem showing that any optimal contract involves full knowledge transfer in finite time and takes the form of an apprenticeship with zero consumption during the training phase. Third, we show that there is a unique optimal contract and we present this contract in closed form. Finally, we discuss the efficiency properties of the optimal contract and the role of δ.

We work with a relaxed problem for the expert in which her own continuation-payoff constraints Π_t are ignored. As shown in corollary 1 below, ignoring such constraint is without loss. The relaxed problem is

$$\max_{\mathcal{C}} \ \sum\nolimits_{t=2}^{\infty} \delta^{t-1} m_t \tag{I}$$

$$s.t$$

$$m_t \leq \frac{1}{\delta} B_{t-1}(\mathcal{C}) \text{ for all } t, \tag{L_t}$$

$$\frac{1}{1-\delta} f(X_{t-1}) + \frac{1}{\delta} B_{t-1}(\mathcal{C}) \leq V_t(\mathcal{C}) \text{ for all } t. \tag{IC_t}$$

We say that contract $\mathcal{C}^* = (X_t^*, m_t^*, c_t^*)$ is optimal if it solves the above problem.

Lemma 1 *Let $\mathcal{C}^*$ be an optimal contract. During every period before the end of the training phase (i.e. every period such that $X_t^* < X_{\sup}(\mathcal{C}^*)$) the novice transfers the full balance of his bank account to the expert (i.e. $m_t^* = \frac{1}{\delta} B_{t-1}(\mathcal{C}^*)$).*

Proof. Suppose toward a contradiction that there is a t such that $X_t^* < X_{\sup}(\mathcal{C}^*)$ and yet $m_t^* < \frac{1}{\delta} B_{t-1}(\mathcal{C}^*)$. There are two possibilities: $X_t^* < X_{t+1}^*$ and $X_t^* = X_{t+1}^*$. The argument that follows considers only the former case ($X_t^* < X_{t+1}^*$). Appendix A considers the general case ($X_t^* \leq X_{t+1}^*$).

Without loss, set $c_k^* = 0$, and therefore $V_k(\mathcal{C}^*) = \delta^{t+1-k} V_{t+1}(\mathcal{C}^*)$, for all $k \leq t+1$. Now consider an alternative contract $\mathcal{C}' = \left\langle X_t^{'}, m_t^{'}, c_t^{'} \right\rangle_{t=1}^{\infty}$ that is identical to $\mathcal{C}^*$ except for the following: for some small $\varepsilon > 0$, (i) $m_t' = m_t^* + \varepsilon$, (ii) $m_{t+1}' = m_{t+1}^* - \frac{1}{\delta}\varepsilon + \frac{1-\delta}{\delta}\varepsilon$, and (iii) $f(X_t') = f(X_t^*) + (1-\delta)\varepsilon$.

Notice that

$$\begin{aligned} V_{t+1}(\mathcal{C}') &= \underbrace{\frac{1}{\delta}\left[\frac{1}{\delta} B_{t-1}(\mathcal{C}') - m_t^{'} + f(X_t^{'})\right]}_{B_t(\mathcal{C}')} + \left[f(X_{t+1}^{'}) - m_{t+1}'\right] + \sum_{\tau \geq t+2} \delta^{\tau-(t+1)} \left[f(X_\tau') - m_\tau'\right] \\ &= \underbrace{\frac{1}{\delta}\left[\frac{1}{\delta} B_{t-1}(\mathcal{C}^*) - m_t^{*} + f(X_t^{*})\right]}_{B_t(\mathcal{C}^*)} + \left[f(X_{t+1}^{'}) - m_{t+1}^*\right] + \sum_{\tau \geq t+2} \delta^{\tau-(t+1)} \left[f(X_\tau') - m_\tau'\right] \\ &= V_{t+1}(\mathcal{C}^*). \end{aligned}$$

$\mathcal{C}'$ satisfies all constraints in the relaxed problem:

1. L_t holds since ε is small.

2. IC_t holds as all terms are unchanged.

3. L_{t+1} holds as both the L.H.S. and R.H.S. of this constraint increase by $-\frac{1}{\delta}\varepsilon + \frac{1-\delta}{\delta}\varepsilon$.

4. IC_{t+1} holds since

$$\frac{1}{1-\delta}f(X_t') + \frac{1}{\delta}\underbrace{\left[\frac{1}{\delta}B_{t-1}(\mathcal{C}') - m_t' + f(X_t')\right]}_{B_t(\mathcal{C}')}$$

$$= \frac{1}{1-\delta}f(X_t^*) + \frac{1}{\delta}\underbrace{\left[\frac{1}{\delta}B_{t-1}(\mathcal{C}^*) - m_t^* + f(X_t^*)\right]}_{B_t(\mathcal{C}^*)}$$

$$\leq V_{t+1}(\mathcal{C}^*) = V_{t+1}(\mathcal{C}').$$

5. All other liquidity and incentive constraints hold as their terms are unaffected.

Finally, note that $\mathcal{C}'$ increases the expert's payoff by

$$\delta^{t-1}\left(m_t' - m_t^*\right) + \delta^t\left(m_{t+1}' - m_{t+1}^*\right) = \delta^{t-1}(1-\delta)\varepsilon > 0,$$

a contradiction. ■

Intuition for lemma 1 is as follows. Suppose there is a period t in which (1) the novice has not yet received the totality of knowledge that the contract will eventually grant him ($X_t^* < X_{\sup}(\mathcal{C}^*)$) and (2) the novice does not transfer the full balance of his bank account to the expert. Then, the expert can increase her payoff by modifying the contract as follows. Request, during the transaction stage of period t, a slightly higher transfer from the novice. In return, and in a manner that leaves the novice exactly indifferent, grant the novice a slightly higher knowledge transfer during period t and request a slightly lower transfer during period $t+1$. By virtue of accelerating the knowledge transfer, this modification leads to higher surplus – which the expert keeps for herself owing to the fact that the novice is left indifferent. Notice, in particular that, in present value, the increase in the period t transfer is larger in magnitude than the reduction in the period $t+1$ transfer.

Lemma 2 *Every optimal contract $\mathcal{C}^*$ has a finite training phase (i.e. there exists a period T such that $X_T^* = X_{\sup}(\mathcal{C}^*)$).*

Proof. Suppose not. Select an arbitrary period K and, without loss, set $c_t^* = 0$, and therefore $V_t(\mathcal{C}^*) = \delta^{K-t}V_K(\mathcal{C}^*)$, for all $t < K$. From lemma 1 we have $m_t^* = \frac{1}{\delta}B_{t-1} = \frac{1}{\delta}f(X_{t-1}^*)$ for all $t < K$, and $m_t^* = \frac{1}{\delta}B_{t-1} = \frac{1}{\delta}f(X_{t-1}^*) - c_{t-1}^*$ for all $t \geq K$. We can write the expert's payoff as

$$\sum\nolimits_{t=2}^{K}\delta^{t-2}f(X_{t-1}^*) + \sum\nolimits_{t=K+1}^{\infty}\delta^{t-2}\left[f(X_{t-1}^*) - c_{t-1}^*\right],$$

where the values of X^*_{t-1} and c^*_{t-1} satisfy

$$\begin{aligned}\frac{1}{1-\delta}f(X^*_{t-1})+\frac{1}{\delta}f(X^*_{t-1}) &\leq \delta^{K-t}V_K(\mathcal{C}^*) \text{ for all } t<K\\ \frac{1}{1-\delta}f(X^*_{t-1})+\frac{1}{\delta}B_{t-1}(\mathcal{C}^*) &\leq V_t(\mathcal{C}^*) \text{ for all } t\geq K\end{aligned}$$

Since V_K is bounded above by $\frac{1}{1-\delta}f(X_{\sup}(\mathcal{C}^*))$, the expert's payoff is bounded above by

$$\begin{aligned}&\sum\nolimits_{t=2}^{K}\delta^{K-1}f(X_{\sup}(\mathcal{C}^*))+\sum\nolimits_{t=K+1}^{\infty}\delta^{t-2}f(X_{\sup}(\mathcal{C}^*))\\ =\ &(K-1)\,\delta^{K-1}f(X_{\sup}(\mathcal{C}^*))+\frac{\delta^{K+1}}{1-\delta}f(X_{\sup}(\mathcal{C}^*)).\end{aligned}$$

When K is arbitrarily large this payoff is arbitrarily close to zero (since $\lim_{K\to\infty}(K-1)\,\delta^{K-1}=0$) and therefore lower than the payoff attained using the contract with two training periods described in the preceding section. This fact contradicts the hypothesis that setting $X^*_t<X_{\sup}(\mathcal{C}^*)$ for all t is optimal. ■

Intuition for lemma 2 is as follows. Consider a contract with an infinitely long training phase. Select an arbitrarily large period K and notice that the knowledge transferred after that period ($X_{\sup}(\mathcal{C}^*)-X^*_K$) is worth essentially zero, in present value, to either party (namely, $\frac{\delta^K}{1-\delta}f(X_{\sup}(\mathcal{C}^*))$ $-\ \frac{\delta^K}{1-\delta}f(X^*_K)$ is arbitrarily close to zero). Alternatively, the expert could transfer that knowledge to the novice early on, for free, in a way that relaxes the novice's liquidity constraints (such knowledge transfer must be properly timed so as to not violate the novice's incentive constraints). Then, by lemma 1, the expert could use the resulting slack in the liquidity constraints to increase her payoff during a period that is less heavily discounted.

Lemma 3 *Let $\mathcal{C}^*$ be an optimal contract. During the last period of the training phase (i.e. the smallest period T such that $X^*_T=X_{\sup}(\mathcal{C}^*)$) the novice transfers the full balance of his bank account to the expert (i.e. $m^*_T=\frac{1}{\delta}B_{T-1}(\mathcal{C}^*)$).*

Proof. See appendix B. ■

Intuition for lemma 3 is as follows. Consider a contract with T training periods in which the expert does not extract the full balance of the novice's bank during the transaction stage of period T (the last active transaction stage in the contract). This contract can only be optimal if an increase in the T period transfer violates the novice's T period incentive constraint. It follows that this must be a contract in which the knowledge transferred during period T ($X_{\sup}(\mathcal{C}^*)-X^*_{T-1}$) is small relative to the novice's T period cash balances. In other words, it is a contract in which substantial knowledge was transferred early on and, as a result, the novice enjoys rents in period T above and beyond the present discounted value of $f(X_{\sup}(\mathcal{C}^*))$. The expert, however, could replicate such substantial early knowledge transfer by merely shortening the length of the contract's training phase and, as a result, eliminating the need for the novice to enjoy those additional rents.

Theorem 1 *Any optimal contract $\mathcal{C}^*$ has the following features:*

1. *It is an apprenticeship with a finite training phase (i.e. $X_t^* < X_T^* = X_{\sup}(\mathcal{C}^*)$ implies $m_t^* = \frac{1}{\delta}B_{t-1}(\mathcal{C}^*)$ and $m_{t+1}^* = \frac{1}{\delta}B_t(\mathcal{C}^*)$).*
2. *Consumption is zero during every period before the end of the training phase (i.e. $X_t^* < X_{\sup}(\mathcal{C}^*)$ implies $c_t^* = 0$).*
3. *During every period of the training phase, the novice's incentive constraint holds with equality.*
4. *All available knowledge is transferred to the novice (i.e. $X_{\sup}(\mathcal{C}^*) = 1$).*

Proof. Part 1 follows from combining lemmas 1-3.

For part 2, suppose toward a contradiction that there is a period t such that $X_t^* < X_{\sup}(\mathcal{C}^*)$ and yet $c_t^* > 0$. Now consider an alternative contract $\mathcal{C}'$ that is identical to $\mathcal{C}^*$ except for the following: (i) $c_t' = 0$; (ii) $c_{t+1}' = c_{t+1}^* + \frac{1}{\delta}c_t^*$. As a result, $B_t(\mathcal{C}') = B_t(\mathcal{C}^*) + c_t^*$, $V_{t+1}(\mathcal{C}') = V_{t+1}(\mathcal{C}^*) + \frac{1}{\delta}c_t^*$, and $B_k(\mathcal{C}') = B_k(\mathcal{C}^*)$ and $V_{k+1}(\mathcal{C}') = V_{k+1}(\mathcal{C}^*)$ for all $k \neq t$. It follows that $\mathcal{C}'$ satisfies all constraints and, since this new contract delivers the same payoff as $\mathcal{C}^*$ to the expert, $\mathcal{C}'$ must be optimal. However, we have

$$\frac{1}{\delta}B_t(\mathcal{C}') - m_{t+1}' = \frac{1}{\delta}B_t(\mathcal{C}^*) + \frac{1}{\delta}c_t^* - m_{t+1}^* > \frac{1}{\delta}B_t(\mathcal{C}^*) - m_{t+1}^* \geq 0,$$

and therefore $m_{t+1}' < \frac{1}{\delta}B_t(\mathcal{C}')$, which from lemmas 1 and 3 implies that $\mathcal{C}'$ is not optimal, a contradiction.

For parts 3 and 4, we can write the expert's payoff as

$$\sum_{t=2}^{T} \delta^{t-1} m_t^* = \sum_{t=2}^{T} \delta^{t-2} f(X_{t-1}^*),$$

where T is the smallest period such that $X_T = X_{\sup}(\mathcal{C}^*)$ and $m_t^* = \frac{1}{\delta}B_{t-1} = \frac{1}{\delta}f(X_{t-1}^*)$ for all $t \in \{2, ..., T\}$ owing to lemmas 1 and 3 and part 2 of the theorem. Moreover, $\mathcal{C}^*$ must satisfy the following incentive constraints for the novice:[6]

$$f(X_{t-1}^*) \leq \delta^{T-(t-1)} f(X_{\sup}(\mathcal{C}^*)) \text{ for all } t \leq T.$$

We first claim that all such constraints must hold with equality. Notice that the R.H.S. of these constraints is increasing in t. Now suppose, contrary to this claim, that at least one such constraint holds with strict inequality. Select the largest $t' \leq T$ such that the corresponding constraint holds

[6] To see why we can write the incentive constraint in this way recall that the original incentive constraint is

$$\frac{1}{1-\delta}f(X_{t-1}^*) + \frac{1}{\delta}B_{t-1}(\mathcal{C}^*) \leq V_t(\mathcal{C}^*).$$

Moreover, from parts 1 and 2 of the theorem, $\frac{1}{\delta}B_{t-1}(\mathcal{C}^*) = \frac{1}{\delta}f(X_{t-1}^*)$ and $V_t(\mathcal{C}^*) = \frac{\delta^{T-t}}{1-\delta}f(X_{\sup}(\mathcal{C}^*))$. The desired inequality follows from substituting these expressions in the original incentive constraint and rearranging terms.

with strict inequality, and so $f(X^*_{t'-1}) < f(X^*_{t'})$. Notice that the expert could have attained a higher payoff by increasing $X^*_{t'}$ by a small amount while holding every other X^*_t constant, a contradiction. This observation establishes part 3.

We may now further simplify the expert's payoff to

$$\sum_{t=2}^{T} \delta^{T-1} f(X_{\sup}(\mathcal{C}^*)).$$

If $X_{\sup}(\mathcal{C})$ was smaller that 1, the expert could have attained a higher payoff by increasing $X_{\sup}(\mathcal{C})$. This observation establishes part 4. ■

Intuition for theorem 1 is as follows. Part 1 merely combines lemma's 1-3. Part 2 is derived from the following observation. Suppose there is a period t in which (1) the novice has not yet received the totality of knowledge that the contract will eventually grant him ($X^*_t < X_{\sup}(\mathcal{C}^*)$) and (2) the novice enjoys \$1 of consumption. Then, the expert could alter the contract, without affecting either party's payoff, by merely requesting that the novice saves, rather than consumes, an additional \$1 during period t and then consumes an additional $\$\frac{1}{\delta}$ during period $t+1$. Since this modification relaxes the novice's period $t+1$ liquidity constraint, by lemmas 1 and 3, the expert could use the resulting slack to increase her payoff.

Part 3 is derived from the following observation. If the incentive constraint for period t is slack, the expert can increase her payoff by increasing X_{t-1} for free (which increases the novice's bank balance for period t) and then requesting a higher payoff from the novice in period t, at which point both the incentive and liquidity constraints are slack.

For part 4, consider an optimal contract $\mathcal{C}^*$ with T training periods. Since this contract prescribes zero consumption and zero cash accumulation during those T periods, the novice's equilibrium payoff from the standpoint of period T is $V_T(\mathcal{C}^*) = \frac{1}{1-\delta} f(X_{\sup}(\mathcal{C}^*)) = \frac{1}{1-\delta} f(X^*_T)$ and therefore his equilibrium payoff from the standpoint of period $t \leq T$ is

$$V_t(\mathcal{C}^*) = \frac{\delta^{T-t}}{1-\delta} f(X^*_T).$$

Note that if X^*_T was smaller than 1 the expert could have relaxed all incentive constraints, without affecting the liquidity constraints and without affecting the payoff of either party, by merely increasing X^*_T. As noted above, however, the expert could use this slack to increase her payoff.

Corollary 1 shows that there is a unique optimal contract and describes such contract in closed form:

Corollary 1 *Let $\mathcal{C}^*$ be an optimal contract. The duration of its training phase, denoted $T(\mathcal{C}^*)$, solves*

$$\max_{T} \ (T-1)\delta^{T-1} f(1).$$

As a result, there is a unique optimal contract with the following features (up to an integer constraint *for $T(\mathcal{C}^*)$):*

1. $T(\mathcal{C}^*) = 1 + |\ln \delta|^{-1}$
2. $f(X_t) = \frac{1}{e}\delta^{-(t-1)} f(1)$ *for all* $t \in \{1, ..., T(\mathcal{C}^*)\}$.
3. $m_t = \delta^{T-t} f(1)$ *for all* $t \in \{2, ..., T(\mathcal{C}^*)\}$ *and* $m_t = 0$ *for every* $t \notin \{2, ..., T(\mathcal{C}^*)\}$.

Proof. For the first part of the corollary, let $\mathcal{C}^*$ be an optimal contract with $T(\mathcal{C}^*)$ training periods. From theorem 1, part 4, we have $X_{\sup}(\mathcal{C}^*) = 1$ and from theorem 1, parts 1-3, the liquidity and incentive constraints deliver, for all $t \in \{2, ..., T(\mathcal{C}^*)\}$,

$$m_t^* = \frac{1}{\delta} B_{t-1}(\mathcal{C}^*) = \frac{1}{\delta} f\left(X_{t-1}^*\right), \text{ and} \tag{L_t}$$

$$\frac{1}{1-\delta} f(X_{t-1}^*) + \frac{1}{\delta} f\left(X_{t-1}^*\right) = V_t(\mathcal{C}^*) = \frac{\delta^{T(\mathcal{C}^*)-t}}{1-\delta} f(1). \tag{IC_t}$$

By combining the above expressions we obtain, for all $t \in \{2, ..., T(\mathcal{C}^*)\}$,

$$m_t^* = \delta^{T(\mathcal{C}^*)-t} f(1).$$

The expert's payoff is therefore

$$\sum_{t=2}^{T(\mathcal{C}^*)} \delta^{t-1} m_t^* = (T(\mathcal{C}^*) - 1)\, \delta^{T(\mathcal{C}^*)-1} f(1).$$

Notice that for $\mathcal{C}^*$ to be optimal, $T(\mathcal{C}^*)$ must maximize the above expression.

For point 1 of the corollary, note that $(T-1)\delta^{T-1} f(1)$ is a concave function of T that is uniquely maximized (up to an integer constraint for T) at $T = 1 + (-\ln \delta)^{-1}$. The remaining points 2 and 3 of the corollary follow from setting $T(\mathcal{C}^*) = 1 + (-\ln \delta)^{-1}$ in constraints (L_t) and (IC_t) above and noting that $\delta^{-\frac{1}{\ln \delta}} = \frac{1}{e}$. ■

Corollary 1 shows that, in the optimal contract, knowledge transfer takes a simple form. In period 1, the novice receives knowledge X_1^* such that $f(X_1^*) = e^{-1} f(1)$, regardless of δ. After that, the novice's total value of knowledge grows at rate $r = \frac{1}{\delta} - 1$ until the training phase is complete.[7]

Efficiency and the role of δ

[7] Remark 1 extends the above results to the case in which $f(0) > 0$:

Remark 1 *If* $f(0) \leq e^{-1} f(1)$ *all results are unaffected. If* $f(0) > e^{-1} f(1)$, *then the optimal contract* $\mathcal{C}^*$ *satisfies all properties in theorem 1. Moreover,*

$$\begin{aligned} X_1^* &= 0 \text{ and} \\ f(X_{t+1}^*) &= (1+r) f(X_t^*) \text{ for all } t < T(\mathcal{C}^*). \end{aligned}$$

Proof. Available upon request. ■

The expert's optimal contract $\mathcal{C}^*$ is socially inefficient as it artificially spreads the transfer of knowledge over multiple periods rather than transferring all knowledge up front. From corollary 1, the total waste in surplus, discounted to period 1, is

$$\sum_{t=1}^{|\ln \delta|^{-1}} \left[\delta^{t-1} - \frac{1}{e} \right] f(1).$$

This waste increases in δ for two reasons. First, a higher δ implies a longer training phase and, with it, a lower rate of knowledge transfer. Second, a higher δ implies that the inefficiencies caused by this lower rate loom larger from the perspective of period 1.

To provide intuition for why the optimal training phase increases with δ we compare a contract with two training periods against a contract with three training periods, both satisfying the properties in theorem 1 (i.e. apprenticeships with zero consumption during their respective training phases, binding incentive constraints, and full knowledge transfer).

In the contract with two training periods, the expert transfers knowledge such that $f(X_1) = \delta f(1)$ and $f(X_2) = f(1)$. In return, the expert receives money transfer $m_2 = f(1)$ in period 2 and her total payoff, discounted to period 1, is

$$\Pi_1 = \delta f(1).$$

In the contract with three training periods, the expert transfers knowledge such that $f(X_1) = \delta^2 f(1)$, $f(X_2) = \delta f(1)$, and $f(X_3) = f(1)$. In return, the expert receives money transfers $m_2 = \delta f(1)$ and $m_3 = f(1)$ and her total payoff, discounted to period 1, is

$$\Pi_1 = 2\delta^2 f(1).$$

(In other words, the trade-off facing the expert is that reducing the speed at which knowledge is transferred leads to lower transfers during the training phase, but also leads to a larger number of such transfer.)

As δ grows, the novice places a higher value on receiving incremental knowledge even if he receives this knowledge slightly farther in the future. Consequently, slowing down the speed of knowledge transfer has less of a negative impact on each money transfer the expert is capable of extracting during the training phase. It follows that using a longer training phase becomes relatively more attractive.

V Constrained-efficient contracts

In this section, we study surplus-maximizing contracts. So that the exercise is nontrivial, in addition to imposing the novice's liquidity and incentive constraints, we impose a new constraint requiring that the expert receives a payoff no lower than an exogenous value $\Pi_0 > 0$ in exchange for training the novice:

$$\Pi_1(\mathcal{C}) \geq \Pi_0.$$

We may interpret Π_0, for example, as a fixed cost incurred by the expert. We assume that Π_0 is low enough so that full knowledge transfer is in fact feasible.[8]

A constrained-efficient contract solves

$$\max_{\mathcal{C}} \ \sum\nolimits_t \delta^{t-1} f(X_t)$$
$$s.t.$$
$$L_t,\ IC_t \text{ for all } t, \text{ and}$$
$$\Pi_1(\mathcal{C}) \geq \Pi_0.$$

Proposition 1 shows that any constrained-efficient contract has several properties that mirror the expert's profit-maximizing contract:

Proposition 1 *Any constrained-efficient contract $\mathcal{C}$ has the following features:*

1. *It is an apprenticeship with a finite training phase.*
2. *Consumption is zero during every period before the end of the training phase.*
3. *During every period of the training phase, the novice's incentive constraint holds with equality.*
4. *All available knowledge is transferred to the novice.*

Proof. Available upon request. ■

The key difference relative to the expert's profit-maximizing contract is its duration:

Corollary 2 *Let $\mathcal{C}$ be a constrained-efficient contract. The duration of its training phase, denoted $T(\mathcal{C})$, solves*

$$\min_{T} \ (T-1)\delta^{T-1} f(1)$$
$$s.t.$$
$$(T-1)\delta^{T-1} f(1) \geq \Pi_0.$$

[8] Specifically, we assume that

$$\Pi_0 \leq \max_T (T-1)\delta^{T-1} f(1) = \frac{1}{e\,|\ln \delta|} f(1),$$

and so the expert's optimal contract delivers a payoff of at least Π_0.

Proof. Let $\mathcal{C}$ be an optimal contract with $T(\mathcal{C})$ training periods. From proposition 1, part 4, we have $X_{\sup}(\mathcal{C}) = 1$ and from proposition 1, parts 1-3, the liquidity and incentive constraints deliver

$$\begin{aligned} m_t &= \delta^{T(\mathcal{C})-t} f(1) \text{ for all } t \in \{2, ..., T(\mathcal{C})\}, \text{ and} \\ f(X_t) &= \delta^{T(\mathcal{C})-t} f(1) \text{ for all } t \in \{1, ..., T(\mathcal{C})\}. \end{aligned}$$

The expert's payoff and total surplus, discounted to period 1, are respectively

$$\Pi_1 = (T(\mathcal{C}) - 1)\, \delta^{T(\mathcal{C})-1} f(1), \text{ and}$$
$$\sum_{t=1}^{\infty} \delta^{t-1} f(X_t) = (T(\mathcal{C}) - 1)\, \delta^{T(\mathcal{C})-1} f(1) + \frac{1}{1-\delta} f(1)$$

The corollary follows from the observation that surplus decreases in $T(\mathcal{C})$ and the requirement that $\Pi_1 \geq \Pi_0$. ■

Corollaries 1 and 2 imply that the training phase of the expert's profit-maximizing contract is at least as long as the training phase of the constrained-efficient contract. Moreover, as δ grows, while the training phase of the expert's profit-maximizing contract grows, the training phase of the constrained-efficient contract shrinks.

VI Extensions

In this section we study three extensions of our baseline model (with each extension considered in isolation from the other two). First, we consider the case in which the novice is endowed, up front, with a positive level of cash. Second, we consider the case in which the expert directly benefits or suffers from the knowledge acquired by the novice, i.e. knowledge transfers produce an externality on the expert. Third, we consider the case in which knowledge has a degree of relationship specificity.

A common theme throughout these extensions is that, under the parametrizations we consider, profit-maximizing contracts give rise to apprenticeships of a finite duration with full knowledge transfer. As in the baseline model, during the training phase of these contracts the value of the novice's knowledge grows at the gross interest rate $(1 + r) = \frac{1}{\delta}$, regardless of other parameter values. The novice's initial cash balance, the size of the externalities, and the degree of knowledge specificity affect, exclusively, the magnitude of the initial knowledge transfer X_1. Moreover, in all three extensions, as players become more patient (δ grows) the profit-maximizing contract involves a (weakly) longer training phase and a (weakly) lower level of efficiency.

A Novice's liquidity

Proposition 1 extends our baseline results to the case in which the novice enters period 1 with a positive bank balance, namely, $\frac{1}{\delta} B_0 > 0$.

Proposition 2 *Suppose* $\frac{1}{\delta}B_0 > 0$.

1. *Case 1:* $\frac{1}{\delta}B_0 \leq \frac{e^{-1}}{1-\delta}f(1)$. *The optimal contract* $\mathcal{C}^*$ *satisfies all properties in theorem 1. Moreover,*

$$\begin{aligned} f(X_1^*) &= e^{-1}f(1), \text{ and} \\ f(X_{t+1}^*) &= (1+r)f(X_t^*) \text{ for all } t < T(\mathcal{C}^*). \end{aligned}$$

   In this case, N's initial balance does not accelerate the rate of knowledge transfer and N receives rents

$$\frac{e^{-1}}{1-\delta}f(1) - \frac{1}{\delta}B_0.$$

2. *Case 2:* $\frac{1}{\delta}B_0 \in \left(\frac{e^{-1}}{1-\delta}f(1), \frac{1}{1-\delta}f(1)\right)$. *The optimal contract* $\mathcal{C}^*$ *satisfies all properties in theorem 1. Moreover,*

$$\begin{aligned} f(X_1^*) &= \frac{1-\delta}{\delta}B_0 > e^{-1}f(1) \text{ and} \\ f(X_{t+1}^*) &= (1+r)f(X_t^*) \text{ for all } t < T(\mathcal{C}^*). \end{aligned}$$

   In this case, N's initial balance accelerates the speed of knowledge transfer exclusively through its positive impact on X_1^* *and N receives zero rents.*

3. *Case 3:* $\frac{1}{\delta}B_0 \geq \frac{1}{1-\delta}f(1)$. *The optimal contract* $\mathcal{C}^*$ *prescribes full knowledge transfer in period* 1 *in exchange for money transfer* $\frac{1}{1-\delta}f(1)$. *N receives zero rents.*

Proof. Available upon request. ■

Intuition is as follows. In case 1, the expert is able to use an ideal arrangement. Namely, she implements the optimal contract from the baseline model and, simultaneously, she extracts the novice's entire balance $\frac{1}{\delta}B_0$. This arrangement is possible because the optimal contract from the baseline model had left rents $\frac{e^{-1}}{1-\delta}f(1)$ in the hands of the novice.

In case 2, the optimal contract is determined by two forces. On the one hand, owing to the forces behind lemma 1, the expert wishes to extract the novice's entire balance in period $\frac{1}{\delta}B_0$, which the novice will only agree to if he receives sufficient knowledge in return. On the other hand, the expert wishes to keep $f(X_1)$ as close as possible to $e^{-1}f(1)$ in order to slow down the overall speed at which knowledge is transferred. The solution is setting $f(X_1)$ to its minimum level such that the novice surrenders $\frac{1}{\delta}B_0$.

In case 3, the expert is able to implement the first-best allocation ($X_1 = 1$) while keeping all surplus ($\frac{e^{-1}}{1-\delta}f(1)$) for herself.

B Externalities

This section considers an extension of the baseline model in which the expert experiences an externality as the novice acquires knowledge. In particular, we assume that during the production/consumption phase of every period t, the expert produces output $\overline{f}(X_t)$ for herself, in addition to the novice producing output $f(X_t)$ for himself. We assume that $\overline{f}$ is a strictly monotonic function. Let $S(X_t) = f(X_t) + \overline{f}(X_t)$ denote the total output produced in period t.

Definition 3 *The novice's level of knowledge X_t causes a positive (resp. negative) externality on the expert if $\overline{f}(X_t)$ is increasing (resp. decreasing) in X_t.*

For a simple example of a positive externality consider a senior partner in a law firm (the expert) who benefits when a junior partner (the novice) acquires knowledge and, as a result, enhances the reputation of the firm in question. For a simple example of a negative externality consider a technology firm (the expert) that loses profits when a competing firm (the novice) acquires knowledge and, as a result, becomes a stronger competitor. Note that when $\overline{f}(X_t)$ is independent of X_t the extended model is equivalent to the baseline model.

Assumption 1 $S(X_t)$ is increasing in X_t.

The expert's relaxed problem is

$$\max_{\mathcal{C}} \ \sum_{t=1}^{\infty} \delta^{t-1} \left[m_t + \overline{f}(X_t)\right] \qquad (II)$$

$$s.t$$

$$(L_t) \text{ and } (IC_t) \text{ for all } t.$$

Proposition 2 presents a partial characterization of optimal contracts under externalities:

Proposition 3 *Consider the model with externalities. Under assumption 1, any optimal contract $\mathcal{C}^*$ has the following features:*

1. *During every period before the end of the training phase the novice transfers the full balance of his bank account to the expert.*

2. *The training phase is finite.*

3. *During every period of the training phase, the novice's incentive constraint holds with equality.*

Proof. For part 1, we note that the proof of lemma 1 remains valid in the model with externalities, except for the following difference. Switching from contract $\mathcal{C}^*$ to contract $\mathcal{C}'$ increases the expert's payoff by

$$\delta^{t-1}\left(m_t' - m_t^* + \overline{f}(X_t') - \overline{f}(X_t^*)\right) + \delta^t\left(m_{t+1}' - m_{t+1}^*\right) =$$
$$\delta^{t-1}(1-\delta)\varepsilon + \delta^{t-1}\left(\overline{f}(X_t') - \overline{f}(X_t^*)\right) =$$
$$\delta^{t-1}\left(f(X_t') - f(X_t^*) + \overline{f}(X_t') - \overline{f}(X_t^*)\right) > 0,$$

where the inequality follows from the fact that $S(X_t)$ is increasing in X_t.[9]

For part 2, we note that the argument in the proof of lemma 2 remains valid, except for the following difference. The expert's payoff is

$$\sum\nolimits_{t=1}^{\infty}\delta^{t-1}\overline{f}(X_t^*) + \sum\nolimits_{t=2}^{K}\delta^{t-2}f(X_{t-1}^*) + \sum\nolimits_{t=K+1}^{\infty}\delta^{t-2}\left[f(X_{t-1}^*) - c_{t-1}^*\right],$$

which is bounded above by

$$\sum\nolimits_{t=1}^{K}\delta^{t-1}\overline{f}(X_t^*) + (K-1)\,\delta^{K-1}f(X_{\sup}(\mathcal{C}^*)) + \frac{\delta^{K+1}}{1-\delta}f(X_{\sup}(\mathcal{C}^*)).$$

Since the second and third terms converge to zero as K grows to infinity, and $\overline{f}$ is strictly monotonic, $\mathcal{C}^*$ can be improved upon using an alternative contract $\mathcal{C}'$ with zero money transfers and either $X_t' = 0$ for all t (if $\overline{f}$ is decreasing) or $X_t' = 1$ for all t (if $\overline{f}$ is increasing).

For part 3, we can write the expert's payoff as

$$\sum\nolimits_{t=2}^{T-1}\delta^{t-2}S(X_{t-1}^*) + \delta^{T-1}\left[\frac{1}{\delta}S(X_{T-1}^*) + \frac{1}{1-\delta}S(X_{\sup}(\mathcal{C}^*)) - V_T(\mathcal{C}^*)\right].$$

where T is the duration of the training phase and the values of X_t^* and c_t^* satisfy the novice's incentive constraints

$$\frac{1}{1-\delta}f(X_{t-1}^*) + \frac{1}{\delta}f(X_{t-1}^*) = \frac{1}{\delta(1-\delta)}f(X_{t-1}^*) \leq \delta^{T-t}V_T(\mathcal{C}^*) \text{ for all } t \leq T.$$

Now suppose, contrary to this claim, that at least one such constraint holds with strict inequality. Select the largest $t' < T$ such that the corresponding constraint holds with strict inequality, and so $f(X_{t'}^*) < f(X_{t'+1}^*)$. Notice that, under assumption 1, the expert could have attained a higher payoff by increasing $X_{t'}^*$ by a small amount while holding every other X_t^* constant, a contradiction. ■

Corollary 3 presents a more detailed characterization of the optimal contract under the added assumption that $\overline{f}$ and f are linearly related:

Corollary 3 *Suppose $\overline{f}(X) \equiv \gamma f(X)$ for some constant γ.*[10] *Under assumption 1, an optimal*

[9] The argument in appendix A also remains valid because remark 1 is unaffected.
[10] Adding a constant to $\overline{f}$ would be immaterial.

contract $\mathcal{C}^*$ delivers payoff

$$f(X_{\sup}(\mathcal{C}^*)) \cdot A(\delta, \gamma),$$

where

$$A(\delta, \gamma) = \max_T \ \delta^{T-1} \left[(T-1)(1+\gamma) + \frac{\gamma}{1-\delta} \right] > 0.$$

As a result, $\mathcal{C}^*$ has the following features:

1. *The cumulative knowledge transfer is complete (i.e. $X_{\sup}(\mathcal{C}^*) = 1$).*

2. *The initial knowledge transfer X_1^* is such that*

$$f(X_1^*) = \frac{1}{e} \cdot \delta^{-\frac{\gamma}{(1+\gamma)(1-\delta)}},$$

   which is increasing in γ, increasing in δ when $\gamma > 0$, and decreasing in δ when $\gamma < 0$.

3. *During the training phase, the novice's output $f(X_t^*)$ grows at rate $r = 1 - \frac{1}{\delta}$.*

Proof. Available upon request. ■

Corollary 3 shows that, in the optimal contract, knowledge transfer takes a simple form. In period 1, the novice is gifted knowledge $X_1^*(\delta, \gamma)$ such that: (i) $f(X_1^*(\delta, 0)) = e^{-1} f(1)$ for all δ (as the special case in which $\gamma = 0$ corresponds to our baseline model); (ii) $\frac{\partial}{\partial \gamma} X_1^*(\delta, \gamma) > 0$ for all δ, γ (namely, the larger the externality, the larger the initial knowledge transfer); and (iii) $\frac{\partial^2}{\partial \gamma \partial \delta} X_1^*(\delta, \gamma) > 0$ for all δ, γ (namely, δ magnifies the effect of γ on X_1^*). After period 1, the novice's value of knowledge $f(X_t^*)$ grows at rate $r = \frac{1}{\delta} - 1$, regardless of γ, until the training phase is complete.

Figure 1 (at the end of this document) depicts the duration of the training phase as a function of (δ, γ). Lighter areas correspond to longer durations. In one extreme, when γ is sufficiently high and δ is sufficiently low, training occurs in a single period. In the other extreme, when either γ approaches -1 or δ approaches 1 (or both), duration approaches infinity.

C Specific knowledge

So far we have focused on cases in which the expert transfers general-purpose knowledge to the novice. In this section we consider the case in which knowledge is complementary to an asset, such as a trademark or a factory, owned by the expert. That is, knowledge is (partially) specific to the relationship.

Formally, the novice's period t output is $f(X_t, A_t)$, where $A_t \in \{0, 1\}$ in as input controlled by the expert and $f(X_t, 1) > f(X_t, 0)$. The expert selects A_t during the production/consumption stage of period t. For every t, a contract prescribes, on the path of play, values $\langle X_t, m_t, c_t, A_t \rangle$ and prescribes, following any past deviation, an interruption of all transactions and $A_t = 0$.

Given contract $\mathcal{C} = \langle X_t, m_t, c_t, A_t \rangle_{t=1}^{\infty}$, the novice's incentive constraints are now as follows. Before the transaction stage of period t, the novice must prefer to carry on with the prescribed actions over walking away with his current stock of knowledge, which is now worth only $f(X_{t-1}, 0)$ per period, together with his current bank account:[11]

$$\frac{1}{1-\delta} f(X_{t-1}, 0) + \frac{1}{\delta} B_{t-1}(\mathcal{C}) \leq V_t(\mathcal{C}). \qquad (IC_t')$$

The expert's problem is

$$\max_{\mathcal{C}} \sum_{t=1}^{\infty} \delta^{t-1} m_t \qquad (II)$$

$$s.t$$

$$(L_t) \text{ and } (IC_t') \text{ for all } t.$$

In what follows, we assume that $f(X_t, 0) \equiv \lambda f(X_t, 1)$ for some $\lambda \in (0, 1)$. Thus, a smaller λ implies that knowledge is more specific to the novice/expert relationship.

Proposition 4 *In the model with specific knowledge, the optimal contract $\mathcal{C}^*$ satisfies all properties in theorem 1. In addition, it has the following properties, which extend corollary 1:*

1. *Up to an integer constraint, the duration of the training phase is*

$$T(\mathcal{C}^*) = |\ln \delta|^{-1} + 1 - \frac{\delta}{(1-\delta)}(1-\lambda),$$

which is increasing in both δ and λ.

2. *During the training phase, the value of knowledge grows at rate r.*

3. *After training is complete, the novice pays the expert a per-period rental fee*

$$(1-\lambda) f(1, 1).$$

Sketch of proof. Consider a contract $\mathcal{C}^*$ that prescribes, on the path of play, an apprenticeship with full knowledge transfer and $A_t = 1$ for every t.

From the standpoint of period $T(\mathcal{C}^*) + 1$ (where $T(\mathcal{C}^*)$ is the first period in which $X_t = 1$), the equilibrium continuation payoff for the novice is:

$$V^* = \frac{1}{1-\delta} f(1, 0) + \frac{1}{\delta} f(1, 1) = \left[\frac{\lambda}{1-\delta} + \frac{1}{\delta}\right] f(1, 1).$$

[11] A simple observation is in order. Given the general-purpose nature of knowledge, it is immaterial for the model whether the residual claimant of output is the novice (as we have assumed) or the expert. In the former case, the novice's key deviation consists in walking away from the relationship during the transaction phase of a given period t before making any money transfer to the principal. In the latter case, the novice's key deviation is identical except for the fact that it is initiated during the production/consumption phase of period $t-1$, with the novice producing output $f(X_{t-1})$ on his own rather than for the expert. In present value, both deviations lead to the same payoff.

Therefore, the novice's incentive constraints during the training phase are

$$\left[\frac{\lambda}{1-\delta}+\frac{1}{\delta}\right] f(X_{t-1}^*,1) \leq \delta^{T-(t-1)} V^* \text{ for all } t \in \{1,...,T(\mathcal{C}^*)\},$$

which simplify to

$$f(X_{t-1},1) \leq \delta^{T-(t-1)} f(1,1) \text{ for all } t \in \{1,...,T(\mathcal{C}^*)\}.$$

By meeting these constraints with equality, the expert obtains the following money transfers during the training phase

$$m_t^* = \frac{1}{\delta} B_{t-1} = \frac{1}{\delta} f(X_{t-1},1) = \delta^{T-t} f(1,1) \text{ for all } t \in \{2,...,T(\mathcal{C}^*)\}. \tag{a}$$

After training is complete ($t \geq T(\mathcal{C}^*)+1$), the maximum per period transfer that the expert can extract from the novice, denoted M^*, satisfies

$$f(1,1)+\frac{\delta}{1-\delta}\lambda f(1,1) = f(1,1)+\frac{\delta}{1-\delta}\left[f(1,1)-M^*\right],$$

and therefore $M^* = (1-\lambda)f(1,1)$. The expert can then set

$$m_t^* = M^* = (1-\lambda)f(1,1) \text{ for all } t \geq T(\mathcal{C}^*)+1. \tag{b}$$

By combining (a) and (b) the expert's payoff becomes

$$\Pi_1 = \sum\nolimits_{t\geq 2} \delta^{t-1} m_t^* = \left[\delta^{T(\mathcal{C}^*)-1}(T(\mathcal{C}^*)-1)+\frac{\delta^{T(\mathcal{C}^*)}}{1-\delta}(1-\lambda)\right] f(1,1).$$

Finally, the first-order condition for $T(\mathcal{C}^*)$ delivers:

$$T(\mathcal{C}^*) = 1+\frac{1}{|\ln\delta|}-\frac{\delta}{1-\delta}(1-\lambda).$$

■

Figure 2 (at the end of this document) depicts the duration of the training phase as a function of (δ,λ). Lighter areas correspond to longer durations. The training phase expands with λ and δ, and, for any given $\lambda > 0$, the training phase expands to infinity as δ converges to 1.

D Remarks on competition: a continuum of experts and novices

In this section we investigate a form of competition across experts. We assume that there is a population with multiple experts and novices, with an equal mass of each. Experts compete for novice's, but are restricted to training at most one novice each.

Formally, consider a continuum of experts, indexed E, and a continuum of novices, indexed N, each distributed uniformly on $[0,1]$. All experts are identical to the expert in the baseline model. All novices are identical to the novice in the baseline model except for their initial cash balances. In particular, we assume that novice N is endowed with balance $B(N)$ at the beginning of period 1, where $B(0) = 0$ and $B'(N) \geq 0$ for all N.

Experts and novices are matched in pairs. Each matched pair agrees on a bilateral relational contract $\mathcal{C}$, as in the baseline model, that governs the corresponding relationship. Let $V(N, \mathcal{C})$ and $\Pi(E, \mathcal{C})$ denote the payoffs obtained from the viewpoint of period 1 by novice N and expert E, respectively, upon agreeing to contract $\mathcal{C}$.[12] Without loss, we assume that novice N is matched with expert $E = N$ and we denote their contract $\mathcal{C}(N) = (X_t(N), m_t(N), c_t(N))_{t=1}^{\infty}$.

We are interested in describing families of contracts that are immune to pairwise deviations (in the spirit of Gale and Shapley, 1962):[13]

Definition 4 *A family of contracts $\mathcal{C}(N)_{N\in[0,1]}$ is* ***stable*** *if (a) each contract is feasible and (b) there does not exist a pair (N, E) and a feasible contract $\mathcal{C}'$ for that pair such that*

$$V(N, \mathcal{C}') > V(N, \mathcal{C}(N)) \text{ and } \Pi(E, \mathcal{C}') > \Pi(E, \mathcal{C}(E)).$$

Remark 2 *In a stable family of contracts $\mathcal{C}(N)_{N\in[0,1]}$ all experts obtain identical payoffs:*

$$\Pi(\mathcal{C}(E), E) = \Pi(\mathcal{C}(E'), E') \text{ for all } E, E' \in [0,1].$$

Proof. Suppose contrary to the remark that $\Pi(\mathcal{C}(E), E) > \Pi(\mathcal{C}(E'), E')$ for some E, E'. Then the pair composed of the novice $N = E$ and expert E' would strictly prefer matching with each other and agreeing to a contract $\mathcal{C}'$ that is identical to $\mathcal{C}(E)$ except for prescribing an arbitrarily smaller level of m_1. ■

In general, there are multiple stable families of contracts, differing in both their efficiency levels and their allocation of surplus.

Definition 5 *A stable family of contracts $\mathcal{C}(N)_{N\in[0,1]}$ is* ***expert-preferred (novice-preferred)*** *if there is no alternative stable family of contracts $\mathcal{C}'(N)_{N\in[0,1]}$ yielding higher payoffs for all experts (novices).*

The simplest example of a stable family of contracts is the unique novice-preferred family. In this family, all experts transfer 100% of their knowledge in the first period to their corresponding novices in exchange for a zero payoff.

[12] We define the novices' payoffs net of their initial cash balances, and so these payoffs reflect incremental payoffs from agreeing to a contract.

[13] Our focus is on pairwise deviations taking place at the beginning of period 1. We assume that after period 1 no new matches are formed (e.g. owing to switching costs)..

The following lemma provides a first step toward characterizing expert-preferred families of contracts by restricting to the simple class of families with full knowledge transfer in at most two training periods:

Lemma 4 *Suppose experts are restricted to offering contracts with full knowledge transfer in at most two training periods. Let $\mathcal{C}(N)_{N\in[0,1]}$ be an expert-preferred, stable family of contracts in that class. We then have:*

1. For all E,
$$\Pi(\mathcal{C}(E), E) = \delta f(1).$$

2. For all N such that $B(N) \leq \delta f(1)$,
$$f(X_1(N)) = \delta f(1) + rB(N) \text{ and } V(N, \mathcal{C}(N)) = \frac{\delta}{1-\delta} f(1) + rB(N).$$

3. For all N such that $B(N) > \delta f(1)$,
$$f(X_1(N)) = f(1) \text{ and } V(N, \mathcal{C}(N)) = \frac{1}{1-\delta} f(1) - \delta f(1).$$

Proof. Let $\mathcal{C}(N)_{N\in[0,1]}$ be an expert-preferred, stable family of contracts in the above class.

Step 1. We argue that, for any given N, $f(X_1(N)) < f(1)$ implies that $m_1(N) = B(N)$ and $m_2(N) = \frac{1}{1-\delta}\left[f(1) - f(X_1(N))\right]$. That $m_1(N) = B(N)$ follows from the proof of lemma 1 and that $m_2(N) \leq \frac{1}{1-\delta}\left[f(1) - f(X_1(N))\right]$ follows from N's period 2 incentive constraint. Now suppose toward a contradiction that $m_2(N) < \frac{1}{1-\delta}\left[f(1) - f(X_1(N))\right]$. Expert $E = N$ and novice N could then simultaneously increase their payoffs by agreeing on an alternative contract $\mathcal{C}'$ that is identical to their original contract $\mathcal{C}(N)$ except for raising both $f(X_1(N))$ and $m_2(N)$ by a small $\varepsilon > 0$.

Step 2. We argue that, for any given N and N' such that both $f(X_1(N))$ and $f(X_1(N'))$ are smaller than $f(1)$ we have
$$f(X_1(N)) - f(X_1(N')) = \frac{1-\delta}{\delta}\left[B(N) - B(N')\right].$$

By combining step 1 with the above remark (i.e. $\Pi(N, \mathcal{C}(N)) = \Pi(N', \mathcal{C}(N'))$) we obtain
$$B(N) + \frac{\delta}{1-\delta}\left[f(1) - f(X_1(N))\right] = \Pi(N, \mathcal{C}(N))$$
$$= \Pi(N', \mathcal{C}(N')) = B(N') + \frac{\delta}{1-\delta}\left[f(1) - f(X_1(N'))\right],$$

which, upon rearranging terms, delivers the desired result.

Step 3. We argue that, for any given N, $f(X_1(N)) = f(1)$ implies that $m_1(N) = \frac{\delta}{1-\delta}[f(1) - f(X_1(0))]$ and $m_2(N) = 0$. That $m_2(N) = 0$ follows from N's period 2 incentive constraint. That

$m_1(N) = \frac{\delta}{1-\delta}\left[f(1) - f(X_1(0))\right]$ follows from noting that

$$m_1(N) = \Pi(N, \mathcal{C}(N)) = \Pi(0, \mathcal{C}(0)) = \frac{\delta}{1-\delta}\left[f(1) - f(X_1(0))\right]$$

and rearranging terms.

For part 1 of the lemma note that the payoff of any given expert E,

$$\Pi(E, \mathcal{C}(E)) = \frac{\delta}{1-\delta}\left[f(1) - f(X_1(0))\right],$$

is decreasing in $f(X_1(0))$ and therefore maximized when $f(X_1(0))$ takes its lowest possible value consistent with the period 2 liquidity constraint of novice $N = 0$, $m_2(0) \leq \frac{1}{\delta} f(X_1(0))$. From step 1 (i.e. $m_2(0) = \frac{1}{1-\delta}\left[f(1) - f(X_1(0))\right]$), such lowest possible value for $f(X_1(0))$ is $\delta f(1)$. It follows that $\Pi(E, \mathcal{C}(E)) = \delta f(1)$ for all E.

For part 2 of the lemma, $f(X_1(N)) = \delta f(1) + rB(N)$ follows from step 2 after setting $N' = 0$ and $f(X_1(0)) = \delta f(1)$, and $V(N, \mathcal{C}(N)) = \frac{\delta}{1-\delta} f(1) + rB(N)$ follows from the fact that

$$\begin{aligned} V(N, \mathcal{C}(N)) &= f(X_1(N)) + \frac{\delta}{1-\delta} f(1) - \Pi(N, \mathcal{C}(N)) \\ &= \frac{\delta}{1-\delta} f(1) + rB(N). \end{aligned}$$

For part 3 of the lemma, $f(X_1(N)) = f(1)$ is implied by the fact that setting $f(X_1(N)) < f(1)$ would contradict part 2 of the lemma, and $V(N, \mathcal{C}(N)) = \frac{1}{1-\delta} f(1) - \delta f(1)$ follows from the fact that

$$\begin{aligned} V(N, \mathcal{C}(N)) &= \frac{1}{1-\delta} f(1) - \Pi(N) \\ &= \frac{1}{1-\delta} f(1) - \delta f(1). \end{aligned}$$

■

For comparison, consider a scenario in which experts do not compete against each other (namely, pairwise deviations are ruled out), while maintaining the assumption of full knowledge transfer in at most two training periods. In that scenario, as a corollary of proposition 2, we obtain:

1. For all N such that $B(N) \leq \frac{\delta}{1-\delta} f(1)$,

$$f(X_1(N)) = \delta f(1) \text{ and } V(N, \mathcal{C}(N)) = \frac{\delta}{1-\delta} f(1) - B(N)$$

2. For all N such that $B(N) \in \left(\frac{\delta}{1-\delta} f(1), \frac{1}{1-\delta} f(1)\right)$,

$$f(X_1(N)) = (1-\delta)B(N) \text{ and } V(N, \mathcal{C}(N)) = 0.$$

3. For all N such that $B(N) \geq \frac{1}{1-\delta} f(1)$,

$$f(X_1(N)) = f(1) \text{ and } V(N, \mathcal{C}(N)) = 0.$$

It follows that competition across experts (in expert-preferred form) has two effects:

First, competition accelerates the speed of knowledge transfer for every novice with an initial cash balance in the interval $\left(0, \frac{1}{1-\delta} f(1)\right)$. Indeed, in the competitive scenario, no matter how small a novice's initial cash balance is, every dollar of additional cash is translated into r dollars of additional output in period 1, whereas in the monopolistic scenario every dollar of additional cash is translated into $\frac{r}{1+r}$ dollars of additional output in period 1 *only* after the novice's initial cash balance exceeds the sizeable level $\frac{1}{r} f(1)$.

Second, competition increases the equilibrium payoff for every novice with a positive cash balance. Indeed, in the competitive scenario, a higher initial cash balance is translated into a (weakly) higher net equilibrium payoff, whereas in the monopolistic scenario a higher initial cash balance is translated into a (weakly) lower net equilibrium payoff.

VII Applications and discussion

A Slow training and rent extraction in professional service firms

Professional service firms provide a wide range of general skills to junior consultants (our apprentices), usually called "associates" (see e.g. Mester, 19993; Richer at al., 2007). Partly, this training may be paid through lower wages, but there are reasons to believe that the training is being slowed down while the consultants "pay their dues", that is, rents are being extracted from their work, in exchange of the promise of more training and promotion.[14] Specifically:

- Management consulting firms train junior consultants as generalists, and do not allow them to specialize in a particular type of assignment or client.[15] While this is usually advertised as leading to the development of "generalist consultants", partners are usually specialists. Inside the firms, it is often discussed that the knowledge should belong to the firm, and it is not wanted that an associate becomes 'good enough' at a particular type of problem or a particular client that he can set up shop on his own, like in our setting.

- In law firms, the training of associates also seems, anecdotally, to proceed at a glacial pace while associates are "paying their dues". Numerous blog posts and articles are dedicated

[14] For an alternative explanation of the partnership, as a way to commit to high quality service in a context of imperfect observability, see Levin and Tadelis (2005).

[15] For example, McKinsey, in its web site, says "Our goal is to give you team assignments that will help you develop new talents and expose you to a diverse portfolio of clients, industries, and challenges." (http://www.mckinsey.com/careers/how_youll_grow/team_based_growth accessed 29.3.2013).

> to describing this feature. For example, a former litigator wrote recently in the New York Times, "this recession may be the thing that delivers them from more 3,000-hour years of such drudgery as changing the dates on securitization documents and shuffling them from one side of the desk to the other"... "it often takes a forced exit to break the leash of inertia that collars so many smart law graduates to mind-numbing work."[16]

As our analysis above suggests, with a large prize at the end, the expert can entice the novice to long training periods with little knowledge added in each period, while extracting the production generated in the process.

B Does the market deliver the optimal duration of apprenticeships?

There is a lot of policy interest in encouraging firms to offer apprenticeships. The G20 ministers meeting in Guadalajara, Mexico declared themselves committed to "promote, and when necessary, strengthen quality apprenticeship systems that ensure high level of instruction and adequate remuneration and avoid taking advantage of lower salaries" (OECD, 2012). The OECD (2012:5) argues that "Quality apprenticeships require good governance to prevent misuse as a form of cheap labour." Is this concern legitimate?

In a word, yes. Our analysis suggests that the expert, in order to extract rents from the apprentice, will artificially slow down the rate of knowledge transfer. Thus the market will not deliver apprenticeships of the optimal duration by making them too long. This is in contrast to the previous literature, which, consistently with the Beckerian insights on general human capital, focuses on the distortions that occur when the trainees quit 'too early' (or threaten to do so), thus limiting the gain for the firm doing the training, reducing their returns, and thus reducing their incentives to undertake training to begin with. For example Malcomson et al. (2003) argue that the optimal regulation should "increase apprenticeship length, coupled with a subsidy for each completed apprenticeship."

In fact, our analysis shows that, if the expert knows the duration is more limited, he has an incentive on his own to train faster, which is efficiency enhancing. In fact, absent technological knowledge transmission constraints, the rate of knowledge transfer increases in a first-order-stochastic-dominance sense as T increases.

C International Joint Ventures as Relational knowledge transfers

International joint ventures between a first world "North" partner and a developing country or "South" one often involve a transfer of knowledge from the North partner in exchange for a cash flow from the South one. Regardless of the actual legal structure of the contract, the contract is in place as long as the parties consider it in their common interest to continue the "knowledge for

[16] "Another View: In Praise of Law Firm Layoffs" July 1, 2009, The New York Times, Dealbook.

cash" swap, as the legal system in the developing country is often not likely to pay much attention to the legal niceties of the contract.

A notable example of this "relational rather than legal" joint venture agreement, which eventually grew to involve the Presidents of both France and China, was the dispute between Danone and Wahaha that started in 2007 and was settled in 2009.

The relationship started in 1996 when Danone, a French drinks and yogurt producer, established a joint venture with the Hangzhou Wahaha group, a Chinese maker of milk drinks for children. For Danone, the venture was a way to profit from the growing Chinese market, while the purpose for Wahaha was to learn Danone's technology. The joint venture was initially very successful, eventually contributing 5%-6% of the entire Danone's group operating profits (FT, Lex blog, April 12, 2007).

However, by 2007 Danone's contribution to the joint venture had run its course, as Wahaha could produce all the drinks on its own. In a public and acrimonious war of words, Danone accused Wahaha of having set up a parallel organization to sell yoghurt and drinks to its clients outside of the joint venture, making profits as large as those of the JV itself (FT, April 12, 2007). Mr. Zong, the Chairman of Wahaha, accused Danone of "trying to take control of parts of the Chinese group not included in the joint venture."

Danone had legally the stronger hand to play, as it owned 51% $\delta$of the joint venture. However, this power was not quite real, since as the press accounts at the time recognized, "the joint venture depends on Mr. Zong's continuing cooperation. Not only is he chairman and general manager of the joint venture, but he is the driving force behind the entire Wahaha organization. Furthermore, in China, employees in private enterprises often feel a stronger loyalty to the boss than the organization itself. Winning in the courts or pushing out Mr. Zong, therefore, are not solutions to Danone's problems." (FT,12 April 2007). Workers were also on Zong's side: "We formally warn Danone and the traitors they hire, we will punish your sins. We only want Chairman Zong. Please get out of Wahaha!" (WSJ 12 June 2007).

Indeed, Danone lost all its court battles in China. In December 2007, Wahaha won a trademark arbitration in China, when the Hangzhou Arbitration Commission, based on Wahaha hometown, accepted Wahaha's request that the trademark transfer agreement be terminated. Simultaneously, a trade union representing the companies workers (FT December 17, 2007) said "it had won an injunction freezing assets of the joint venture in Shandong province as a result of a lawsuit accusing Danone of bad faith." Also on December 2007, China's state news agency Xinhua reported that a court in far-western Xinjiang region had dismissed a lawsuit against Wahaha by a Danone subsidiary." Finally, on August 2008 (see FT, 06 Aug 2008), Danone lost the appeal on the arbitration, when the Hagnzhou People's courts rejected a bid by Danone to overturn the Chinese arbitration commission's ruling in favour of Wahaha in their battle over ownership of the Wahaha trademark.

The dispute ended on October 2009 with a cash settlement after an Stockholm arbitration court ruling (FT 1 Oct 2009) terminating all legal proceedings and leaving all the IP with Wahaha in exchange for a $300m cash transfer. This settlement was considered by observer a "pyrrhic victory" for Danone (FT November 10, 2009). For the FT Lex column (1 oct 2009) there were two flaws in Danone's JV: first, the trademark structure (the JV did not owned the trademark). Second, to try to settle the dispute in arbitration in Stockholm. "Such arrangements, typical of foreign joint ventures in China, work fine for routine complaints, such as one party seeking damages from the other for a faulty product. But as a mechanism to settle complex arguments on obligations under Chinese intellectual property law - forget it."

This type of case is far from unique, and there is a lot of anecdotal evidence that time these disputes have been quite common. For example, Ingersoll-Rand (IR) claimed Zhengchang Liyang Machinery Company Ltd. (ZC) had breached the Joint Venture Agreement by manufacturing and selling imitation processing equipment based on IR's patents (2000 U.S. Dist. LEXIS 18449). Again the government of China concluded that IR had defrauded ZC by falsifying the value of its contribution to the joint venture. The disputes between IR and ZC grew so severe that ZC built a five-foot wall down the middle of the factory used by the joint venture to remedy the "short changing" that it felt it had been receiving from IR.

Our analysis illuminates multiple aspects of this type of cases. Like in our model, these are cases where there effectively are no legally enforceable contracts; no commitment on novice's (Wahaha's) side, and knowledge is transmitted over time. Danone completely misread the situation, relying on its legal rights rather than on what the parties could bring to the relationship. Knowing about the lack of enforceability, Danone should have taken actions to reduce from the start the outside option of apprentice: IPs/Trademarks.

The weak institutions in China transform the relationship between the two companies to one which is equivalent to the one between two human beings were human capital is being transmitted.

VIII Conclusion

An extensive literature has grappled with the paradox that firms often pay for general training. A key insight in that literature is that firms pay for that type of training because competition for human capital is far from perfect, allowing firms to extract sufficient rents. Here, we provide an alternative answer: the teacher trains the student in the expectation of future payment, while the student pays the teacher, using the output produced by means of his acquired knowledge, due to the expectation of subsequent training.

We show that the optimal contract takes the form of an apprenticeship in which the student accumulates no savings and uses all his output to pay for additional knowledge until training is complete. During training, paying the student with knowledge, rather than cash, is preferred by both the student and teacher, as it increases total surplus. Moreover, training lasts for a finite number of periods. The reason is that during every training period the teacher extracts all the output produced one period before, and since output increases as knowledge increases, knowledge must be transferred at an increasing rate.

As we show, the teacher inefficiently delays the knowledge transfer. Intuitively, since future knowledge transfers are the only way to reward the student, the teacher must keep enough knowledge in reserve so that the student has the incentive to return for further training. Moreover, the more patient the players, the lower the transfer rate, the longer the training phase, and the larger the associated deadweight losses. Indeed, when players are more patient, the expert can keep the student around with a smaller immediate transfer–future transfers are viewed as more valuable–which in turn allows the expert to extract larger future transfers–also viewed as more valuable.

While the bulk of our analysis discusses a novice who wishes to acquire general knowledge and has no cash to pay for it, we show that our core findings are robust. Specifically we show that the fundamental contract form and the associated distortions are qualitatively unchanged if the novice arrives with some cash up front, or if the knowledge is partially specific, so that when the novice quits, his knowledge is less valuable than when used in association with the expert. We also study the impact of training externalities, that is when the expert directly suffers, e.g. because of competition, or benefits, e.g. because of team production, from the novice's knowledge acquisition. We show that these variations affect the speed of knowledge transfer, but not the core properties of the optimal contract. Finally, we introduce competition between experts for novices with different amounts of initial cash. We show that competition accelerates the speed of knowledge transfer, with those novices with higher initial cash balances obtaining faster training and receiving higher rents.

Our model has a range of implications. The most general one is perhaps the prediction that the training of juniors is artificially delayed in the sense that it takes too long relative to the underlying technological constraints. An instance of this inefficiency are careers in professional services, such as consulting and the law, in which it appears, anecdotally, that juniors spend many years "paying their dues" to the partners. During those years, juniors are involved in wasteful drudgery, rather than in maximizing their learning rate.

Beyond apprenticeships, our model has implications for knowledge transfers in international alliances and joint ventures. The imperfect contractibility that results from poor contractual enforcement in many developing countries means that contracts between companies exhibit the same lack of commitment characteristic of human capital. We study two specific examples of companies developing partnerships in China and argue that the speed of knowledge transfer must trade off, also in those cases, the amount of output generated against the ability of the teacher, in this case the developed-country partner, to extract rents using the promise of further training.

More generally, if patience decreases the rate of knowledge transfer, then features that are traditionally considered to affect this parameter, such as a higher probability of interaction, a longer horizon, or a more reliable partner, will lead, contrary to what might be expected, to larger inefficiencies and slower transfer rates. Some of the empirical implications of our analysis allow a possible falsification of the model. For example, looking at a cross section of industries, the model is falsified if the industries with the lower initial training see larger training rates per period.

The model is highly tractable and can be used as a building block for other models in which human capital acquisition is relevant. In future work, we expect to study training hierarchies, where an expert can train a number of other agents who in turn can train others. Also, in order to emphasize the role of the student's incentive constraints, we have assumed that knowledge transfers are not subject to technological constraints, and so players could transfer all knowledge in one go if they so desired. Future work may consider the implications of the interaction between incentive and technological constraints (such as communication costs, varying languages, and organizational codes) for the rate of knowledge transfer and for the growth of productivity.

References

[1] Acemoglu, Daron (1997). "Training and Innovation in an Imperfect Labour Market." Review of Economic Studies, 64, 445–464.

[2] Acemoglu, Daron and Pischke, Jorn-Steffen (1998). "Why Do Firms Train? Theory and Evidence." Quarterly Journal of Economics, 113, 79–119.

[3] Acemoglu, Daron and Pischke, Jorn-Steffen (1999a). "Beyond Becker: Training in Imperfect Labour Markets." Economic Journal, 106, F112–F142.

[4] Acemoglu, D. and Pischke, J.-S. (1999b). "The structure of wages and investment in general training." Journal of Political Economy, 107(3), 539-572.

[5] Baker, G., R. Gibbons and K. Murphy (1994). "Subjective Performance Measures in Optimal Incentive Contracts." Quarterly Journal of Economics, 109, 1125-1156.

[6] Becker, Gary S. 1994 (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

[7] Biais, B., T. Mariotti, G. Plantin, and J.-C. Rochet. 2007. "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." Review of Economic Studies, 74, 345–90.

[8] Bull, C. (1987). "The Existence of Self-Enforcing Implicit Contracts." Quarterly Journal of Economics, 102, 147-160.

[9] Burdett, Kenneth and Smith, Eric (1996). "Education and Matching Externalities." In Acquiring Skills. Booth, Alison L. and Dennis J. Snower, eds. Cambridge: Cambridge University Press: 65-80.

[10] Chang, Chun and Wang, Yijiang (1996). "Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited." Journal of Labor Economics, 14, 505–519.

[11] DeMarzo, P. M., and Y. Sannikov. 2006. "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model." Journal of Finance, 61, 2681–724.

[12] DeMarzo, P. M., and M. J. Fishman. 2007. "Optimal Long-Term Financial Contracting." The Review of Financial Studies, 20, 6, 2079-2128.

[13] Elbaum, Bernard (1989). "Why Apprenticeship Persisted in Britain But Not in the United States." Journal of Economic History, 49, 337–349.

[14] Fudenberg, D., D. Levine, and E. Maskin (1994). "The Folk Theorem with Imperfect Public Information." Econometrica, 62, 997-1039.

[15] Gale, D., and L. S. Shapley (1962). "College Admissions and the Stability of Marriage." The American Mathematical Monthly, 69, 9-15.

[16] Heckman, James (1971). "Three Essays on Household Labor Supply and the Demand For Market Goods," Chapter 1, Ph.D. Dissertation, Princeton University.

[17] Hörner, Johannes, and Andrzej Skrzypacz (2010). "Selling Information." Cowles Foundation Discussion Paper No. 1743.

[18] Kahn, Charles and Huberman, Gur (1988). "Two-Sided Uncertainty and "Up-or-Out" Contracts." Journal of Labor Economics, 6, 423–444.

[19] Katz, Eliakim and Ziderman, Adrian (1990). "Investment in General Training: The Role of Information and Labour Mobility." Economic Journal, 100, 1147–1158.

[20] Killingsworth, M. (1982) "Learning By Doing and Investment in Training: A Synthesis of Two Rival Models of the Life Cycle," Review of Economic Studies, 49(2), #156, 263-272.

[21] Levin, Jonathan (2003). "Relational Incentive Contracts." American Economic Review, 93, 835-857.

[22] Levin, Jonathan, and Steven Tadelis (2005). "Profit sharing and the role of professional partnerships." The Quarterly Journal of Economics 120, 1, 131-171.

[23] Loewenstein, M., A. and Spletzer, James R (1998). "Dividing the Costs and Returns to General Training." Journal of Labor Economics, 16, 142–171.

[24] Maister, D. H. (1993). *Managing the Professional Service Firm*. Free Press, New York, NY.

[25] Malcomson, J. M., J. W. Maw, B. McCormick (2003)."General training by firms, apprentice contracts, and public policy." European Economic Review, Volume 47, Issue 2, 197–227.

[26] MacLeod, W. B. and J. Malcomson (1989). "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." Econometrica, 57, 447-480.

[27] MacLeod, W. B. and J. Malcomson (1998). "Motivation and Markets." American Economic Review, 88, 388-411.

[28] "OECD Note on quality apprenticeships for the G20 task force on employment." September, 2012. http://www.oecd.org/els/emp/OECD%20Apprenticeship%20Note%2026%20Sept.pdf.

[29] Pearce, D. G. and E. Stacchetti (1998). "The Interaction of Implicit and Explicit Contracts in Repeated Agency." Games and Economic Behavior, 23, 75-96.

[30] Prendergast, Canice (1993). "The Role of Promotion in Inducing Specific Human Capital Acquisition." Quarterly Journal of Economics, 108, 523-534.

[31] Richter, Ansgar, Michael Dickmann, and Michael Graubner (2008). "Patterns of human resource management in consulting firms." Personnel review 37, no. 2, 184-202.

[32] Rosen, Sherwin (1972). "Learning and Experience in the Labor Market," The Journal of Human Resources, 7(3), 326-342.

[33] Shaw, K. (1989), "Lifecycle Labor Supply with Human Capital Accumulation," International Economic Review, 30, 431-456.

[34] Spear, S. E. and Srivastava, S. (1987). "On Repeated Moral Hazard with Discounting." Review of Economic Studies, 54, 599-617.

[35] Stevens, Margaret (1994). "Labour Contracts and Efficiency in On-the-Job Training." Economic Journal, 104, 408–419.

[36] Waldman, Michael (1990). "Up-or-Out" Contracts: A Signaling Perspective." Journal of Labor Economics, 8, 230–250.

[37] Weiss, Y. (1972). "On the Optimal Lifetime Pattern of Labour Supply." Economic Journal, 82, 328, 1293-1315.

IX Appendix A

This appendix completes the proof of lemma 1. We begin with a remark.

Remark 3 *Suppose $\mathcal{C}^*$ is an optimal contract such, for some t, $X_t^* < X_{\sup}(\mathcal{C}^*)$ and $m_t^* < \frac{1}{\delta}B_{t-1}(\mathcal{C}^*)$. Then, there exists a contract $\mathcal{C}'$ with the following features: (1) it delivers the same payoff as $\mathcal{C}^*$; (2) $m_t' = \frac{1}{\delta}B_{t-1}(\mathcal{C}')$; and (3) IC_{t+1} is slack.*

Proof. Let $\frac{1}{\delta}B_{t-1}(\mathcal{C}^*) - m_t^* = \Delta > 0$ and, without loss, set $c_k^* = 0$, and therefore $V_k(\mathcal{C}^*) = \delta^{t+2-k}V_{t+2}(\mathcal{C}^*)$, for all $k \leq t+2$.

Now let $\mathcal{C}'$ be identical to $\mathcal{C}^*$ except for the following changes: (1) $m_t' = m_t^* + \Delta$; (2) $m_{t+1}' = m_{t+1}^* - \frac{1}{\delta}\Delta$. Note that $V_k(\mathcal{C}') = V_k(\mathcal{C}^*)$ for all k, $B_k(\mathcal{C}') = B_k(\mathcal{C}^*)$ for all $k \neq t$, and $\frac{1}{\delta}B_t(\mathcal{C}') = \frac{1}{\delta}B_t(\mathcal{C}^*) - \frac{1}{\delta}\Delta$.

Under the new contract $\mathcal{C}'$, L_t holds with equality, IC_t is unaffected, L_{t+1} continues to hold (as $\frac{1}{\delta}B_t$ falls by $\frac{1}{\delta}\Delta$ and m_{t+1} also falls by $\frac{1}{\delta}\Delta$), and IC_{t+1} is relaxed because $\frac{1}{\delta}B_t$ falls. ■

Now suppose toward a contradiction that $\mathcal{C}^*$ is an optimal contract and yet there exists a period k such that $X_k^* < X_{\sup}(\mathcal{C}^*)$ and $m_k^* < \frac{1}{\delta}B_{k-1}(\mathcal{C}^*)$. Owing to the above remark we can replace $\mathcal{C}^*$ with an optimal contract $\mathcal{C}'$ with the same knowledge transfers as $\mathcal{C}^*$ such that: (i) IC_{k+1} is slack; and (ii) for all t, $X_t' < X_{\sup}(\mathcal{C}')$ implies $m_t' = \frac{1}{\delta}B_{t-1}(\mathcal{C}')$. From the proof of lemma 2, $\mathcal{C}'$ must have a finite training phase and, from the proof of lemma 3, every incentive constraint during the training phase of $\mathcal{C}'$ (which includes $k+1$) must bind, which contradicts the fact that $\mathcal{C}'$ is optimal in the first place.

X Appendix B: proof of lemma 3

Proof. Let T denote the last period of the training phase. Without loss, set $c_t^* = 0$, and therefore $V_t(\mathcal{C}^*) = \delta^{T-t}V_T(\mathcal{C}^*)$, for all $t < T$.

From step 1,

$$\begin{aligned} m_t^* &= \frac{1}{\delta}f(X_{t-1}^*) \text{ for all } t < T, \text{ and} \\ V_T(\mathcal{C}^*) &= \frac{1}{\delta}B_{T-1} - m_T^* + \frac{1}{1-\delta}f(X_{\sup}(\mathcal{C}^*)) = \frac{1}{\delta}f(X_{T-1}^*) - m_T^* + \frac{1}{1-\delta}f(X_{\sup}(\mathcal{C}^*)). \end{aligned}$$

Therefore, the expert's payoff is

$$\sum_{t=2}^{T-1} \delta^{t-2} f(X_{t-1}^*) + \delta^{T-1} m_T^* =$$
$$\sum_{t=2}^{T-1} \delta^{t-2} f(X_{t-1}^*) + \delta^{T-1} \left[\frac{1}{\delta} f(X_{T-1}^*) + \frac{1}{1-\delta} f(X_{\sup}(\mathcal{C}^*)) - V_T(\mathcal{C}^*) \right].$$

where the values of X_t^* and c_t^* satisfy the novice's incentive constraints

$$\frac{1}{1-\delta} f(X_{t-1}^*) + \frac{1}{\delta} f(X_{t-1}^*) = \frac{1}{\delta(1-\delta)} f(X_{t-1}^*) \leq \delta^{T-t} V_T(\mathcal{C}^*) \text{ for all } t \leq T.$$

First, we claim that all such constraints must hold with equality. Notice that the R.H.S. is increasing in t. Now suppose, contrary to this claim, that at least one such constraint holds with strict inequality. Select the largest $t' < T$ such that the corresponding constraint holds with strict inequality, and so $f(X_{t'}^*) < f(X_{t'+1}^*)$. Notice that the expert could have attained a higher payoff by increasing $X_{t'}^*$ by a small amount while holding every other X_t^* constant, a contradiction.

It follows that the expert's payoff further simplifies to

$$\sum_{t=2}^{T-1} \delta^{T-1}(1-\delta) V_T(\mathcal{C}^*) + \delta^{T-1} \left[(1-\delta) V_T(\mathcal{C}^*) + \frac{1}{1-\delta} f(X_{\sup}(\mathcal{C}^*)) - V_T(\mathcal{C}^*) \right]$$
$$= \delta^{T-1} \left[(T-1)(1-\delta) V_T(\mathcal{C}^*) + \frac{1}{1-\delta} f(X_{\sup}(\mathcal{C}^*)) - V_T(\mathcal{C}^*) \right],$$

where $V_T(\mathcal{C}^*)$ satisfies two restrictions: (1) $V_T(\mathcal{C}^*) \geq \frac{1}{1-\delta} f(X_{\sup}(\mathcal{C}^*))$ (which follows from the fact that $\frac{1}{\delta} B_{T-1} - m_T^* \geq 0$); and (2) $V_T(\mathcal{C}^*) < \frac{1}{\delta(1-\delta)} f(X_{\sup}(\mathcal{C}^*))$ (which follows from the incentive constraint for period T holding with equality and, simultaneously, $f(X_{T-1}^*)$ being lower than $f(X_{\sup}(\mathcal{C}^*))$). Notice that the expert's payoff is linear in $V_T(\mathcal{C}^*)$ and she can freely vary the value of $V_T(\mathcal{C}^*)$ (subject to the two restrictions above) by varying m_T^*.

Second, we claim that $(T-1)(1-\delta) < 1$. Suppose not. Then, given our genericity assumption (i.e. $n(1-\delta) < 1$ for every $n = 1, 2, ...$) we must have $(T-1)(1-\delta) > 1$. But in this case, the expert could increase her payoff by increasing $V_T(\mathcal{C}^*)$ by a small amount while still satisfying the two restrictions above, a contradiction.

Third, and finally, given that $(T-1)(1-\delta) < 1$, the optimality of $V_T(\mathcal{C}^*)$ requires this value to be equal to $\frac{1}{1-\delta} f(X_{\sup}(\mathcal{C}^*))$ (i.e. the lowest value allowed by the restrictions above). It follows that $m_T^* = \frac{1}{\delta} f(X_{T-1}^*)$, which in turn implies $m_T^* = \frac{1}{\delta} B_{T-1}$. ■

Figure 1: Optimal duration with externalities (level curves):

Level curves are a function of γ (the externality) and δ (the discount factor). Lighter areas correspond to longer durations.

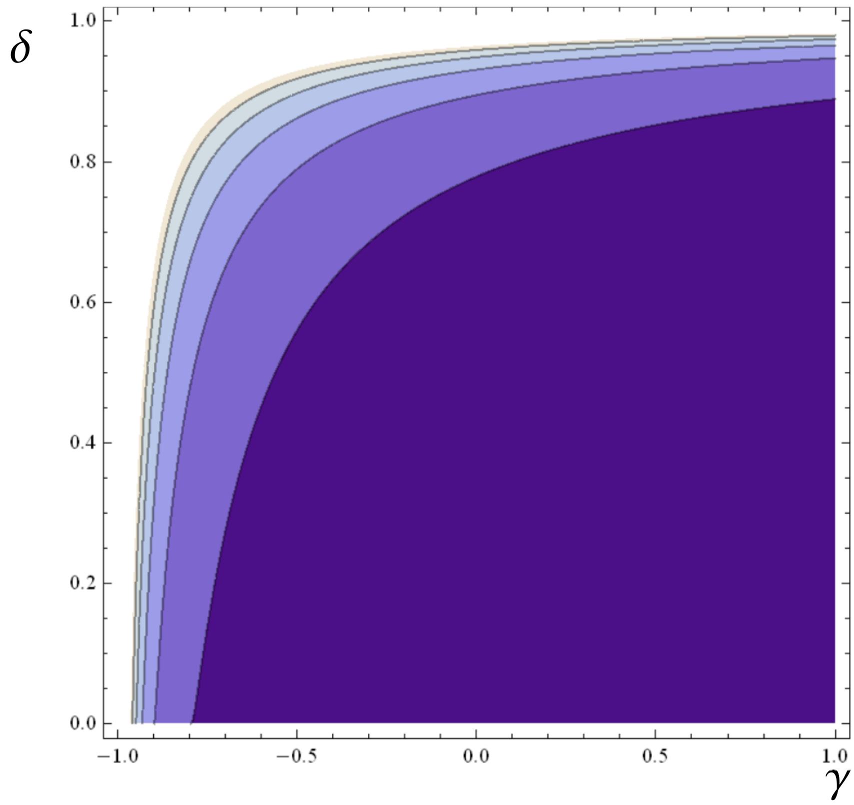

Figure 2: Optimal duration with specific knowledge (level curves)

Level curves are a function of λ (degree of generality of knowledge) and δ (the discount factor). Lighter areas correspond to longer durations.