Repeated Selection with Heterogenous Individuals and Relative Age Effects

Herbert Dawid
Gerd Muehlheusser

Repeated Selection with Heterogenous Individuals and Relative Age Effects

Abstract

In contexts such as education and sports, skill-accumulation of individuals over time crucially depends on the amount of training they receive, which is often allocated on the basis of repeated selection. We analyze optimal selection policies in a model of endogenous skill formation where, apart from their ability to transform training into skills, individuals also differ with respect to relative age. The latter has been identified by recent empirical research as a major determinant for performance differentials within cohorts. We find that the optimal policy is pro-competitive at later selection stages in the sense of selecting the individuals with the higher skill signals. All eventual corrections due to relative age occur at early stages, where selection is either counter-competitive (i.e. individuals with low skill signals are selected) or even avoided at all. Thereby, the induced selection quality is non-monotone in the degree of ex-ante asymmetry due to relative age. Finally, the (empirical) observation of persistent relative age effects does in general not hint at suboptimal selection policies.

JEL-Code: J240, M530, I250, I280.

Keywords: skill formation, human capital, selection, heterogeneity, age effects, training, education.

Herbert Dawid
Department of Economics
Bielefeld University
Bielefeld / Germany
hdawid@wiwi.uni-bielefeld.de

Gerd Muehlheusser
Department of Economics
University of Hamburg
Hamburg / Germany
gerd.muehlheusser@wiso.uni-hamburg.de

1 Introduction

Motivation In many contexts such as education and sports, individuals accumulate skills over a long period of time. In doing so, they differ with respect to the amount of training resources and opportunities received, which in turn is often determined by a sequence of selection (or *streaming*) decisions to the effect that a subset of (selected) individuals is provided with more resources and better opportunities to accumulate skills than others. Examples in each of the two contexts include access to more advanced or fast-track curriculum groups and better coaches, more practice time, and stronger opponents, respectively.

The need to select is often driven by the scarcity of resources such that not every individual can be trained with high intensity. As a result, efficiency commands to devote these scarce resources to those individuals where they are most effective in enhancing skills.

In practice, allocating these resources efficiently across individuals is a major challenge for decision makers because of two confounding factors both of which are of great empirical importance in the contexts of education and sports: First, individuals are typically heterogeneous with respect to their *ability* to transform the training opportunities received into skills, and this ability is typically unobservable to the decision maker. Second, there are often other factors which obscure the eventual "gleaming" (Allen and Barnsley, 1993) of ability, e.g. through observable performance or skill signals. In the context of education and sports, one prominent factor of this type is the *relative age* of individuals which gives older cohort members (i.e. those who are born shortly after the cutoff date for cohort membership) a head-start advantage over younger ones. The existence of potentially sizeable relative-age effects is well-documented by a large body of recent empirical research, according to which older cohort members perform systematically and considerably better than younger ones (for evidence from a large number of OECD countries, see e.g. Bedard and Dhuey, 2006).

Furthermore, relative age does not only seem to matter during childhood where age differences (of up to one year) might indeed entail sizeable differences in cognitive and non-cognitive skills, thereby potentially explaining the better performance of older chil-

dren. Rather, there is also evidence for age effects to persist even in adolescence and adulthood, where the direct effect of relative age on skill levels should be negligible: For example, Bedard and Dhuey (2006) and Sampaio, Da Matta, Ribas, and Sampaio (2011) find that younger cohort members are significantly less likely to attend university than older ones. Similarly, Fredriksson and Öckert (2005) find evidence that younger cohort members receive lower wages on average upon entering the labor market. Finally, in many sports such as soccer, ice-hockey or tennis, the birth day distribution of top athletes is often markedly different compared to the underlying population (see e.g. Barnsley and Thompson, 1988; Dudink, 1994; Helsen, Van Winckel, and Williams, 2005; Cobley, Schorer, and Baker, 2008).

The persistence effect of relative age raises the question concerning the channel through which it feeds through the whole process of skill formation. In this respect, it is typically argued (see e.g. Bedard and Dhuey, 2006) that when selection occurs repeatedly and on a competitive basis, relative age tends to systematically favor older cohort members with respect to the availability of skill formation opportunities. For example, because of their initial edge, older cohort members might systematically receive more high-intensity training, better access to fast-track or advanced curriculum programs. This allows them to accumulate more skills than younger members, and it might even make later human capital investments more valuable (Cunha and Heckman, 2007). As a result of both effects, an initial advantage due to relative age also translates into a permanent one, in which case an eventual gleaming of ability seems less likely to occur.

More precisely, under the 'non-astrology' (Allen and Barnsley, 1993) assumption that birth dates and ability are uncorrelated, competitive selection policies will tend to be erroneous in the sense that training resources will not be devoted to those individuals for which they are most valuable, but to those with the strongest signals, thereby systematically favoring older cohort members. Such errors do not only occur in early stages of the skill formation process (where it is often hard to distinguish between ability and maturity), but they also persist over time because of path-dependencies which arise when the effect of training is sufficiently strong such that old cohort members with low-ability

who do receive high-intensity training continue to outperform their young counterparts with high-ability who do not.

Surprisingly, while the prevalence of relative-age effects and their persistence are by now empirically well-documented, the issue of how scarce resources should be allocated within cohorts to efficiently promote skill formation has received virtually no attention so far. In this paper, we aim at narrowing this gap by analyzing optimal selection strategies in a model of endogenous skill formation and age effects.

Framework and Results We consider a dynamic model where individuals accumulate skills over time, which depend on the amount of training received, the (unobservable) ability (or talent) to transform training into skills, and their relative age, the latter giving older cohort members a head-start advantage. In each period, upon receiving a noisy signal about each individual's skill, a planner decides on how much training to provide to each individual. Due to a budget constraint, she can either avoid selection and train all individuals with equal (intermediate) intensity. Alternatively, she can select a subset of individuals and provide high-intensity training to these (and low-intensity training to the remaining ones). We focus on simple selection rules which are either pro- or counter-competitive in the sense that those individual with high or low skill signals are selected, respectively, or random in the sense that selection occurs randomly, independent of the skill signal. The planner's aim is to maximize the total (or average) skills in the population at the end of the training process.

Our main results can be summarized as follows: The optimal selection policy is always pro-competitive at later stages of the selection process, but not necessarily at the beginning, where all eventual corrections for the asymmetries induced by the initial age effect take place. Not selecting pro-competitively throughout avoids irreversible path-dependencies in the sense that those who perform better at the beginning (where old individuals enjoy a head-start advantage) will also, via the high-intensity training received upon being selected, be very likely to do so in the future. In this respect, if the initial age advantage is weak, then a "wait-and-see" policy is optimal under which early selection is avoided and all individuals receive the same amount of training. Intuitively, unless

hampered by an erroneous selection, ability will eventually 'gleam', thereby leading to efficient selection decision at later stages. In contrast, when the relative-age advantage is strong, a more active, counter-competitive policy is optimal which, to ensure a level playing field, is heavily biased at the beginning in favor of young individuals. By doing so, efficient selection in later stages is only mildly hampered. Finally, when the relative-age advantage is medium, the optimal police is also competitive in the beginning which, because of the induced path-dependencies, leads to a lower selection quality compared to weak and strong relative-age advantage. Interestingly, this points at a non-monotonic relationship between the degree of ex-ante bias of skill signals (relative-age advantage) and the induced selection quality in the sense that the latter is highest when the asymmetry is either small or large, while it is lower in-between.

As a further result, we also find that random selection is never optimal. Last, but not least, while it is possible to fully avoid a persistent influence of relative age in the course of the selection process (referred to as relative-age *effect*), it is not optimal to do so, and the relative-age effect will always be non-zero (positive of negative) under the optimal rule. Therefore, the empirical observation of relative-age effects cannot per se be taken as an indication for the use of non-optimal selection schemes.

Relation to the Literature To the best of our knowledge, this paper is the first to provide a full-fledged theoretical framework for analyzing optimal selection strategies in the context of skill formation, where individuals are heterogeneous with respect to ability and relative age. It is therefore related to the following lines of literature:

Firstly, a large part of the literature is concerned with documenting both the existence and the persistence of relative-age effect in the context of education (e.g. Bedard and Dhuey, 2006; McEwan and Shapiro, 2008; Puhani and Weber, 2008; Elder and Lubotsky, 2009; Billari and Pellizzari, 2008; Sprietsma, 2010; Mühlenweg and Puhani, 2010; Sampaio, Da Matta, Ribas, and Sampaio, 2011; Jürges and Schneider, 2011) and sports (e.g. Barnsley and Thompson, 1988; Dudink, 1994; Helsen, Starkes, and Van Winckel, 1998; Helsen, Van Winckel, and Williams, 2005; Wattie, Baker, Cobley, and Montelpare, 2007; Cobley, Schorer, and Baker, 2008; Till, Cobley, Wattie, O'Hara, Cooke, and Chapman,

2010). Our paper complements this line of research by providing a theoretical framework for a number of resulting policy implications. For example, our results are by and large consistent with proposals that early stages of the selection process should (i) not be too competitive (Helsen, Van Winckel, and Williams, 2005), (ii) avoid irreversible decisions with respect to educational tracks (Sprietsma, 2010; Jürges and Schneider, 2011), and (iii) show more leniency with respect to binding cutoff dates (Bedard and Dhuey, 2006).

Secondly, with respect to selection errors, our paper formalizes and extends the analysis of Allen and Barnsley (1993) who point attention to the social costs due to inefficient allocation of training resources in the presence of age effects. In particular, the optimal selection rules we characterize are precisely those under which those selection errors are minimized. Moreover, our analysis reveals under which conditions ability can be expected to eventually 'gleam', in which case the optimal policy is to avoid selection in early periods.

Thirdly, apart from age and ability, the process of skill formation in our framework is crucially driven by the intensity of training which individuals receive. In this respect, our framework is in line with a prominent body of research in social psychology which stresses the crucial role of training and "deliberate practice" (Ericsson, Krampe, and Tesch-Römer, 1993) in accumulating expert skills, as opposed to being mainly driven by innate (born) talent.[1]

Finally, in our model training and ability are complements in the sense that the former is more effective for high-ability individuals. Complementarities also play an important role in the framework by Cunha and Heckman (2007), where the effectiveness of human capital investments is the higher the stock of skills accumulated in the past.[2] While both frameworks address the issue of skill formation over time, apart from the different forms of complementarities considered, they differ in a number of further aspects: First, we explicitly consider age effects. Second, Cunha and Heckman (2007)

---

[1]See also Ericsson (2006). This line of research has also sparked a considerable public interest, see e.g. Dubner and Levitt (2006), Colvin (2008), Coyle (2009), and Gladwell (2008).

[2]See Cunha, Heckman, Lochner, and Masterov (2006) for an overview how this theory can be used to reconcile a large number of empirical findings in the context of skill formation over the life-cycle.

focus on a given individual's optimal inter-temporal investment profile so that, because of the complementarities between previous and current investments, optimal investment levels should be particularly high at the beginning of the skill formation process in early childhood. In contrast, we analyze optimal inter-temporal *selection* policies for a given cohort of individuals, which is heterogenous with respect to ability and relative age. In at least one important aspect, however, we view our analysis as complementary: Our finding that all eventual deviations from a competitive selection policy should occur at early rounds of selection also points at the crucial importance of the initial phase of the skill formation process. To further investigate the relationship between the two frameworks, in Section 5 we extend our basic model to also exhibit complementarities of the type considered in Cunha and Heckman (2007). Our results appear to be qualitatively robust also with respect to this extension, where values of the elasticity of substitution between past and current human capital investments are chosen as estimated in Cunha, Heckman, and Schennach (2010).

The remainder of the paper is organized as follows: The model framework is laid out in Section 2 and analyzed in Section 3, where we first provide a number of preliminary results and then characterize the optimal selection rule. Section 4 considers the persistence of relative-age effects. As a robustness check, Section 5 analyzes a number of model extensions using a computational approach. Finally, Section 6 discusses our findings and concludes. All proofs are in the Appendix.

2   The Model

There are two individuals $i = 1, 2$ and throughout, we use the convention that individual 1 is old and individual 2 is young. Individuals are ex ante identical with respect to their (unobservable) ability $a_i \in \{\underline{a}, \overline{a}\}$ which is a random variable where $Prob(a_i = \underline{a}) = Prob(a_i = \overline{a}) = \frac{1}{2}$ for $i = 1, 2$. Low ability $\underline{a}$ is normalized to 1 and we write $\overline{a} = (1 + \gamma)\underline{a} = (1 + \gamma)$, so that $\gamma > 0$ measures the degree of heterogeneity with respect to ability in the population.

The sequence of events is as follows (see also Figure 1): Individuals aggregate skills
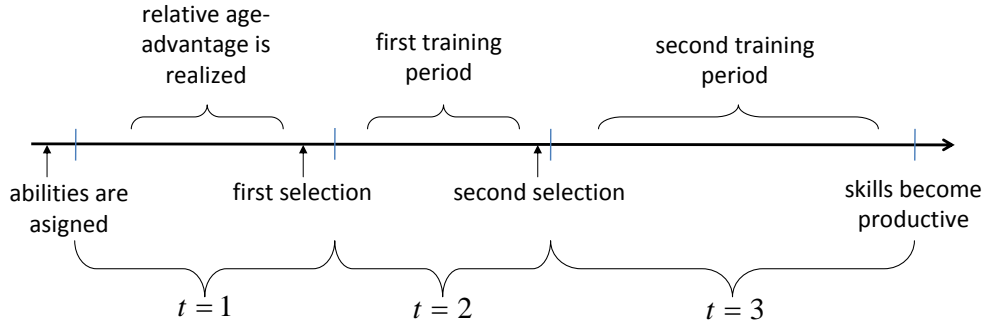
Figure 1: Sequence of Events

over the course of 3 periods $t = 1, 2, 3$, where the (unobservable) skill s_i^t of individual i in period t is given by

$$s_i^t = \begin{cases} a_i \cdot \delta_i^0 & \text{for } t = 1, \\ s_i^{t-1} + a_i \cdot \delta_i^t & \text{for } t = 2, \\ s_i^{t-1} + \lambda a_i \cdot \delta_i^t & \text{for } t = 3. \end{cases} \tag{1}$$

In period $t = 1$, δ_i^0 denotes the amount of training received by individual i *before* the start of the formal skill formation process considered here (e.g. pre-school education at home), where $\delta_2^0 = 1$ and $\delta_1^0 = \beta \geq 1$ which reflects the relative-age advantage of the old individual 1.

In the subsequent periods 2 and 3, δ_i^2 and δ_i^3 measure the intensity with which individual i is trained during periods 2 and 3, respectively. Thereby the increase in skills due to training in period 3 is enlarged by a parameter $\lambda \geq 1$ relative to period 2 which captures the idea that the second selection stage is the more important one. For example, in the context of education, the more rigid forms of (formal) streaming (e.g. academic versus vocational track) are typically preceded by less rigid ones such as math or reading groups or enrichment programs in primary school. As documented in Bedard and Dhuey (2006, p. 1470), this is the case for 16 out of the 19 countries considered.[3]

_____

[3]As we will show in Section 5 below, when allowing for more than two (equally important) rounds of selection, λ can also be viewed as to capture all training decisions from period $t = 3$ onwards, which naturally leads to $\lambda > 1$. Note also that $\lambda \geq 1$ refers to a property of the skill accumulation technology; it does not reflect any time preference for the different periods in which training occurs.

Importantly, in all periods ability and training intensity are complements in increasing skills, as training is more effective for high-ability individuals. As a result, the aggregate level of skills in the population will depend on how the training intensities δ_i^2 and δ_i^3 are assigned across individuals. These assignments are carried out by the planner at the end of dates $t = 1$ and $t = 2$, respectively, subject to the budget constraint

$$\delta_1^t + \delta_2^t \leq 2\delta. \tag{2}$$

Thereby, either both individuals are trained equally, i.e. $\delta_1^t = \delta_2^t = \delta$, or one of the two individuals is *selected* and receives more intensive training than the other, i.e. $\delta_i^t = \delta + x, \delta_j^t = \delta - x, \; i \neq j$ for some positive x.[4] Given the setup of our model, the budget constraint (2) can safely considered to be binding throughout.

Before assigning training intensities, the planner receives a noisy skill signal $\hat{s}_i^t = s_i^t + \eta_i^t$, where η_i^t is normally distributed with mean 0 and variance σ_t^2. In what follows, we confine attention to selection rules y^t which are simple in the sense that they can at most (but do not have to) condition on these two skill signals, but not on other variables such as the identity and the age of individuals or previous skill signals or selection decisions, i.e.

$$y^t : (\hat{s}_1^t, \hat{s}_2^t) \rightarrow (\delta_1^t, \delta_2^t) \in \{\delta, \delta + x, \delta - x\}^2 \quad \text{for} \quad t = 1, 2.$$

Moreover, as for the dependency on the skills signals, we consider only rules which are binary in the sense of only conditioning on whether or not the difference of skill signals is positive or negative, thereby excluding rules which depend on this difference (or even on other operators) in more sophisticated ways. These restrictions leave us with the following four selection rules to be considered throughout:

1. *pro-competitive selection* (P): select the individual with the higher signal,[5]

---

[4]Alternatively, one could think of instructors (teachers, trainers) of different qualities $\delta + x$, δ, and $\delta - x$ which the planner assigns to each individual.

[5]In the context of education, selection decisions, such as admission to particular secondary schools, are often strongly influenced by an applicant's performance in a competitive ability test (such as ACT or SSAT). Furthermore, in countries, such as United Kingdom and Japan, there also exist

2. *counter-competitive selection* (C): select the individual with the lower signal,

3. *random selection* (R): ignore the signal and select each individual with probability $\frac{1}{2}$,

4. *equal training* (E): do not select any individual, but provide equal training of δ to each of them.

In what follows, we focus on the case where the planner maximizes the total expected skills after training is over:

$$\max_{y^1, y^2} \ \mathbb{E}[s_1^3 + s_2^3] \quad \text{s. t. budget constraint (2)} \tag{3}$$

In line with the strand of literature discussed above which argues that to a large extent, "(stars) are made, not born" (Dubner and Levitt, 2006), we assume throughout that although ability is an important factor for skill accumulation, the primary determinant of the speed of skill accumulation is the intensity of training received:

Assumption 1 *A low-ability individual which has been selected in a given period acquires more skills during that period than a high-ability individual which has not been selected, i.e. $\delta + x > (1 + \gamma)(\delta - x)$ which is equivalent to $\gamma < \gamma^m := \frac{2x}{\delta - x}$.*

Hence, the assumption gives an upper bound on the heterogeneity with respect to ability. Furthermore, we rule out scenarios where the relative age advantage is so strong that it dominates all effects of ability and subsequent training:

Assumption 2 *A low-ability old individual with low-intensity training in period 1 has a lower skill at the end of period 2, than a high-ability young individual with high-intensity training in period 1, i.e. $\beta + \delta - x < (1 + \gamma)(1 + \delta + x)$ which is equivalent to $\beta < \beta^m := 1 + 2x + \gamma(1 + \delta + x)$.*

---

numerous schools, where admission is based *solely* on the performance in entrance exams. One example for the latter are grammar schools in Birmingham and Black County in the U.K. (see www.birmingham.gov.uk/schooladmissions).

3 Characterizing the Optimal Selection Policy

In characterizing the optimal selection policy, we proceed in three steps: We first provide some crucial preliminary results (Section 3.1), then consider a benchmark case with no relative-age advantage (Section 3.2) and finally analyze the case with relative-age advantage (Section 3.3).

3.1 Preliminaries

Define first the following two ability configurations $\overline{A} := \{a_1 = \overline{a}, a_2 = \underline{a}\}$ and $\underline{A} := \{a_1 = \underline{a}, a_2 = \overline{a}\}$. Moreover, denote by

$$\overline{\mu}_1 := Prob(s_1^1 > s_2^1 | \overline{A}) \quad \text{and} \quad \underline{\mu}_1 := Prob(s_1^1 > s_2^1 | \underline{A}),$$

the probabilities that the old individual sends the higher signal in period 1 under ability configurations $\underline{A}$ and $\overline{A}$, respectively. Analogously, for period 2, define

$$\overline{\mu}_2^z := Prob(s_1^2 > s_2^2 | \overline{A}, z) \quad \text{and} \quad \underline{\mu}_2^z := Prob(s_1^2 > s_2^2 | \underline{A}, z)$$

as the probabilities that the old individual sends the higher signal in period 2 under ability configuration $\underline{A}$ and $\overline{A}$ and selection history $z \in \{s, n, e\}$, which denotes the amount of training which individual 1 has received in period 1, where $z = s$, $z = n$, and $z = e$ indicate high-intensity $(\delta + x)$, low-intensity $(\delta - x)$ and equal (δ) training, respectively.

Lemma 1

(i) $\underline{\mu}_1 < \overline{\mu}_1$, $\quad \overline{\mu}_1 > \frac{1}{2}$ $\quad and \quad \underline{\mu}_1 + \overline{\mu}_1 \geq 1$

(ii) $\underline{\mu}_2^z < \overline{\mu}_2^z$, $\quad \forall z \in \{s, n, e\}$

(iii) $\overline{\mu}_2^n < \overline{\mu}_2^e < \overline{\mu}_2^s$ $\quad and \quad \underline{\mu}_2^n < \underline{\mu}_2^e < \underline{\mu}_2^s$

Intuitively, as for the part (i), (the old) individual 1 is more likely to send the higher signal in period 1 under ability of configuration $\overline{A}$ compared to $\underline{A}$. Under configuration $\overline{A}$, this probability is strictly larger than $\frac{1}{2}$ because the individual benefits from both

high ability and a relative-age advantage. The latter also gives individual 1 an overall advantage across ability configurations. Part (ii) of the Lemma establishes that, for a given selection history, the probability for individual 1 sending the higher signal in period 2 is increasing in individual 1's ability. Finally, as for part (iii), under each ability configuration, the probability for individual 1 sending the higher signal in period 2 is increasing in the intensity of training received by individual 1 in period 1.

Consider next in more detail the properties of the different rule types and denote by $S^{y^1 y^2}$ the expected total skills under rules y^1 and y^2 in the first and second round of selection, respectively. This can be more generally expressed as

$$S^{y^1 y^2} = Z + \frac{1}{2} \left[(\overline{\omega}_1 - \underline{\omega}_1) K + (\overline{\omega}_2^s \overline{\omega}_1 + \overline{\omega}_2^n (1 - \overline{\omega}_1) - \underline{\omega}_2^s \underline{\omega}_1 - \underline{\omega}_2^n (1 - \underline{\omega}_1)) \lambda K \right] \quad (4)$$

where $Z := \frac{2+\gamma}{2} \cdot [(1+\beta) + 2\delta(1+\lambda)]$ are the expected total skills when selection is either fully random ($y^1 = y^2 = R$) or does not occur at all ($y^1 = y^2 = E$). Thereby, the terms $\frac{2+\gamma}{2}$, $(1+\beta)$ and $2\delta(1+\lambda)$ reflect the expected ability of each individual, the total skill level before training starts, and the total skill increase due to training under either no and random selection, respectively.

The remaining terms in Eqn. (4) refer to the expected deviations due to the (non-random) selection process in periods 1 and 2, where $K := x\gamma$ captures the additional gain in expected skills if, in a given period, the high-ability individual is selected.[6] As for the probabilities of selection, we define

$$\overline{\omega}_1 := Prob(\delta_1^1 = \delta + x | \overline{A}) \quad \text{and} \quad \underline{\omega}_1 := Prob(\delta_1^1 = \delta + x | \underline{A}),$$

$$\overline{\omega}_2^z := Prob(\delta_1^2 = \delta + x | \overline{A}, z) \quad \text{and} \quad \underline{\omega}_2^z := Prob(\delta_1^2 = \delta + x | \underline{A}, z)$$

for a given period 1 selection history $z = s, n, e$. Clearly, whether or not these probabilities coincide with those for individual 1 sending the higher signal (μ-notation) depends on the rule considered: For example, for $y^1 = P$, we have $\overline{\omega}_1 = \overline{\mu}_1$ and $\underline{\omega}_1 = \underline{\mu}_1$, for $y^1 = C$, we have $\overline{\omega}_1 = 1 - \overline{\mu}_1$ and $\underline{\omega}_1 = 1 - \underline{\mu}_1$, and for $y^1 = R$, we have $\overline{\omega}_1 = \underline{\omega}_1 = \frac{1}{2}$.

---

[6]To see this, note that under both rules EE or RR, the aggregate skill increase in period 2 is $(1+\gamma)\delta + \delta = (2+\gamma)\delta$. Under efficient training, we get $(1+\gamma)(\delta + x) + \delta - x = (2+\gamma)\delta + \gamma x$, leading to a difference of γx.

Although selection does not occur under rule E, Eqn. (4) nevertheless encompasses the resulting expected skills as $y^1 = E$ leads to $\overline{\omega}_1 = \underline{\omega}_1 = 0$, and $y^2 = E$ to $\overline{\omega}_2^s = \overline{\omega}_2^n = 0$ and $\underline{\omega}_2^s = \underline{\omega}_2^n = 0.$[7]

From this discussion, it follows that maximizing $S^{y^1 y^2}$ is just equivalent to choosing the policy which maximizes the probability of selecting the high-ability individual in each period. In doing so, in our setup with two rounds the planner can choose from a set of $4^2 = 16$ rules. The following result establishes that this set can be narrowed down considerably:

Proposition 1 (Optimal rules) *No rule other than PP, CP or EP can be optimal.*

Note first that random selection (rule R) is never part of the optimal policy, while equal training (rule E) can be.[8] Moreover, our result that rule P is always optimal in period 2 suggests that later selection decisions should be made at a competitive basis (and it implies that $\overline{\omega}_2^z = \overline{\mu}_2^z$ and $\underline{\omega}_2^z = \underline{\mu}_2^z$ throughout for all $z = s, n, e$). Intuitively, the expected gain from selecting correctly under configuration $\overline{A}$ outweighs the loss from erroneous selection under configuration $\overline{A}$, so that rule P outperforms all other rules in the second selection stage. It is worth noting that, as there is no effort choice in our model, this result seems qualitatively different from those obtained in the literature on tournaments and contests in the spirit of Lazear and Rosen (1981) and Tullock (1980), where competitive and symmetric environments are often desirable, because they elicit strong effort incentives (see e.g. Amann and Leininger, 1996; Krishna and Morgan, 1997; Moldovanu and Sela, 2006; Feess, Muehlheusser, and Walzl, 2008).

To ensure that this second selection process is not too unbalanced, all eventual corrections to account for ex ante asymmetries should occur at the first selection stage.

---

[7] This highlights a crucial difference between rules E and R: For example, while each individual receives an *expected* amount of training of δ under both rules $y^1 = E$ and $y^1 = R$, (random) selection does occur under rule R leading to $\overline{\omega}_2^s \neq \overline{\omega}_2^n$ and $\underline{\omega}_2^s \neq \underline{\omega}_2^n$. As a result, even though the expected amount of training in period 1 is the same under these two rules, they are not equivalent.

[8] The non-optimality of random rules casts serious doubt on the desirability of a recent policy by the state government of Berlin (Germany) which uses a lottery to allocate a large part of the slots for the top school track 'Gymnasium'.

Therefore, in a next step, we analyze under which conditions, and in which form, such corrections are desirable. To do so, we compare the performance of the three candidate rules PP, CP and EP with respect to their induced selection quality resp. the resulting expected total skills which are given as follows:

$$S^{PP} = Z + \frac{1}{2}\left[(\overline{\mu}_1 - \underline{\mu}_1)K + (\overline{\mu}_2^s\overline{\mu}_1 + \overline{\mu}_2^n(1-\overline{\mu}_1) - \underline{\mu}_2^s\underline{\mu}_1 - \underline{\mu}_2^n(1-\underline{\mu}_1))\lambda K\right] \quad (5)$$

$$S^{CP} = Z + \frac{1}{2}\left[-(\overline{\mu}_1 - \underline{\mu}_1)K + (\overline{\mu}_2^s(1-\overline{\mu}_1) + \overline{\mu}_2^n\overline{\mu}_1 - \underline{\mu}_2^s(1-\underline{\mu}_1) - \underline{\mu}_2^n\underline{\mu}_1)\lambda K\right] \quad (6)$$

$$S^{EP} = Z + \frac{1}{2}\left[(\overline{\mu}_2^e - \underline{\mu}_2^e)\lambda K\right] \quad (7)$$

We refer to a rule $(y^1 y^2)$ as *superior* to rule $(\tilde{y}^1\tilde{y}^2)$ in period t if

$$\mathbb{E}[(s_1^{t+1} + s_2^{t+1}) - (s_1^t + s_2^t)|(y^1 y^2)] > \mathbb{E}[(s_1^{t+1} + s_2^{t+1}) - (s_1^t + s_2^t)|(\tilde{y}^1\tilde{y}^2)],$$

which leads to the following result:

Proposition 2 (Superiority)

(i) *In period 1, rule PP is superior to both rule CP and EP.*

(ii) *In period 2, rule PP is superior to rule CP if and only if*

$$(\overline{\mu}_1 - (1-\overline{\mu}_1)) \cdot (\overline{\mu}_2^s - \overline{\mu}_2^n) + ((1-\underline{\mu}_1) - \underline{\mu}_1) \cdot (\underline{\mu}_2^s - \underline{\mu}_2^n) > 0. \quad (8)$$

(iii) *In period 2, rule PP is superior to rule EP if and only if*

$$\left(\overline{\mu}_2^s\overline{\mu}_1 + \overline{\mu}_2^n(1-\overline{\mu}_1) - \underline{\mu}_2^s\underline{\mu}_1 - \underline{\mu}_2^n(1-\underline{\mu}_1)\right) - (\overline{\mu}_2^e - \underline{\mu}_2^e) > 0. \quad (9)$$

(iv) *In period 2, rule EP is superior to rule CP if and only if*

$$(\overline{\mu}_2^e - \underline{\mu}_2^e) + \left(\overline{\mu}_2^s\overline{\mu}_1 + \overline{\mu}_2^n(1-\overline{\mu}_1) - \underline{\mu}_2^s\underline{\mu}_1 - \underline{\mu}_2^n(1-\underline{\mu}_1)\right) > 0. \quad (10)$$

While the choice of rule is irrelevant when both individuals have the same ability of transforming training received into skills, it does matter when abilities differ (i.e. under configurations $\overline{A}$ and $\underline{A}$), as training is more effective under high-ability. Hence, any differences in resulting surplus stems from each rule's performance with respect

to inducing an efficient selection decision under these two ability configurations. More precisely, if a rule manages to select the high-ability individual, an efficiency gain of size $K = x\gamma$ is realized compared to no (or random) selection.

As for period 1 (part (i) of the Proposition), selection does not occur under rule EP, and both low- and high ability individuals receive the same amount of training, so that the additional gain in skills is zero. In contrast, this gain is positive under rule PP: Under configuration $\overline{A}$, (the high-ability) individual 1 is more likely to be selected than (the low-ability) individual 2, as $\overline{\mu}_1 > 1 - \overline{\mu}_1$ (Lemma 1), leading to an expected gain compared to rule EP. Under configuration $\underline{A}$ using PP might lead to an expected loss compared to EP (for $\underline{\mu}_1 > \frac{1}{2}$), which however is smaller than the expected gain under $\overline{A}$ because $\overline{\mu}_1 > \underline{\mu}_1$ (again by Lemma 1). Finally, as for rule CP, the expected gains and losses in period 1 are just reversed compared to rule PP resulting in an expected net loss relative to both rules PP and EP.

As for period 2 (parts (ii) – (iv) of the Proposition), the same type of comparison applies, although the resulting conditions as stated in the Proposition are less straightfor-wardly interpreted. Intuitively, since all three potentially optimal rules exhibit $y^2 = P$, their relative performance in period 2 needs to be assessed in terms of selecting high-ability individuals in the course of the competitive selection process in the second round. For each rule, this is can be expressed as the difference of individual 1's selection proba-bility under configurations $\overline{A}$ and $\underline{A}$, respectively, conditional on the selection history of the first round.

In this respect, rule PP performs well under configuration $\overline{A}$, as individual 1 is already more likely to be selected in the first round ($\overline{\mu}_1 > \frac{1}{2}$). Together with the resulting high-intensity training, this makes selection likely to occur also in the second. More precisely, the likelihood of selection in the second round is $\overline{\mu}_1 \overline{\mu}_2^s + (1 - \overline{\mu}_1)\overline{\mu}_2^n$ as opposed to $(1 - \overline{\mu}_1)\overline{\mu}_2^s + \overline{\mu}_1 \overline{\mu}_2^n$ under rule CP, leading to a difference of $(2\overline{\mu}_1 - 1)(\overline{\mu}_2^s - \overline{\mu}_2^n) > 0$ (Lemma 1).

However, although less so, individual 1 remains advantaged also under configuration $\underline{A}$, and his likelihood of (erroneous) selection under rule PP is $\underline{\mu}_1 \underline{\mu}_2^s + (1 - \underline{\mu}_1)\underline{\mu}_2^n$, so that the difference to rule CP becomes $(2\underline{\mu}_1 - 1)(\underline{\mu}_2^s - \underline{\mu}_2^n) \gtreqless 0$. When individual 1's

relative-age advantage is sufficiently large such that $\underline{\mu}_1 > \frac{1}{2}$ (which holds for $\beta > 1 + \gamma$, see Lemma 1), then this difference is again positive, in which case rule CP outperforms rule PP as it leads to a lower probability of (erroneously) electing (low-ability) individual 1. In this case, the benefit of a more leveled playing field in the second round might even over-compensate the loss induced by rule CP in the first round of selection.

Finally, as no selection occurs under the "wait-and-see" rule EP in the first round, there is no danger of enlarging existing asymmetries, but it does also not aid in correcting them. Clearly, to determine which rule is ultimately optimal, both rounds of selection need to be considered, also taking into account that the second round is the more important one ($\lambda \geq 1$).

Moreover, apart from the parameters β, γ, λ and x, the exact characterization of the optimal rule will also hinge on the stochastic properties of the signal technology. In this respect, it seems realistic to assume that signals become less noisy as time progresses (i.e. $\sigma_1^2 > \sigma_2^2$). In what follows, we will consider the extreme case $\sigma_2^2 \to 0$, which allows for an analytical characterization of the optimal selection rule.[9]

Lemma 2 (Limit properties of second round signals) *Assume that $\sigma_2^2 \to 0$. Then,*

(i)

$$\overline{\mu}_2^e \to 1, \ \overline{\mu}_2^s \to 1, \ \underline{\mu}_2^s \to 1, \ \underline{\mu}_2^n \to 0$$

(ii) there exists a threshold $\beta^e := 1 + \gamma(1 + \delta)$ such that $\underline{\mu}_2^e \to 1$ if $\beta > \beta^e$ and $\underline{\mu}_2^e \to 0$ if $\beta < \beta^e$. Moreover, β^e is increasing in γ.

(iii) there exists a threshold $\beta^n := \frac{1 + 2x - \gamma(\delta - x)}{1 + \gamma}$ such that $\overline{\mu}_2^n \to 1$ if $\beta > \beta^n$ and $\overline{\mu}_2^n \to 0$ if $\beta < \beta^n$. Moreover, β^n is decreasing in γ.

Intuitively, the Lemma simply allows to further highlight the properties of the different rules in terms of their selection quality in round 2.[10] Hence, because of $\overline{\mu}_2^e \to 1$, rule

---

[9]Using a computational framework, we will verify in Section 5 below that our main results do not qualitatively depend in this extreme assumption.

[10]Recall from Proposition 1 that $\underline{\omega}_2^z = \underline{\mu}_2^z$ and $\overline{\omega}_2^z = \overline{\mu}_2^z$ holds for all $z \in \{s, n, e\}$, as all three relevant rules exhibit $y^2 = P$.

EP ensures that individual 1 is always selected in the second round under configuration $\overline{A}$. Moreover, since $\overline{\mu}_2^s \to 1$ and $\underline{\mu}_2^s \to 1$, rule PP is "path-dependent" in the sense that it always selects individual 1 in the second round again, if 1 was already selected in the first round. Clearly, this is efficient under configuration $\overline{A}$, but not under configuration $\underline{A}$. Finally, since $\underline{\mu}_2^n \to 0$, by biasing the first round against individual 1 (so that the selection history is indeed likely to be $z = n$), rule CP is helpful in avoiding that individual 1 is inefficiently selected in the second round.

While all limit values in part (i) are independent of the size of the relative-age advantage (β), parts (ii) and (iii) consider the two remaining cases $\underline{\mu}_2^e$ and $\overline{\mu}_2^n$, for which the limit value does depend on β; and which in turn will have crucial implications for the characterization of the optimal rule below.

As for part (ii), at $\beta = \beta^e$, both individuals have the same skill (and hence the same expected signal) after one round of equal training under configuration $\underline{A}$ (i.e. $\beta^e + \delta = (1 + \gamma)(1 + \delta)$). Hence, for all $\beta < \beta^e$, individual 2 will always be selected under EP ($\underline{\mu}_2^e \to 0$). Therefore, in combination with $\overline{\mu}_2^e \to 1$ from part (i), this implies that rule EP will always select efficiently in round 2 as long as long as β is small ($\beta < \beta^e$). In contrast, for $\beta > \beta^e$ individual 1's skill (and expected signal) is higher under both ability configurations $\overline{A}$ and $\underline{A}$ after one period of equal training, so that individual 1 will *always* be selected under rule EP in the second round (i.e. even under configuration $\underline{A}$). Finally, β^e is increasing in γ, since an increase in β is required to compensate the larger ability advantage of individual 2.

As for part (iii), at $\beta = \beta^n$, both individuals have the same skill (and hence the same expected signal) under configuration $\overline{A}$ when individual 2 was selected in the first round (i.e. $(1 + \gamma)(\beta^n + \delta - x) = 1 + \delta + x$). Hence, for $\overline{A}$ and all $\beta > \beta^n$, individual 1 will always be selected in the second round, regardless of the selection history $z = n, s, e$ (by definition of β^n this holds even for $z = n$, and hence a fortiori for $z = s$ and $z = e$). In this case, rule CP always efficiently selects individual 1 under configuration $\overline{A}$. As for configuration $\underline{A}$ whether or not 2 is selected depends on the selection history z: for $z = s$ efficient selection does not occur ($\underline{\mu}_2^s \to 1$), but for $z = n$ it does ($\underline{\mu}_2^n \to 0$) and, as shown above, this latter history is likely to occur under rule CP. Finally, β^n is decreasing in γ,

because with a larger γ a lower β is required to compensate for individual 1's training disadvantage in the first round.

Given the two thresholds β^e and β^n, it is useful to categorize the range of admissible values $\beta \in [1, \beta^m)$ as follows:

Definition 1 (Weak, medium and strong relative-age advantage) *The relative-age advantage (β) is called **weak** for $1 \le \beta < \beta^e$, **medium** for $\beta^e < \beta < \max(\beta^e, \beta^n)$, and **strong** for $\max(\beta^e, \beta^n) < \beta < \beta^m$.*[11]

In the light of the discussion of Lemma 2, the definition can be usefully interpreted as follows: when the relative-age advantage is weak, one period of equal training (thereby avoiding selection) suffices to get an efficient selection decision in period 2 under both ability configurations. When it is no longer weak (i.e. $\beta > \beta^e$), selection is warranted and this leads to the issue of "path-dependencies", i.e. whether or not an individual which has been selected in the first round will also *always* be selected in the second, thereby potentially perpetuating any erroneous choice made in period 1. In this respect, when the relative-age advantage is medium, such path-dependencies necessarily exist as an individual is selected in the second round *if only if* it was already selected in the first round (as $\overline{\mu}_2^s \to 1$, $\overline{\mu}_2^n \to 0$, $\underline{\mu}_2^s \to 1$, and $\underline{\mu}_2^n \to 0$); when it is strong, however, they can be avoided (as $\underline{\mu}_2^n \to 1$). As one implication, the selection quality can in fact be *increasing* in the strength of the relative-age advantage (see section 3.3).

Furthermore, comparing β^e and β^n, since β^e (β^n) is increasing (decreasing) in γ, $\beta^e < \beta^n$ holds if and only if

$$\gamma < \frac{x}{1 + \delta}. \tag{11}$$

Clearly, as γ increases, the importance of training becomes smaller compared to ability. In this respect, a further definition turns out to be useful for the characterization of the optimal rule:

Definition 2 (Weak and strong heterogeneity) *The heterogeneity with respect to ability (γ) is **weak** if $\gamma < \frac{x}{1+\delta}$, while it is **strong** for $\gamma \ge \frac{x}{1+\delta}$.*

---

[11]Note that for $\max(\beta^e, \beta^n) = \beta^e$, the set of values with intermediate relative-age advantage is empty. Moreover, it follows from the definitions of β^n and β^m (see Assumption 2) that $\beta^n < \beta^m$ holds.

Note that $\frac{x}{1+\delta} < y^m$, so that the Definition does not interfere with Assumption 1.

3.2   Benchmark: No Relative Age Advantage ($\beta = 1$)

As it allows to highlight some crucial properties of the three rules PP, CP and EP, it is instructive to discuss first a benchmark case without relative-age advantage ($\beta = 1$), so that there is only heterogeneity with respect to ability (γ).[12] Moreover, as will become clear in subsection 3.3 below, the optimal rule characterized next is valid not only for $\beta = 1$, but for the whole range of weak relative-age advantage:

Proposition 3 (Benchmark: No relative-age advantage) *Assume that $\beta = 1$.*

  (i) There exists a λ^* such that rule EP is optimal if $\lambda > \lambda^*$ and σ_2^2 sufficiently small.

  (ii) If $\gamma > \Phi^{-1}\left(\frac{3}{4}\right)\sqrt{2\sigma_1^2}$ then $\lambda^* > 1$ and rule PP is optimal for $1 \leq \lambda < \lambda^*$ and σ_2^2 sufficiently small, where Φ is the distribution function of a standard Gaussian distribution with mean zero and variance one.

(iii) Rule CP is never optimal.

As explained above, rule CP is particularly beneficial for large levels of relative-age advantage, and hence never optimal in the case where $\beta = 1$ considered here.

The comparison of rules EP and PP is driven by both the degree of heterogeneity with respect to ability (γ) and the relative importance of the two selection stages (λ). In this respect, recall first from the discussion of part (i) of Proposition 2, that any potential gain from exploiting the higher complementarities between ability and training for high-ability individuals in the first round of selection ($K = x\gamma$) is forfeited under rule EP, but not under rule PP. Obviously, this gain increasing in γ. Moreover, as γ increases, ability is more likely to 'gleam' in the first selection stage, as $\overline{\omega}_1$ and $\underline{\omega}_1$ are increasing and decreasing in γ, respectively, so that the difference $\overline{\omega}_1 - \underline{\omega}_1$ is increasing in γ. Both effects make rule PP more desirable compared to rule EP. On the other

---

[12]There is no need to consider another benchmark case with $\gamma = 1$ (and $\beta > 1$) since selection does not matter in that case.

	relative-age advantage (β)		
	weak	medium	strong
	$\beta \in [1, \beta^e)$	$\beta^e < \beta < \max(\beta^e, \beta^n)$	$\beta > \max(\beta^e, \beta^n)$
heterogeneity (γ)			
weak ($\gamma < \frac{x}{1+\delta}$)	EP	PP	CP
strong ($\gamma \geq \frac{x}{1+\delta}$)	EP	n.a.	CP

Table 1: The optimal selection rule (Proposition 4)

hand, the crucial advantage of rule EP over PP is that it always selects the high-ability individual in the second round as long as the relative-age advantage is weak (i.e. for $\beta < \beta^e$), while rule PP fails to accomplish this if the low-ability individual has been (erroneously) selected in the first round (as $\underline{\mu}_2^s \to 1$, see Lemma 2). The optimal rule therefore reflects a trade-off between all of these effects and rule EP is optimal whenever the second round is sufficiently important, i.e. when λ is sufficiently large. As argued above, this seems to be the more relevant case in our context(s), such that we will focus on it throughout.

3.3 The Optimal Rule Under Relative Age Advantage ($\beta > 1$)

Consider now the case where individuals are not only heterogeneous with respect to ability, but where there is also a relative-age advantage of individual 1 ($\beta > 1$). Building on our previous results, the optimal rule can then be characterized as follows:

Proposition 4 (Optimal rule with relative-age advantage) *Assume that σ_2^2 is sufficiently small, and that λ is sufficiently large.*

(i) *When the relative-age advantage is weak, rule EP is optimal.*

(ii) *When the relative-age advantage is strong, rule CP is optimal.*

(iii) *When the heterogeneity with respect to ability is weak and the relative-age advantage is medium, rule PP is optimal.*

19

Proposition 4 is illustrated in Figure 2 and summarized in Table 1. As for the intuition of part (i), the highly desirable feature of the "wait-and-see" rule EP of *always* selecting the high-ability individual in the second round of selection (after providing equal training in the first round) holds for all $\beta < \beta^e$ (Lemma 2). As a result, for the whole range $[1, \beta^e)$ of weak relative-age advantage, the results from the benchmark with $\beta = 1$ (Proposition 3) are qualitatively preserved so that, for λ large, rule EP remains optimal.

As for cases (ii) and (iii) where the relative-age advantage is either medium or strong, rule EP ceases to be optimal as it is highly erroneous in that always selects individual 2 in the second round of selection. In these cases, recall that for $\beta > \beta^n$, rule CP always selects individual 1 under configuration $\overline{A}$, independent of the selection decision in the first round. Furthermore, as long as $\beta < \beta^m$, it also efficiently selects individual 2 under configuration $\underline{A}$, provided that the selection history was $z = n$, which for $y^1 = C$ occurs with probability $\overline{\mu}_1 > \frac{1}{2}$.

Hence, under strong heterogeneity (case (ii) of the Proposition) where $\beta^e > \beta^n$, rule CP is optimal for all $\beta > \beta^e$ when the second round of selection is sufficiently important (i.e. for λ large). Similarly, under weak heterogeneity (case (iii) of the Proposition), $\beta^e < \beta^n$, so that rule CP is optimal for all $\beta > \beta^n$. Note that strongly favoring the young individual 2 under CP is optimal here, even though the planner is by assumption not concerned with the distribution of final skills across individuals, but he only cares about the sum.

Finally, (only) under weak heterogeneity, there exists an interval (β^e, β^n) of medium relative-age advantage which exhibits a path dependency in the sense that the individual which is selected in period 1 is *always* also selected in period 2.[13] As a consequence, rule CP performs badly under configuration $\overline{A}$, since with probability $\overline{\mu}_1 > \frac{1}{2}$ individual 2 is selected in period 1, and therefore is selected again in period 2. Accordingly, in this interval, rule PP is optimal: While favoring individual 1 instead, it does so at a less extreme level and therefore outperforms rule CP.

---

[13] As $\beta^n < \beta^e$ holds under strong heterogeneity, the set of values where the relative-age advantage is *intermediate* is empty.

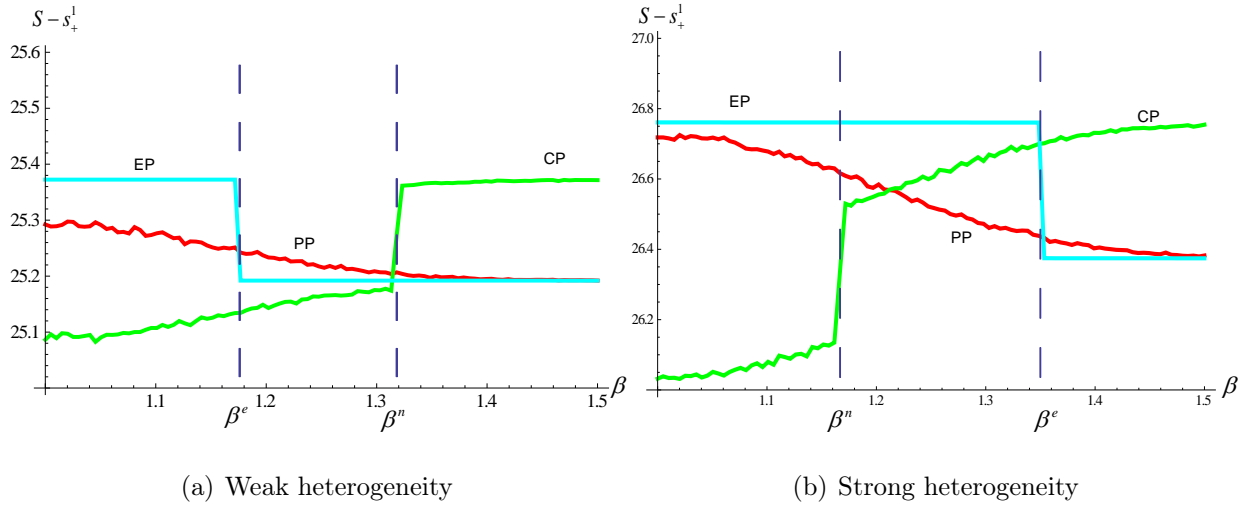(a) Weak heterogeneity   (b) Strong heterogeneity

Figure 2: Illustration of optimal rule

Consider now in more detail the impact of the relative-age advantage on the selection quality. In Figure 2, each panel depicts the induced total expected surplus due to training under each of the three relevant rules.[14] For the sake of illustration, we subtract $s_+^1 :=$ $\mathbb{E}\left[s_1^1 + s_2^1\right] = \frac{(2+\gamma)(1+\beta)}{2}$ from the toal expected surplus function, thereby eliminating the direct (positive) effect of a higher β in period 1 (i.e. before selection starts), such that only the indirect effect due to selection remains. Note first that in both panels (a) and (b), the respective function for rules EP and PP are decreasing in β throughout. This is intuitive, since a larger value of β increases the head-start advantage of (old) individual 1, thereby leading to more erroneous selection under configuration $\underline{A}$.

In contrast, under rule CP it is *increasing* in β. The intuition for this – somewhat counter-intuitive– result is that for $y^1 = C$ the probability of efficiently selecting individual 2 in period 1 (and hence also in period 2) under configuration $\underline{A}$ is the higher, the *larger* the head-start advantage of individual 1. Of course, this also makes erroneous selection in period 1 more likely under configuration $\overline{A}$. But in the range $\beta > \max(\beta^e, \beta^n)$ where rule CP is optimal, this has no detrimental effect on the expected total surplus, as individual 1 is nevertheless always selected in period 2 by definition of β^n.

---

[14]The parameter values used throughout for the case of weak and strong heterogeneity are $\gamma = 0.1$ and $\gamma = 0.2$, respectively. Moreover, for both cases we use $\delta = 0.85$, $x = 0.15$, $\sigma_1^2 = 0.1$ and $\lambda = 15$.

4 Relative Age Effect under the Optimal Rule

As discussed in the Introduction, one main issue in the empirical literature is whether initial advantages due to (relative) age tend to be perpetuated throughout the overall process of skill accumulation. In order to address this issue within our framework, we therefore distinguish between age-related skill differences which are due to a relative age *advantage* (β), and those arising as a result of subsequent selection decisions, referred to relative age *effect*.

As a measure for the relative-age effect, we use the expected difference in skills between an old and a young individual as induced between periods 1 (i.e. before training starts) and 3 (i.e. after training is over). Hence, we are again interested only in differences as induced by the selection decisions, and therefore subtract the (initial) advantage of the old individual in period 1 due to relative age, $s_-^1 := \mathbb{E}\left[s_1^1 - s_2^1\right] = \frac{(2+\gamma)(\beta-1)}{2} > 0$.[15] This leads to the following measure for the relative-age effect as induced under rules y^1 and y^2:

$$B^{y^1 y^2} = \mathbb{E}\left[s_1^3 - s_2^3\right] - s_-^1$$

Note first that full elimination of the relative-age effect (i.e. $B^{y^1 y^2} = 0$) is easily achieved either by rule RR (random selection) or EE (no selection at all), thereby on average neither enlarging nor reducing any skill differences due to relative-age advantage. However, as has been shown above, neither of these rules is optimal and therefore a relative age *effect* will arise under the optimal rule, and – through the induced selection decisions – the size of which does depend on the (initial) relative age *advantage*:

Proposition 5 *Assume that σ_2^2 is sufficiently small, and that λ is sufficiently large.*

(i) *When the relative-age advantage is weak so that rule EP is optimal, $B^{EP} > 0$.*

(ii) *When the relative-age advantage is medium so that rule PP is optimal, $B^{PP} > 0$ and increasing in β*

---

[15]In the empirical literature discussed above, the term s_-^1 is typically not subtracted (presumably due to lack of the respective information), and only the final outcomes are compared.

(iii) When the relative-age advantage is strong so that rule CP is optimal, $B^{CP}(\beta = \beta^n) < B^{PP}(\beta = \beta^n)$ might hold, and B^{CP} is decreasing in β
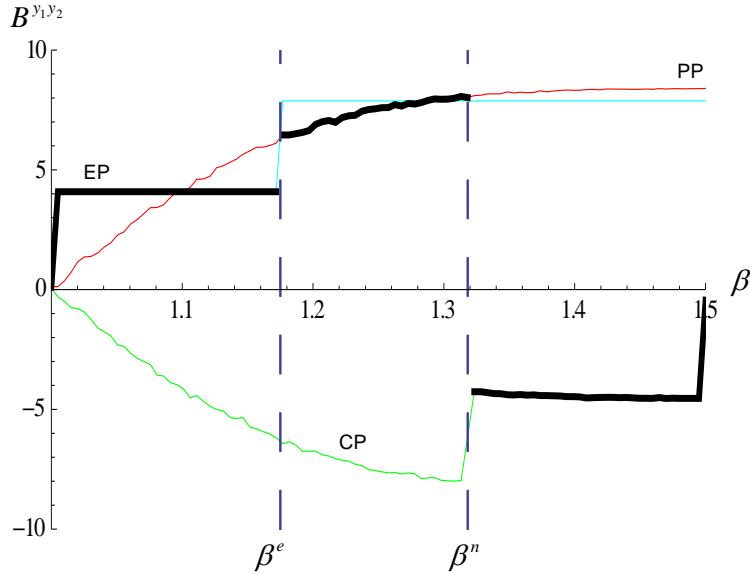

Figure 3: The size of the relative-age effect under the optimal rule

The proposition is illustrated in Figure 3 which depicts the relative-age effect for the relevant rules, using the same parametrization as in panel (a) of Figure 2. As already noted, the relative-age effect is always non-zero under the optimal rule. In particular, it is strictly positive when the relative-age advantage is either weak or medium such that rules EP and PP are optimal. Moreover, while almost constant as long as rule EP is optimal due to efficient selection, it is increasing in β under rule PP, because of the induced path-dependency which tends to favor individual 1.

In contrast, when the relative-age advantage becomes strong such that the optimal rule switches to CP (which occurs at $\beta = \beta^n$), then the relative-age effect typically exhibits a downward jump (and it can even become negative as in the case depicted in Figure 3) which is due to the fact that rule CP now heavily favors individual 2 in the first selection decision such that the relative age advantage of the old individual is overcompensated prior to the second selection stage.[16] Moreover, in the whole range

---

[16]It should be noted that $B^{CP} < 0$ for some $\beta > \beta^n$ does not hold for all permissible parameter constellations. In particular for σ_1^2 sufficiently large, we have $B^{CP} \approx B^{PP} > 0$

where rule CP is optimal, the relative-age effect is *decreasing* in β. Intuitively, a higher β increases the likelihood of selecting individual 2 in period 1, thereby further reducing the relative-age effect.

Overall, apart from the fact that it is not optimal to induce $B = 0$, also values $B \lesseqgtr 0$ are not necessarily a good indicator for the induced selection quality: to see this, recall from Proposition 4, that the induced selection quality is particularly high when either rules EP or rule CP are optimal, where $B^{EP} > 0$ always holds and where $B^{CP} < 0$ can hold. Hence, our results suggest that the mere empirical observation of a relative age effect is per se no indication of a suboptimal selection scheme.
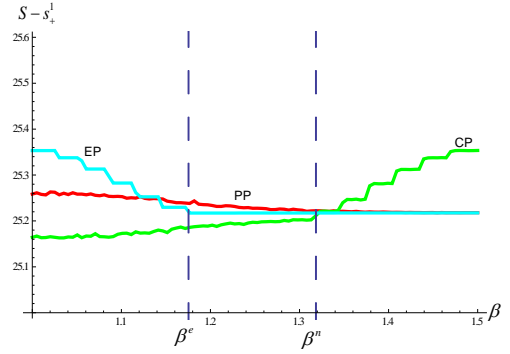
5 Robustness

In order to derive our previous results analytically, we have made several simplifying assumptions. The aim of this section is to show computationally that our main analytical results (see Proposition 4) are qualitatively robust, when several key assumptions are relaxed. In particular we will consider the following generalizations: (a) non-negligible noise of the second-stage skill signal ($\sigma_2^2 >> 0$); (b) more than two ability types; (c) cohort size larger than two; (d) more than two selection stages; (e) in addition to ability, the productivity of training also depends on current skills. In order to be able to distinguish between the implications of each of the considered generalizations, we consider each extension separately. Moreover, we confine attention to the (richer) case of weak heterogeneity (see Table 1) and, for the sake of comparison and where appropriate, also use the same parametrization as for panel (a) of Figure 2 and Figure 3.

As illustrated in Figure 4, our analytical results appear to be largely robust with respect to these extensions: In panel (a) we increase the level of noise in the second selection stage to $\sigma_2^2 = 0.03$, while keeping $\sigma_1^2 = 0.1$ so that our assumption that signals become less noisy over time is smaller remains satisfied. Panel (b) depicts a case with 7 ability types, where the range and the mean of the ability distribution are the same as in our basic model. In panel (c), we extend the cohort size to 10, where each individual has high ability with probability 0.5. Hence, under $y^t = P(C), t = 1, 2$ the five individuals
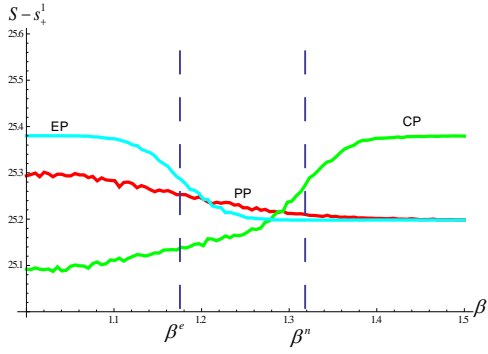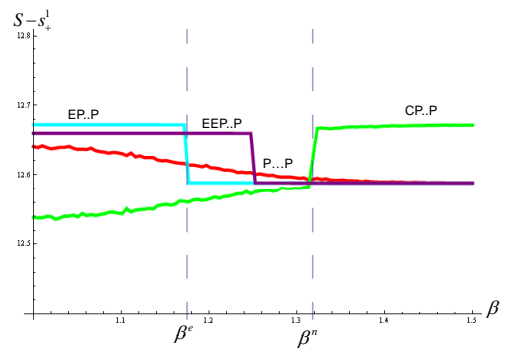
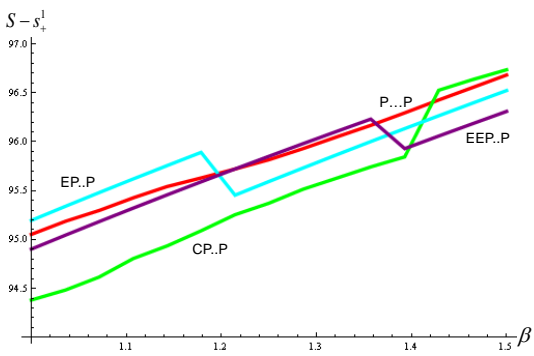(a) Noisier Signal in Stage 2 ($\sigma_2^2 = 0.03$)

(b) More Ability Types (7)
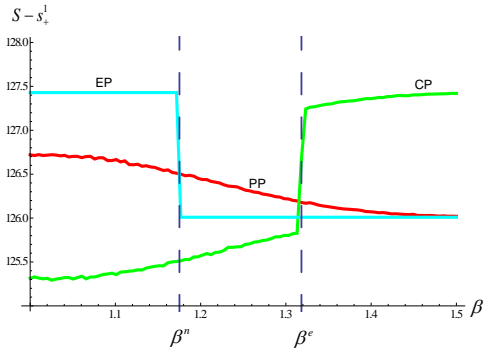
(c) Larger Cohort Size (10)

(d) Multiple Periods (8)

(e) Inter-temporal Complementarities

Figure 4: Robustness Checks

25

with the highest (lowest) signals are selected. For all three extensions we have verified numerically that the set of optimal rules remains $\{PP, EP, CP\}$ and, as observed in Figure 4 the transition between the optimal rules is very similar to the basic model, and then even the threshold values of β, where the transitions occur, differ only slightly.

Panel (d) depicts the case where the number of selection periods is increased to eight, and where equal weight $\lambda_t = 1$ is put on all periods $t = 1...8$. Again, the main qualitative features of the basic model are preserved. First, all eventual corrections occur at early stages of the selection process, and rule P is optimal for all, but at most the first two selection periods. This feature provides an additional motivation for assuming $\lambda > 1$ in the basic model, as it can be interpreted as a reduced form to capture all of these later selection periods where rule P is optimal. As for these early periods of selection, the transition pattern is also in accordance with the basic model, where rule E is optimal when the relative-age advantage is weak, while rule C is optimal when it is strong. The only difference occurs in the case of medium relative-age advantage, where the transition to rule P is delayed by one further period of equal training.

Last, but not least, as for extension where the skill formation exhibits inter-temporal complementarities in the spirit of Cunha and Heckman (2007) and Cunha, Heckman, and Schennach (2010), we consider the following CES-function of skill formation:

$$ s_{i,t+1} = \left(s_{i,t}^{\alpha} + (\lambda_t \delta_{i,t} a_i)^{\alpha} \right)^{1/\alpha}, \quad i = 1, 2, \ t = 1, .., T, $$

where $\alpha \in (\infty, 1]$ denotes the (constant) elasticity of substitution between past and current human capital investments (and where $\alpha = 1$ would lead back to our basic model). The now nonlinear nature of the production function again requires multiple periods of selection, and we consider six such periods where $\lambda_t = 3$ for all $t = 2, .., 6$. Panel (e) depicts the results for $\alpha = 0.6$ which is close to the value obtained by Cunha, Heckman, and Schennach (2010) in their estimations for cognitive skills (see p. 908, Table 1). Again, our main qualitative results appear robust also with respect to this extension, in particular the optimality of rule P in the later selection periods, and the transition from rule E to rule P and finally to rule C in the early ones as β increases.

6 Conclusion

This paper has characterized optimal selection rules in a stylized model of repeated selection in the presence of heterogeneity of individuals with respect to both ability and age within a cohort. Assuming that the objective is to maximize expected total skills at the end of the training process, it has been shown that deviations from a competitive rule – which selects the individuals with the higher skill signals for high intensity training – should only occur at the initial selection round(s). For low levels of relative age advantage, no selection should be made, whereas for high levels, the individuals with the lowest skill signals should be selected for high intensity training. It is noteworthy that this last result, which has a flavor of an equalizing policy, is obtained although (in)equality considerations are not assumed to be part of the objective of the planner. Moreover, the larger the degree of heterogeneity with respect to ability within the cohort, the larger the parameter range of relative age advantage for which counter-competitive selection is optimal.

One might be tempted to predict that a stronger initial relative age advantage (β), which distorts the link between ability and skill signals, should make the allocation of high-intensity training to high-ability individuals more difficult. However, it turns out that under the optimal selection rule the total expected skills attained during the training period are U-shaped with respect to the size of relative-age advantage (β), and therefore actually increasing in β when β is relatively large. Finally, our analysis also shows that the existence of a positive relative age effect is per se no indication for the use of a sub-optimal selection scheme for the allocation of training to individuals.

Although the robustness of our qualitative findings with respect to the relaxation of several assumptions has already been confirmed in this paper, several variations of the considered setup would be worth exploring. In particular, it would be interesting to analyze alternative objective functions, taking into account equality considerations or short term objectives of the teachers (or trainers) in charge of the training in the different periods. Also, assuming constant returns to training for each individual throughout the training process is quite strong, and it should be explored whether the qualitative insights remain when considering decreasing returns instead. Finally, although the simple 'cutoff-

type' selection rules considered in this paper are frequently used in the real world, from a theoretical perspective it would be interesting to explore the implications of allowing the planner to use more complex rules for allocating training intensities to individuals.

Appendix

Proof of Lemma 1 Part (i): Note that

$$\underline{\mu}_1 = \mathbb{P}[\beta + \eta_1^1 > (1+\gamma) + \eta_2^1] = \mathbb{P}[\Delta\eta_1 < \beta - (1+\gamma)],$$

$$\overline{\mu}_1 = \mathbb{P}[\beta(1+\gamma) + \eta_1^1 > 1 + \eta_2^1] = \mathbb{P}[\Delta\eta_1 < \beta(1+\gamma) - 1],$$

where $\Delta\eta_1 := \eta_2^1 - \eta_1^1$ is normally distributed with mean 0 and variance $2\sigma_1^2$. The first two claims follow directly from $\beta(1+\gamma) - 1 > \max[0, \beta - (1+\gamma)]$. The third claim is due to $\beta - (1+\gamma) + \beta(1+\gamma) - 1 = (\beta - 1)(2+\gamma) \geq 0$ (ii): Follows directly from the fact that for a given selection decision z in period 1 the difference in expected skills $\mathbb{E}[s_1^2 - s_2^2]$ is always higher in case $\overline{A}$ compared to $\underline{A}$. Part (iii) follows directly from the fact that for given abilities of both individuals the difference in expected skills $\mathbb{E}[s_1^2 - s_2^2]$ is increasing in the amount of training individual 1 receives in period 1. ∎

Proof of Proposition 1 As a first step we show that in the second selection period it is never optimal to choose either the R or the E rule. Furthermore, $y^2 = C$ can only be optimal if $y^1 = C$, i.e. if rule C has already chosen in the first selection period. To see this, note that

$$\mathbb{E}\left[(s_1^3 + s_2^3) - (s_1^2 + s_2^2)|\hat{s}_1^2, \hat{s}_2^2\right] = \begin{cases} \lambda(2+\gamma)\delta & y^2 = R, E \\[2mm] \lambda(2+\gamma)\delta + 2\lambda K & \\ \left(\mathbb{P}[\hat{s}_1^2 > \hat{s}_2^2|\overline{A}] - \mathbb{P}[\hat{s}_1^2 > \hat{s}_2^2|\underline{A}]\right) & y^2 = P \\[2mm] \lambda(2+\gamma)\delta - 2\lambda K & \\ \left(\mathbb{P}[\hat{s}_1^2 > \hat{s}_2^2|\overline{A}] - \mathbb{P}[\hat{s}_1^2 > \hat{s}_2^2|\underline{A}]\right) & y^2 = C. \end{cases}$$

This implies that choosing E or R is dominated either by P or by C. Furthermore, P is optimal in period 2, whenever

$$\mathbb{P}[\hat{s}_1^2 > \hat{s}_2^2|\overline{A}] - \mathbb{P}[\hat{s}_1^2 > \hat{s}_2^2|\underline{A}] \geq 0. \tag{12}$$

The exact expression for these two probabilities depends on the selection rule used in the first stage (y^1). In particular for the four possible choices in period 1 condition (12)

reduces to

$$\overline{\mu}_2^s + \overline{\mu}_2^n \;\geq\; \underline{\mu}_2^s + \underline{\mu}_2^n \qquad\qquad y^1 = R$$

$$\overline{\mu}_2^e \;\geq\; \underline{\mu}_2^e \qquad\qquad y^1 = E$$

$$\overline{\mu}_1\overline{\mu}_2^s + (1-\overline{\mu}_1)\overline{\mu}_2^n \;\geq\; \underline{\mu}_1\underline{\mu}_2^s + (1-\underline{\mu}_1)\underline{\mu}_2^n \qquad\qquad y^1 = P$$

$$(1-\overline{\mu}_1)\overline{\mu}_2^s + \overline{\mu}_1\overline{\mu}_2^n \;\geq\; (1-\underline{\mu}_1)\underline{\mu}_2^s + \underline{\mu}_1\underline{\mu}_2^n \qquad\qquad y^1 = C.$$

It follows directly from Lemma 1 that the condition is satisfied for $y^1 = R$ and $y^1 = E$ so that $y^2 = P$ is optimal in period 2 if E or R have been chosen in the first period. Considering the inequality for $y^1 = P$, we have

$$
\begin{aligned}
\overline{\mu}_2^s\overline{\mu}_1 + \overline{\mu}_2^n(1-\overline{\mu}_1) - \underline{\mu}_2^s\underline{\mu}_1 - \underline{\mu}_2^n(1-\underline{\mu}_1) & \qquad (13)\\
= \;\; (\overline{\mu}_2^s - \underline{\mu}_2^s) + (1-\overline{\mu}_1)\underbrace{(\overline{\mu}_2^n - \overline{\mu}_2^s)}_{<0} + (1-\underline{\mu}_1)\underbrace{(\underline{\mu}_2^s - \underline{\mu}_2^n)}_{>0} &\\
> \;\; (\overline{\mu}_2^s - \underline{\mu}_2^s) + (1-\underline{\mu}_1)(\overline{\mu}_2^n + \underline{\mu}_2^s - \underline{\mu}_2^n - \overline{\mu}_2^s) &\\
= \;\; \underline{\mu}_1\underbrace{(\overline{\mu}_2^s - \underline{\mu}_2^s)}_{>0} + (1-\underline{\mu}_1)\underbrace{(\overline{\mu}_2^n - \underline{\mu}_2^n)}_{>0} &\\
> \;\; 0, &
\end{aligned}
$$

which implies that P is also optimal in period 2. Finally, if C was chosen in period 1, then P is optimal in period 2 if

$$(1-\overline{\mu}_1)\overline{\mu}_2^s + \overline{\mu}_1\overline{\mu}_2^n - (1-\underline{\mu}_1)\underline{\mu}_2^s - \underline{\mu}_1\underline{\mu}_2^n \geq 0. \qquad (14)$$

Otherwise, CC yields higher total expected skills than CP. However, it can be shown that PP always dominates CC: To see this observe that expected total skills under PP and CC are given by

$$S^{PP} \;=\; Z + \frac{1}{2}\left[(\overline{\mu}_1 - \underline{\mu}_1)K + (\overline{\mu}_2^s\overline{\mu}_1 + \overline{\mu}_2^n(1-\overline{\mu}_1) - \underline{\mu}_2^s\underline{\mu}_1 - \underline{\mu}_2^n(1-\underline{\mu}_1))\lambda K\right] \quad (15)$$

$$S^{CC} \;=\; Z + \frac{1}{2}\left[-(\overline{\mu}_1 - \underline{\mu}_1)K - (\overline{\mu}_2^s(1-\overline{\mu}_1) + \overline{\mu}_2^n\overline{\mu}_1 - \underline{\mu}_2^s(1-\underline{\mu}_1) - \underline{\mu}_2^n\underline{\mu}_1)\lambda K\right] \quad (16)$$

Hence, we obtain

$$S^{PP} - S^{CC}$$
$$= \frac{1}{2}\left[(\overline{\mu}_1 - (1-\underline{\mu}_1))K\right]$$
$$+ \frac{1}{2}\left(\overline{\mu}_2^s\overline{\mu}_1 + \overline{\mu}_2^n(1-\overline{\mu}_1) - \underline{\mu}_2^s\underline{\mu}_1 - \underline{\mu}_2^n(1-\underline{\mu}_1) + \overline{\mu}_2^s(1-\overline{\mu}_1) + \overline{\mu}_2^n\overline{\mu}_1 - \underline{\mu}_2^s(1-\underline{\mu}_1) - \underline{\mu}_2^n\underline{\mu}_1\right)\lambda K$$
$$= \frac{1}{2}\left[2(\overline{\mu}_1 - \underline{\mu}_1)K + \left(\overline{\mu}_2^s + \overline{\mu}_2^n - \underline{\mu}_2^s - \underline{\mu}_2^n\right)\lambda K\right] > 0,$$

as both terms in the square bracket are positive (by parts (i) and (ii) of Lemma 1). Therefore, rule CC is never optimal, and hence any optimal rule stipulates $y^2 = P$.

To complete the proof, it remains to rule out that rule RP can be optimal. Calculating the expected total skills for both rules RP and CP gives

$$S^{RP} = Z + \frac{1}{2}\left[\frac{1}{2}(\overline{\mu}_2^s + \overline{\mu}_2^n - \underline{\mu}_2^s - \underline{\mu}_2^n)\lambda K\right] \tag{17}$$

$$S^{CP} = Z + \frac{1}{2}\left[-(\overline{\mu}_1 - \underline{\mu}_1)K + (\overline{\mu}_2^s(1-\overline{\mu}_1) + \overline{\mu}_2^n\overline{\mu}_1 - \underline{\mu}_2^s(1-\underline{\mu}_1) - \underline{\mu}_2^n\underline{\mu}_1)\lambda K\right] \tag{18}$$

Therefore, $S^{RP} = \frac{1}{2}(S^{PP} + S^{CP})$, which implies $S^{RP} \leq \max[S^{PP}, S^{CP}]$. ∎

Proof of Proposition 2 Taking the difference between (5) and (6) yields

$$S^{PP} - S^{CP} = \frac{1}{2}\left[2(\overline{\mu}_1 - \underline{\mu}_1) + \lambda(2\overline{\mu}_1 - 1)\cdot(\overline{\mu}_2^s - \overline{\mu}_2^n) + (1 - 2\underline{\mu}_1)\cdot(\underline{\mu}_2^s - \underline{\mu}_2^n)\right]K. \tag{19}$$

By part (i) of Lemma 1, the first term of the sum in the bracket, which corresponds to the skill increases in period 2, is positive, and therefore rule PP is superior to rule CP in period 1. The second term corresponding to skill gains in period 3 is positive if and only if (8) holds. Analogous reasoning applies to the comparison of rule EP (see (7)) with rules PP and CP. ∎

Proof of Lemma 2 As for part (i), note first that for $\sigma_2^2 \to 0$ the probability that individual 1 sends the higher signal in period 2 goes to 1 if $\mathbb{E}[s_1^2 - s_2^2] > 0$, and to 0 if

$\mathbb{E}[s_1^2 - s_2^2] < 0$. The Lemma then follows from:

$$E[s_1^2 - s_2^2] = \begin{cases} (1+\gamma)(\beta+\delta) - (1+\delta) > 0 & \text{if } \overline{A}, z = e \\[2mm] (1+\gamma)(\beta+\delta+x) - (1+\delta-x) > 0 & \text{if } \overline{A}, z = s \\[2mm] (\beta+\delta+x) - (1+\gamma)(1+\delta-x) > 0 & \text{if } \underline{A}, z = s \\[2mm] (\beta+\delta-x) - (1+\gamma)(1+\delta+x) < 0 & \text{if } \underline{A}, z = n, \end{cases}$$

where the inequalities in lines three and four follow from Assumptions 1 and 2.

Concerning part (ii), we have for $\underline{A}$ and $z = e$

$$E[s_1^2 - s_2^2] = (\beta+\delta) - (1+\gamma)(1+\delta),$$

which is positive if and only if $\beta > \beta^e$. Moreover, we have $\frac{\partial \beta^e}{\partial \gamma} = 1+\delta > 0$.

Similarly, as for part (iii) we get for $\overline{A}$ and $z = n$

$$E[s_1^2 - s_2^2] = (1+\gamma)(\beta+\delta-x) - (1+\delta+x)$$

which is positive if and only if $\beta > \beta^n$. Moreover, we have $\frac{\partial \beta^n}{\partial \gamma} = \frac{-1-\delta-x}{(1+\gamma)^2} < 0$. ∎

Proof of Proposition 3 As for part (i), using Lemma 2 and the fact that for $\beta = 1$ we have $\lim_{\sigma_2^2 \to 0} \underline{\mu}_2^e = \lim_{\sigma_2^2 \to 0} \overline{\mu}_2^n = 0$ we obtain that for small σ_2^2

$$\begin{aligned} S^{EP} - S^{PP} &\approx \frac{1}{2}\left[0 + \lambda(1-0)\right]K - \frac{1}{2}\left[(\overline{\mu}_1 - \underline{\mu}_1) + \lambda(\overline{\mu}_1 + 0 - \underline{\mu}_1 - 0)\right]K \\ &= \frac{1}{2}\left[\lambda - (1+\lambda)(\overline{\mu}_1 - \underline{\mu}_1)\right]K. \end{aligned}$$

Define $\lambda^* = \frac{\overline{\mu}_1 - \underline{\mu}_1}{1 - (\overline{\mu}_1 - \underline{\mu}_1)}$. It is now obvious that for $\lambda < \lambda^*$ PP dominates EP for sufficiently small σ_2^2, whereas the opposite statement holds for $\lambda > \lambda^*$.

As for part (ii), taking into account that $\overline{\mu}_1 - \underline{\mu}_1 = 2\mathbb{P}(\Delta\eta_1 \leq \gamma) - 1$ and that $\frac{\Delta\eta_1}{\sqrt{2\sigma_1^2}}$ has a Gaussian distribution with mean zero and variance one, it is easy to see that $\lambda^* \geq 1$ is equivalent to

$$2\Phi\left(\frac{\gamma}{\sqrt{2\sigma_1^2}}\right) \geq \frac{1}{2} + 1.$$

Part (ii) of the Proposition follows directly.

As for part (iii), comparing PP with CP for $\beta = 1$ and small σ_2^2 we get as above

$$S^{PP} - S^{CP} \approx (1+\lambda)(\overline{\mu}_1 - \underline{\mu}_1)K > 0 \quad \forall \lambda \geq 1,$$

and therefore rule CP is not optimal. ∎

Proof of Proposition 4 To prove the three claims of this Proposition, we first consider how the signs of the differences $S^{PP} - S^{CP}$, $S^{PP} - S^{EP}$ and $S^{EP} - S^{CP}$ depend on β.

Comparing the expected total skills under rules PP and CP we obtain from (5) and (6)

$$S^{PP} - S^{CP} \;=\; (\overline{\mu}_1 - \underline{\mu}_1)K + \frac{\lambda}{2}\left((2\overline{\mu}_1 - 1)(\overline{\mu}_2^s - \overline{\mu}_2^n) + (1 - 2\underline{\mu}_1)(\underline{\mu}_2^s - \underline{\mu}_2^n)\right)K \quad (20)$$

Expression (20) is positive if the bracket in the second term is positive. Conversely, if the bracket is negative, then (20) is negative for sufficiently large λ. Using Lemma 2 this bracket can be approximated for small σ_2^2 by

$$(2\overline{\mu}_1 - 1)(1 - \overline{\mu}_2^n) + (1 - 2\underline{\mu}_1).$$

It is straightforward to see that for $\beta > \beta^e$ we have $\underline{\mu}_1 > \frac{1}{2}$ and therefore $1 - 2\underline{\mu}_1 < 0$. Taking further into account that $\overline{\mu}_1 > \underline{\mu}_1$ (part (i) of Lemma 1) it follows directly from part (iii) of Lemma 2 that this expression reduces to $2(\overline{\mu}_1 - \underline{\mu}_1) > 0$ for $\beta < \beta^n$ and to $1 - 2\underline{\mu}_1 < 0$ relative-age advantage is strong ($\beta > \beta^n$). This implies that CP dominates PP relative-age advantage is strong and λ is large. Furthermore, if (11) holds, and therefore $\beta^n > \beta^e$, both for medium relative-age advantage ($\beta \in (\beta^e, \beta^n)$) and for weak relative-age advantage ($\beta < \beta^e$) rule PP dominates CP.

Comparing rules rules PP and EP we obtain from (5) and (7)

$$S^{PP} - S^{EP} \;=\; \frac{1}{2}(\overline{\mu}_1 - \underline{\mu}_1)K + \frac{\lambda}{2}\left(\overline{\mu}_2^s\overline{\mu}_1 + \overline{\mu}_2^n(1 - \overline{\mu}_1) - \underline{\mu}_2^s\underline{\mu}_1 - \underline{\mu}_2^n(1 - \underline{\mu}_1) - \overline{\mu}_2^e + \underline{\mu}_2^e\right)K.$$

Using again Lemma 2 we approximate the bracket in the second term by

$$\overline{\mu}_1 + \overline{\mu}_2^n(1 - \overline{\mu}_1) - \underline{\mu}_1 - 1 + \underline{\mu}_2^e. \quad (21)$$

For $\beta > \beta^e$ it follows from part (ii) of Lemma 2 that this expression is equal to $\overline{\mu}_1 - \underline{\mu}_1 + \overline{\mu}_2^n(1 - \overline{\mu}_1) > 0$ and therefore $S^{PP} > S^{EP}$ for all λ. Together with the discussion above this shows that PP is optimal if the relative-age advantage is medium, whereas CP is optimal for a strong relative-age advantage which proves parts (ii) and (iii) of the Proposition.

To prove part (i) we consider note first that it follows from above that if the relative-age advantage is weak, then EP dominates PP for large λ. Comparing EP and CP we

obtain

$$S^{EP} - S^{CP} \;=\; \frac{1}{4}(\overline{\mu}_1 - \underline{\mu}_1)K + \frac{\lambda}{4}\left(\overline{\mu}_2^e - \underline{\mu}_2^e - \overline{\mu}_2^s(1 - \overline{\mu}_1) - \overline{\mu}_2^n\overline{\mu}_1 + \underline{\mu}_2^s(1 - \underline{\mu}_1) + \underline{\mu}_2^n\underline{\mu}_1\right)K.$$

For the case of weak relative-age advantage the bracket in the second term can under consideration of Lemma 2 be approximated for small σ_2^2 by $(1 + \overline{\mu}_1(1 - \overline{\mu}_2^n) - \underline{\mu}_1) > 0$. Hence, rule EP is optimal if the relative-age advantage is weak and λ is large. ∎

Proof of Proposition 5 Direct calculations show that for $\sigma_2^2 \to 0$ and $\beta < \beta^e$ the relative age effect under the EP rule converges to

$$\bar{B}^{EP} = \frac{\lambda(2 + \gamma)x}{2} > 0,$$

which proves part (i).

In oder to express the value of B^{PP} and B^{CP} we have to extend our notation for the probability that the old individual is selected in stage $t = 1, 2$ to the scenarios $\bar{\bar{A}} := \{a_1 = a_2 = \bar{a}\}$ and $\underline{\underline{A}} := \{a_1 = a_2 = \underline{a}\}$. Obviously, for any $\beta > 1$ we have

$$\bar{\bar{\mu}}_1 > \underline{\underline{\mu}}_1 > \frac{1}{2}, \lim_{\sigma_2^2 \to 0} \bar{\bar{\mu}}_2^s = \underline{\underline{\mu}}_2^s = 1$$

and for $\beta < \beta^n$

$$\lim_{\sigma_2^2 \to 0} \bar{\bar{\mu}}_2^n = \lim_{\sigma_2^2 \to 0} \underline{\underline{\mu}}_2^n = 0.$$

Using this notation we obtain that for $\beta < \beta^n$ the relative age effect under the PP rule converges to

$$\bar{B}^{PP} = \left((2 + \gamma)\frac{\underline{\mu}_1 + \overline{\mu}_1 - 1 + \bar{\bar{\mu}}_1}{2}x - (\bar{\bar{\mu}}_1 - \underline{\underline{\mu}}_1)x\right) + \lambda\left[(2 + \gamma)\frac{\overline{\mu}_1 + \underline{\mu}_1 + 2(\bar{\bar{\mu}}_1 - 1)}{4}x - (\bar{\bar{\mu}}_1 - \underline{\underline{\mu}}_1)x\right].$$

Taking into account that each of the expressions $\overline{\mu}_1, \underline{\mu}_1, \bar{\bar{\mu}}_1$ and $\underline{\underline{\mu}}_1$ is increasing in β and they all have positive coefficients in the square bracket term, which is multiplied by λ, it is clear that $\bar{B}^{PP}$ is increasing in β for sufficiently large λ. Furthermore, it should be noted that for any probability $\bar{\mu}_2^z, z \in \{s, n, r, e\}$ with $\lim_{\sigma_2^2 \to 0} \mu_2^z = 0$ or $\lim_{\sigma_2^2 \to 0} \mu_2^z = 1$ we have $\lim_{\sigma_2^2 \to 0} \frac{\partial \mu_2^z}{\partial \beta} = 0$. Together, this implies that B^{PP} is increasing in β.

Concerning the sign of $\bar{B}^{PP}$ we obtain for the second term in the sum

$$\left[(2+\gamma)\frac{\overline{\mu}_1 + \underline{\mu}_1 + 2(\bar{\bar{\mu}}_1 - 1)}{4}x - (\bar{\bar{\mu}}_1 - \underline{\underline{\mu}}_1)x\right]$$

$$= \frac{x}{4}\left[(2+\gamma)\left(\overline{\mu}_1 + \underline{\mu}_1\right) + (1+\gamma)2(\bar{\bar{\mu}}_1 - 1) + 2(\bar{\bar{\mu}}_1 - 1) - 4(\bar{\bar{\mu}}_1 - \underline{\underline{\mu}}_1)\right]$$

$$= \frac{x}{4}\left[\gamma\left(\overline{\mu}_1 + \underline{\mu}_1\right) + 2\left(\overline{\mu}_1 + \underline{\mu}_1 - \bar{\bar{\mu}}_1\right) + (1+\gamma)2(\bar{\bar{\mu}}_1 - 1) + \left(4\underline{\underline{\mu}}_1 - 2\right)\right]$$

$$\geq 0$$

Therefore the coefficient of λ in $\bar{B}^{PP}$ is positive and it is easy to see that the first term is even larger than this expression. Hence, $B^{PP} > 0$, which proves (ii).

As for part (iii), we obtain that for $\beta > \beta^n$ the relative age effect under rule CP converges to

$$\bar{B}^{CP} = \left((2+\gamma)\frac{2 - \underline{\mu}_1 - \overline{\mu}_1 - 2\underline{\underline{\mu}}_1}{2}x - (1+\gamma)(\bar{\bar{\mu}}_1 - \underline{\underline{\mu}}_1)x\right)$$
$$+ \lambda\frac{x}{2}\left[(2+\gamma)(2 - \underline{\mu}_1) - 2\underline{\underline{\mu}}_1 - (1+\gamma)\bar{\bar{\mu}}_1\right].$$

The observation that this expression can be smaller than $\bar{B}^{PP}$ and even negative for $\beta = \beta^n$ can be made from figure 3, where an example of such a situation is depicted. Concerning monotonicity, we observe that all three summands in the square bracket multiplied with λ are decreasing with respect to β. Hence $\bar{B}^{CP}$ is decreasing in β for sufficiently large λ and by the same arguments as used in the proof of part (ii) this implies that also B^{CP} in decreasing with respect to β. ∎

References

ALLEN, J., AND R. BARNSLEY (1993): "Streams and tiers: The interaction of ability, maturity, and training in systems with age-dependent recursive selection," *Journal of Human Resources*, 28(3), 649–659.

AMANN, E., AND W. LEININGER (1996): "Asymmetric All-Pay Auctions with Incomplete Information: The Two-Player Case," *Games and Economic Behavior*, 14, 1–18.

BARNSLEY, R., AND A. THOMPSON (1988): "Birthdate and success in minor hockey: The key to the NHL," *Canadian Journal of Behavioural Science*, 20(2), 167–176.

BEDARD, K., AND E. DHUEY (2006): "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects," *Quarterly Journal of Economics*, 121(4), 1437–1472.

BILLARI, F., AND M. PELLIZZARI (2008): "The younger, the better? Relative age effects at university," *IZA Discussion Paper No. 3795*.

COBLEY, S., J. SCHORER, AND J. BAKER (2008): "Relative age effects in professional German soccer: A historical analysis," *Journal of Sports Sciences*, 26(14), 1531–1538.

COLVIN, G. (2008): *Talent is Overrated*. Fortune.

COYLE, D. (2009): *The Talent Code: Greatness Isn't Born. It's Grown. Here's How.* Bantam.

CUNHA, F., AND J. HECKMAN (2007): "The technology of skill formation," *American Economic Review*, pp. 31–47.

CUNHA, F., J. HECKMAN, L. LOCHNER, AND D. MASTEROV (2006): "Interpreting the evidence on life cycle skill formation," *Handbook of the Economics of Education*, 1, 697–812.

CUNHA, F., J. HECKMAN, AND S. SCHENNACH (2010): "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78(3), 883–931.

DUBNER, S., AND S. LEVITT (2006): "A Star is Made," *New York Times*, May 6, 2006.

DUDINK, A. (1994): "Birth date and sporting success.," *Nature*, 368, 592.

ELDER, T., AND D. LUBOTSKY (2009): "Kindergarten entrance age and children's achievement," *Journal of Human Resources*, 44(3), 641–683.

ERICSSON, K., R. KRAMPE, AND C. TESCH-RÖMER (1993): "The role of deliberate practice in the acquisition of expert performance.," *Psychological Review*, 100(3), 363–406.

ERICSSON, K. E. A. (2006): "The influence of experience and deliberate practice on the development of superior expert performance," *The Cambridge Handbook of Expertise and Expert Performance*, pp. 683–703.

FEESS, E., G. MUEHLHEUSSER, AND M. WALZL (2008): "Unfair Contests," *Journal of Economics*, 93(3), 267–291.

FREDRIKSSON, P., AND B. ÖCKERT (2005): "Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labour Market Performance," *IZA Working Paper 1659*.

GLADWELL, M. (2008): *Outliers: The story of success*. Little, Brown and Co.

HELSEN, W., J. STARKES, AND J. VAN WINCKEL (1998): "The influence of relative age on success and dropout in male soccer players," *American Journal of Human Biology*, 10(6), 791–798.

HELSEN, W., J. VAN WINCKEL, AND A. WILLIAMS (2005): "The relative age effect in youth soccer across Europe," *Journal of Sports Sciences*, 23(6), 629–636.

JÜRGES, H., AND K. SCHNEIDER (2011): "Why Young Boys Stumble: Early Tracking, Age and Gender Bias in the German School System," *German Economic Review*, 12, 371–394.

KRISHNA, V., AND J. MORGAN (1997): "An Analysis of the War of Attrition and the All-Pay Auction," *Journal of Economic Theory*, 72, 343–362.

LAZEAR, E., AND S. ROSEN (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), 841–864.

McEWAN, P., AND J. SHAPIRO (2008): "The Benefits of Delayed Primary School Enrollment," *Journal of Human Resources*, 43(1), 1–29.

MOLDOVANU, B., AND A. SELA (2006): "Contest Architecture," *Journal of Economic Theory*, 126(1), 70–97.

MÜHLENWEG, A., AND P. PUHANI (2010): "The evolution of the school-entry age effect in a school tracking system," *Journal of Human Resources*, 45(2), 407–438.

PUHANI, P., AND A. WEBER (2008): "Does the early bird catch the worm? Instrumental variable estimates of early educational effects of school entry in Germany," *The Economics and Training of Education*, pp. 105–132.

SAMPAIO, B., R. DA MATTA, R. RIBAS, AND G. SAMPAIO (2011): "The effect of age on college entrance test score and enrollment: A regression-discontinuity approach," *SSRN Working Paper No. 1471686*.

SPRIETSMA, M. (2010): "The effect of relative age in the first grade of primary school on long-term scholastic results: International comparative evidence using PISA 2003," *Education Economics*, 18(1), 1–32.

TILL, K., S. COBLEY, N. WATTIE, J. O'HARA, C. COOKE, AND C. CHAPMAN (2010): "The prevalence, influential factors and mechanisms of relative age effects in UK Rugby League," *Scandinavian Journal of Medicine & Science in Sports*, 20(2), 320–329.

TULLOCK, G. (1980): "Efficient Rent Seeking," in *Towards a Theory of the Rent-Seeking Society*, ed. by J. Buchanan. Texas A&M University Press, Collge Station.

WATTIE, N., J. BAKER, S. COBLEY, AND W. MONTELPARE (2007): "Tracking relative age effects over time in Canadian NHL players," *International Journal of Sport Psychology*, 38, 1–9.