



## Family background, self-confidence and economic outcomes

Antonio Filippin<sup>a,b,\*</sup>, Marco Paccagnella<sup>c,d,\*\*</sup>

<sup>a</sup> University of Milan, Department of Economics, Via Conservatorio 7, 20122 Milano, Italy

<sup>b</sup> Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany

<sup>c</sup> Bank of Italy, Trento Branch, Piazza A. Vittoria 6, 38122 Trento, Italy

<sup>d</sup> Bocconi University, Department of Economics, Via Röntgen 1, 20136 Milano, Italy

### ARTICLE INFO

#### Article history:

Received 13 December 2011

Received in revised form 4 June 2012

Accepted 11 June 2012

#### JEL classification:

I24

D83

J24

J62

#### Keywords:

Human capital

Educational economics

Socio-economic status

### ABSTRACT

In this paper we analyze the role played by self-confidence, modeled as beliefs about one's ability, in shaping task choices. We propose a model in which fully rational agents exploit all the available information to update their beliefs using Bayes' rule, eventually learning their true type. We show that when the learning process does not converge quickly to the true ability level, small differences in initial confidence can result in diverging patterns of human capital accumulation between otherwise identical individuals. If differences in self-confidence are correlated with socio-economic background (as a large body of empirical literature suggests), self-confidence can be a channel through which education and earning inequalities perpetuate across generations. Our theory suggests that cognitive tests should take place as early as possible, in order to avoid that systematic differences in self-confidence among equally talented people lead to the emergence of gaps in the accumulation of human capital.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Gaps in economic outcomes (like educational attainments and earnings) tend to persist across generations, and it is well-known that parents' socio-economic status is usually a good predictor of the outcomes of their offspring.

Bowles, Gintis, and Osborne (2001) stress that “the advantages of the children of successful parents go considerably beyond the benefits of superior education, the inheritance of wealth, or the genetic inheritance of cognitive ability.” They argue that the inclusion of variables comparable to what now goes under the label of

“non-cognitive skills” can help explaining a larger fraction of the variance of earnings.<sup>1</sup> They also claim that the contribution of parental socio-economic status to their offspring's earnings is partly determined by such non-cognitive skills, genetically transmitted or learned from parents that act as role models.

Since then many other authors have emphasized the role played by non-cognitive skills; the most recent literature tends to treat them as inputs that enter the “black-box” of the skill production function. Cunha and Heckman (2007) propose a particular formulation of the technology of skill formation featuring self-productivity

\* Corresponding author at: University of Milan, Department of Economics, Via Conservatorio 7, 20122 Milano, Italy. Tel.: +39 02 503 21536; fax: +39 02 5032 1505.

\*\* Corresponding author at: Bank of Italy, Trento Branch, Piazza A. Vittoria 6, 38122 Trento, Italy. Tel.: +39 0461 212 232.

E-mail addresses: [antonio.filippin@unimi.it](mailto:antonio.filippin@unimi.it) (A. Filippin), [marco.paccagnella@bancaditalia.it](mailto:marco.paccagnella@bancaditalia.it) (M. Paccagnella).

<sup>1</sup> Brunello and Schlotter (2011) define non-cognitive skills as “personal-ity traits that are weakly correlated with measures of intelligence.” While in some papers the orthogonality between the two types of skills derives from statistical conditions that ensure identification in classical factor analysis, Deke and Haimson (2006) do find that key personality traits are rather poorly associated both with cognitive skills and among themselves, while the correlation among measures of cognitive skills is much stronger.

and dynamic complementarity among a multidimensional vector of cognitive and non-cognitive skills. They argue that insufficient investment in some of these skills early in life has long-lasting consequences that are very difficult or costly to revert. Heckman, Stixrud, and Urzua (2006), Cunha and Heckman (2007, 2008) and Cunha, Heckman, and Schennach (2010) have shown that gaps between children from different backgrounds open up very early in life, as soon as in pre-school age, and then tend to persist and stay roughly constant over the lifetime. This finding clearly locates the rising of the problem in a period in which the role of the parents is the most important.

In this paper we want to analyze the role possibly played by a single non-cognitive skill, namely self-confidence, that we define as the beliefs over one's unknown level of cognitive ability. Hence, our model entails the simplest possible multidimensional vector of skills, containing only two elements: a cognitive skill (innate ability) and a non-cognitive one (self-confidence). The use of such a framework is neither meant to deny the importance of other skills, nor to ignore the well-established fact that cognitive and non-cognitive abilities are multidimensional in nature, nor to downplay the interaction among them. It simply reflects our goal to isolate and highlight a very specific mechanism, i.e. the role that self-confidence plays by distorting task choices. In other words, our purpose is to go into the "black-box" of the skill production function, identifying a precise and specific channel through which inherited differences in self-confidence can endogenously (i.e. through individual choices) explain the emergence and persistence of gaps in the accumulation of human capital. Heckman et al. (2006) precisely quantify the role played by non-cognitive skills. In their empirical model, they assume that a single latent factor captures cognitive ability, and that another single latent factor captures non-cognitive ability.<sup>2</sup> Their empirical measures of non-cognitive skills are scores in the Rotter Locus of Control Scale and in the Rosenberg Self-Esteem Scale, which are both clearly linked to the concept of self-confidence.<sup>3</sup> They show that, even though cognitive ability explains a larger share of wage variance, non-cognitive ability is similarly important, since moving persons from the top to the bottom of the distribution has similar effects for both types of skills.

The working idea of our model is that, by acting as role models, parents transmit to their children beliefs about their (unknown) ability. Such beliefs affect educational choices and, through this channel, contribute to widen the gap in human capital accumulation while the learning process (of actual ability) proceeds. The consequences of

initially "wrong" beliefs can thus have long lasting effects, even if agents eventually learn their true level of ability.

For self-confidence to have important effects we do not need to assume that agents enjoy holding a good image of themselves (i.e. that self-confidence enters directly the utility function), something that would imply that some degree of overconfidence is optimal.<sup>4</sup> Our theoretical framework assumes full rationality, given that agents extract all the available information from the signals received in order to update their beliefs, and this implies that they eventually learn their true type. Similarly, we exclude any other form of self-deception. The Bayesian learning mechanism is based on observing success or failure in the endeavor undertaken, given that the probability of success depends on the true level of ability as well as on the difficulty of the task, which is chosen endogenously in accordance with (updated) beliefs about one's ability.

We finally simulate the model with a bootstrapping procedure, showing that choices distorted by underconfidence (while all the other sources of heterogeneity are neutralized) lead to a significant gap in the accumulation of human capital during the learning process of the true level of ability. As long as it correlates with the family background, self-confidence constitutes therefore a channel through which gaps in educational attainments and earnings perpetuate across generations. This finding also helps to explain why the early gaps based on the socio-economic background do not narrow when the role of the family becomes less important relative to other factors (like school and teacher quality, or peers' characteristics). This also suggests that policies aimed at providing early and accurate feedbacks on the cognitive skills of disadvantaged children can be important in promoting inter-generational income mobility.

The outline of the paper is as follows. In Section 2 we compare our theoretical approach with others in the literature. We also provide evidence supporting both the importance of the role played by self-confidence and the correlation between self-confidence and family background. In Section 3 we present a simple and parsimonious theoretical model that highlights how self-confidence can affect the accumulation of human capital via task choice. In Section 4 we run a simulation of our model in order to better assess its implications in terms of the emergence of gaps in educational attainments between people from different backgrounds. Section 5 comments upon the results and draws some conclusions.

## 2. Motivation

In this section we survey the related theoretical and empirical literature to motivate the relevance of our work. In Section 2.1 we document how incorrect beliefs about ability are indeed very common, and we discuss the possible consequences of this fact. Moreover, we motivate our choice to focus on a definition based on the *levels* of beliefs,

<sup>2</sup> Heckman et al. (2006) also assume independence between latent cognitive and non-cognitive abilities. Such assumptions have then been relaxed in Cunha and Heckman (2008) and Cunha et al. (2010), who use a wider range of psychological measures, mainly related to social behavior and behavioral problems.

<sup>3</sup> In particular, self-esteem reflects a person's overall evaluation or appraisal of his or her own worth, encompassing beliefs and emotions. Locus of control is instead defined as the extent to which individuals believe that they can control events that affect them: people with an external locus of control tend to think that luck or fate (rather than their own actions) are the main determinants of success or failure, and this is likely to be related to low levels of confidence in own ability.

<sup>4</sup> Such an assumption is quite common in the behavioral economics literature (e.g. in Köszegi, 2006 and Weinberg, 2009). We discuss this issue in more details in Section 2.2.

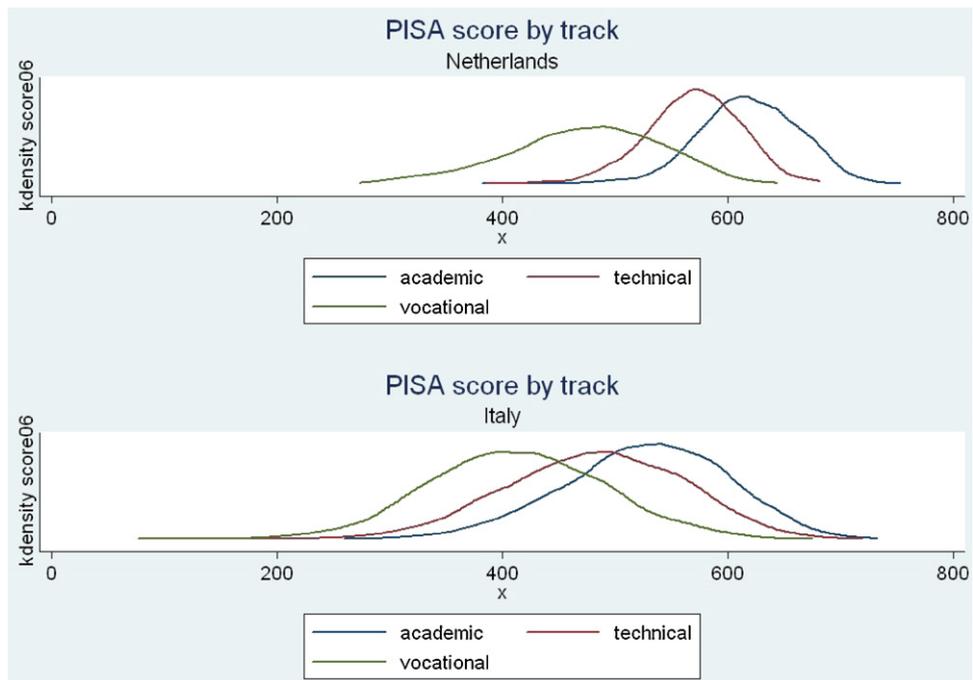


Fig. 1. Netherlands (top), Italy (bottom).

rather than on their *precision*. In Section 2.2 we discuss an assumption commonly made in the literature, i.e. that people actually care about their beliefs, and explain why we decided not to make this assumption. Finally, in Section 2.3 we justify an important implicit assumption of our model by providing evidence that suggests the existence of a relevant link between self-confidence and the socio-economic background.

### 2.1. Imperfect knowledge of one's ability

We define self-confidence as the beliefs an agent holds about his own ability, following Bénabou and Tirole (2002), Hvide (2002), Köszegi (2006), Sjögren and Sällström (2004), and Weinberg (2009) among others. This definition is based on the assumption that ability is unknown to the agent instead of being his private information as in standard signaling models.<sup>5</sup>

There is an extensive literature showing that agents hold a rough estimate of their cognitive skills. Dunning, Heath, and Suls (2004) survey the psychological literature documenting the presence of a weak correlation between actual and perceived performance in several domains, while Falk, Huffman, and Sunde (2006) provide experimental evidence that people are substantially uncertain about their relative ability and that this has indeed important consequences on search decisions.

In Fig. 1 we plot the distribution of PISA 2006 test scores across pupils enrolled in different high-school tracks. Comparing the top and the bottom panel (which refer to the Netherlands and Italy, respectively) we see quite a different degree of overlapping in the two countries. The main difference between the two educational systems is that in Italy students and parents are perfectly free to choose the high-school track, while in the Netherlands there is a nationwide standardized aptitude test,<sup>6</sup> administered at age 12, that has gained a considerable influence in recommending the secondary track most suitable for the pupil. While we recognize that there might be many other reasons to choose different tracks, the level of ability should also play an important role. Fig. 1 shows that the degree of overlapping across tracks, which should in part be driven by ability mismatch, is much more pronounced in Italy, where students receive less precise signals about their ability.

Another important issue is the definition of confidence in terms of the mean vs. the spread of the distribution of beliefs.<sup>7</sup> The former implies that an overconfident holds too high an estimate of his ability. The latter refers to an evaluation that is too precise, and better fits a situation in which an investor underestimates the variance of future returns. As long as the true level of ability is a point estimate, as in our case, it would be meaningless to talk about over- or underconfidence in these terms.

<sup>5</sup> Along this dimension, our approach differs from that of Keane and Wolpin (1997), who propose a structural model in which individual ability in the initial period is assumed to be exogenously given and known by the agents.

<sup>6</sup> The so-called Cito test.

<sup>7</sup> Both definitions are used in the literature. The first by Hvide (2002), Bénabou and Tirole (2002), and Weinberg (2009), among others; the second by Sjögren and Sällström (2004), Köszegi (2006) and Belzil (2007) use both.

The focus on the level rather than on the precision of beliefs is one of the main difference between our model and the one proposed by Sjögren and Sällström (2004). A second major difference is that, in our framework, agents eventually discover their true type, while in Sjögren and Sällström (2004) agents can remain “trapped” with wrong beliefs due to insufficient experimentation and learning.<sup>8</sup>

Imperfect information about ability is conceptually very similar to imperfect information about the market returns to education due to incomplete information.<sup>9</sup> This aspect has been recently investigated by Nguyen (2008), Attanasio and Kaufmann (2009), Kaufmann (2010) and Jensen (2010), with a particular focus on developing countries. However, a model in which agents have biased information on expected returns correlated with family background, would be observationally equivalent to our model only as far as initial decisions are concerned. By making educational choices agents then receive many signals about their ability, while the same does not happen about returns to education. Therefore, the evolution of beliefs (and, as a consequence, the accumulation of human capital) would likely be very different.

Our work is similar in spirit to a recent paper by Trachter (2011), who also assumes imperfect knowledge of academic ability (and not of returns to education) and estimates the option value contained in the choice of enrolling in academic 2-year colleges, often used as stepping stones toward 4-year colleges by students with initially low academic expectations.

## 2.2. Self-confidence in the utility function

In our paper we focus on the role that confidence plays through task choices and on how this channel shapes human capital acquisition. Self-confidence has no direct influence, as it would be if agents derived utility from holding a positive self-image (Köszegi, 2006; Weinberg, 2009). While such models rationalize many interesting features of human behavior and show that moderate levels of overconfidence turn out to be optimal, we decide to stick to a simpler theoretical framework in which this does not happen. The main reason is that once agents are supposed to enjoy holding a good self-image, they should also be capable of tailoring the information acquisition during their learning process in such a way to preserve it, for instance by means of beliefs that are “pragmatic” (Hvide, 2002) or, more generally, self-serving, as well as with selective memory (Bénabou & Tirole, 2002).<sup>10</sup> Manipulating the information acquisition can only be effective *in the short run*, with the only transitory effect of slowing

down the learning process, unless agents end up stuck in a self-confirming equilibrium in which the learning process reaches a fixed point.<sup>11</sup> Although such an outcome cannot be excluded, we find more interesting to analyze the effect of holding a wrong self-image when agents eventually discover their true type. We therefore prefer to adhere to a framework in which agents are characterized by standard preferences and unbiasedly exploit all the available information.

## 2.3. The inter-generational transmission of confidence

Recent empirical findings provide support for one of the key assumptions of our model, namely that self-confidence is correlated with family background. Cesarini, Johannesson, Lichtenstein, and Wallace (2009), using Swedish data on a sample of twins and defining overconfidence as the difference between the perceived and actual rank in cognitive ability, argue that genetic differences explain 16–34% of the variation in overconfidence, and that common environmental differences explain an additional 5–11%. A series of studies on different longitudinal UK datasets (collected in Goodman & Gregg, 2010) find a strong intergenerational correlation not only in cognitive skills, but also in a variety of attitudes that can be considered proxies of confidence. In particular, Gregg and Washbrook (2011) find that, even controlling for family background and prior attainment, children are more likely to perform well in tests at age 11 if they have strong beliefs in their own ability and a more internal locus of control, and that children from poorer families are less likely to possess these attributes. Chowdry, Crawford, and Goodman (2011) find that richer parents have higher expectations about their children’s educational attainments and that young people from poorer families have lower ability beliefs, a more external locus of control and lower educational aspirations and expectations. After controlling for attainment at age 11, 15% of the socio-economic gap in attainment at age 16 is accounted for by children’s attitudes, and an additional 12% by parents’ attitudes.<sup>12</sup> Chevalier, Gibbons, Thorpe, Snell, and Hoskins (2009) find that working class undergraduates underestimate their performance relative to others, but also that working class secondary school pupils are relatively more confident and have a relatively more positive self-evaluation of their math ability: this finding may be due to differences in reference groups (the so-called “big fish, small pond effect”).

<sup>8</sup> To achieve this result, they have to assume the existence of non-informative tasks, in which the probability of success is equal to one.

<sup>9</sup> This can be seen as a possible application of robust control models *à la* Hansen and Sargent (2001), in which agents take rational choices having an imperfect knowledge of the state of nature.

<sup>10</sup> Bénabou and Tirole (2002) also assume that discount rates are lower at shorter horizons than at more distant ones (time-inconsistency). Belzil (2007), however, find a predominance of the future component of intertemporal utility over the present component in schooling decision, and interpret it as evidence supportive of the standard time-consistent model.

<sup>11</sup> Models in Köszegi (2006) and Weinberg (2009) are characterized by a small number of periods. Hvide (2002) justifies pragmatic beliefs in the long run with a thought experiment in which “the agent takes into account what pays rather than what is true.”

<sup>12</sup> Starting from Lazear (1977), some authors have considered education as a normal consumption good. As long as ability is also positively correlated with family background, this would constitute an alternative mechanism capable of explaining the positive correlation between family background, confidence, and the persistence of gaps in educational outcomes. However, the fact that the results of Chowdry et al. (2011) are observed during compulsory education constitutes evidence in favor of a direct inter-generational transmission of confidence rather than an indirect effect mediated by tastes for schooling, which are more likely to matter for tertiary education enrollment.

In what follows we provide additional evidence about the link between socio-economic background and self-confidence using data from the OECD-PISA study. This dataset contains what we believe to be a good proxy for self-confidence, namely “Science Self-Efficacy,” an index built from student’s answers to questions about the ease with which they believe they could perform eight science-related tasks. This variable is a good proxy for beliefs about academic ability because it is meant to go “beyond how good students think they are in subjects such as science. It is more concerned with the kind of confidence that is needed for them to successfully master specific learning tasks, and is therefore not simply a reflection of a student’s abilities and performance” (OECD, 2009).<sup>13</sup>

We regress science self-efficacy on the PISA index of economic, social and cultural status, adding controls at the individual, school and family level; results are presented in Table 1.

The relationship between self-efficacy and family background is statistically significant and positive (as expected) and convex. In the second column we control for the score obtained by the student in the Science section of the test. This is a proxy for “true” ability, comparable across students in different countries and unobserved by the student at the time of filling in the questionnaire. The inclusion of the PISA score captures some variance of self-efficacy, but the positive relationship with family background remains strong. Notice that controlling for the PISA score is likely to bias downward the role played by self-confidence, because if our model is correct the PISA score already encompasses a gap in the human capital accumulated up to that point and that is partly due to differences in self-confidence itself. In other words, two students with the same innate ability but characterized by a different initial self-confidence should also display a different PISA score.

Adding further controls at the student level (column 3) and at the parent and school level (column 4) does not change significantly the results, which we interpret as suggestive evidence that family background has a direct impact on self-confidence, over and above the one operating through the transmission of cognitive skills.<sup>14</sup>

### 3. The model

In this section we present a multi-period model based on the assumptions already discussed in Sections 2.1–2.3. Agents choose a task on the basis of their beliefs, which are updated in a Bayesian manner after observing the outcome of every choice. Our purpose is to highlight the role played

by confidence in explaining educational attainments via task choice.

As already explained, we assume that children do not know their own ability  $a$  and hold a belief represented by the density function  $\mu(a)$ . We define confidence as the perceived ability  $\hat{\mu}(a) = \int a\mu(a)da$ , and underconfident a student who underestimates her ability:  $\hat{\mu}(a) < a$ . Similarly, the overconfident is characterized by  $\hat{\mu}(a) > a$ . Students make educational choices by choosing “tracks” ( $\psi$ ). We think of tracks as a rather general concept, encompassing either “real” school tracks (e.g. academic vs. vocational high schools) or any goal that the student sets herself. In the latter sense, for instance, a student choosing a more difficult “track” could be a student choosing to study for many exams at the same time (with the “risk” of failing or doing poorly in all of them), or choosing to delve into a subject by devoting a lot of time to it (with the “risk” of not getting that much out of it in the end).

Under both interpretations, more difficult tracks are more costly in terms of effort, but they also yield higher payoffs in case of success. A failure could be interpreted either as a true failure in a “real” track (e.g. the student drops out or must repeat a grade) or as the chance that, in trying to deeply understand some difficult material, the student wastes energy and time, ending up learning less than she would have done had she been less ambitious.

We assume that the probability of success is given by

$$p(s) = f(a, \psi) \quad (1)$$

where  $\psi$  represents how difficult is the track chosen. The probability of success is assumed to be increasing in ability ( $f'(a) > 0$ ) and decreasing in the difficulty of the track ( $f'(\psi) < 0$ ).

Students have then the possibility of updating their beliefs using Bayes’ rule, when additional information can be derived from the outcome of their choice. Given a generic density of prior beliefs  $\mu(a)$ , posterior beliefs after receiving the signal implicit in the outcome  $o = \{s; f\}$  are equal to:

$$\mu(a|o) = \frac{p(o)\mu(a)}{\int p(o)\mu(a)da} \quad (2)$$

Successful outcomes ( $s$ ) in the track chosen allow the agents to add human capital  $k(\psi|s)$  to working life productivity, and agents maximize their instantaneous utility by choosing the track that optimally balances the expected acquisition of human capital with a convex cost of acquiring it  $U[p(s)k(\psi) - \psi^2]$ , given their confidence about unobserved ability.

If the track chosen is totally uninformative ( $p(s)=1$ ) the student does not gather evidence that contradicts her wrong beliefs. For instance, this may happen when there is a discrete set of tracks and the less able students self-select into the easiest track characterized by no probability of failure. This is admittedly a limit situation, and therefore we prefer to concentrate on what happens to the gap in the accumulation of human capital when agents do learn from observed outcomes and proceed with Bayesian updating of their beliefs until their confidence eventually converges to the true value of ability.

<sup>13</sup> See Ferla, Valcke, and Cai (2009) for a discussion on the differences between Self-Efficacy and Self-Concept. Since Self-Efficacy solicits goal-referenced evaluations and does not ask students to compare their ability to that of others, we believe it is a better proxy for the notion of confidence that we use in the model of Section 3.

<sup>14</sup> We obtained similar results using a sample of Bocconi students, a much more homogeneous population at a later stage of their academic career. Results are not reported to save space but are available upon request and are consistent with recent findings by Delaney, Harmon, and Redmond (2011).

**Table 1**

Results: science self-efficacy.

	(1) Baseline	(2) Pisa score	(3) Effort	(4) Parents
Index of socio-ec. status	0.318*** [0.014]	0.145*** [0.014]	0.111*** [0.015]	0.119*** [0.026]
Index of socio-ec. status <sup>2</sup>	0.033*** [0.009]	0.022* [0.009]	0.030** [0.009]	0.026 [0.014]
Female	-0.157*** [0.012]	-0.141*** [0.011]	-0.157*** [0.011]	-0.031 [0.017]
PISA score in science		0.004*** [0.000]	0.004*** [0.000]	0.003*** [0.000]
Out of school – science			0.112** [0.009]	
Self study – science			0.114*** [0.006]	
Interest in learning science				0.226*** [0.012]
Personal value of science				0.222*** [0.013]
Parents' value of science				0.001 [0.013]
Science career motivation				-0.029** [0.009]
Science activities at age 10				0.062*** [0.008]
School-level characteristics	No	No	No	Yes
R <sup>2</sup>	0.119	0.230	0.255	0.355
Observations	225,098	225,098	216,304	29,970

BRR standard errors in brackets. All regressions include country dummies and controls for immigrant status, tracking and the interaction between tracking and the socio-economic status. In column 4 we also control for school-level variables like school size, student-teacher ratio, ability sorting and a dummy for public schools.

\*  $p < 0.1$ .  
 \*\*  $p < 0.05$ .  
 \*\*\*  $p < 0.01$ .

To achieve this goal we make some simplifying assumptions. First, we assume that the probability of success is linear in ability. The reason is that, as anticipated in Section 2.1, we concentrate on the role played by the *level* of one's perceived ability and not by the *precision* of such belief.<sup>15</sup> Hence, the probability of success is defined as:

$$p(s) = \psi a + (1 - \psi). \quad (3)$$

This specification implies that the importance of ability is proportional to the difficulty of the track. Notice that for the probability of success to be properly defined ability must have a finite support. For the sake of simplicity we assume  $a \in [0, 1]$  and  $\psi \in (0, 1]$ . The extreme value  $\psi = 0$  would correspond to the case mentioned above

in which ability does not matter and the signal is totally uninformative.<sup>16</sup>

We also assume that more difficult tracks allow students to acquire more human capital if successful, and in particular that the level of capital is equal to

$$k(\psi, \mu(a), a|s) = \frac{\psi}{1 + f(m)}, \quad (4)$$

<sup>15</sup> This is a major difference with respect to the model in Sjögren and Sällström (2004), who assume that the probability of successfully acquiring skills of type  $c_1$  is  $p(s) = a^{c_1}$ , where  $a \in [0, 1]$  is the agent's unknown ability, while  $c_1$  measures the ability elasticity of success. In such a framework the precision of the signal is crucial, because uncertainty about ability makes riskier options more or less attractive depending on whether the probability of success is convex or concave in ability. For instance, what could happen with a convex probability of success is that a totally uncertain agent could think to have more chances of succeeding than an agent characterized by quite a precise belief of being above the average. We chose to remove such discontinuities by assuming linearity in ability in Eq. (1) and to focus on the effect of the level of confidence.

<sup>16</sup> The probability of success that characterizes our model implies that our agents make decisions under uncertainty. Our work can thus be linked to the literature that regards education as a risky investment and that investigates the role played by risk aversion. Belzil (2007) notes that the degree of confidence is a measure of ex-ante risk, caused by agents' ignorance about own ability. He finds evidence of frequent (but moderate) over-estimation and cases of severe underestimation (particularly among the most able individuals). Belzil and Leonardi (2007) develop a dynamic econometric model of sequential schooling decisions in which the dropout rate depends on parental background, individual-specific heterogeneity and individual-specific risk aversion, departing from a more standard representative-agent framework. They find that risk aversion can be a deterrent to investing in education, but that differences in risk attitudes account for a modest portion of the probability of entering higher education. Trachter (2011) analyzes the effect of providing full insurance on decisions about enrollment in postsecondary education. In his model, the provision of insurance increases enrollment in 4-year colleges (which is the riskiest option an individual can take) by 81%, and total enrollment by 40%. In our framework, the effect of underconfidence would be confounded with that of risk aversion. We therefore prefer, in order to keep the model more tractable, to isolate the role of confidence by assuming risk neutrality.

where  $m = (a - \hat{a})$  represents the mismatch between the actual level of ability of the student ( $a$ ) and the optimal level of ability for that track, i.e. the level of ability that characterizes a perfectly informed student that maximizes her utility by choosing exactly that track ( $\hat{a}$ ). We assume that  $f(0) = 0$ , i.e. that human capital coincides with the difficulty of the track when ability perfectly fits. Alternatively, the amount of human capital actually acquired is discounted, with the shape of  $f(m)$  crucially affecting the results when  $m \neq 0$ . In particular, we assume that  $f'(|m|) \geq 0$  meaning that neither under- nor overconfidence can increase human capital beyond  $\psi$ . This assumption might appear counter-intuitive at first glance, but it has the great advantage of preventing self-deception. Consider the case in which in the same track the human capital is lower for the overconfident successful student, because her ability is lower than what optimal for such a track, while the opposite happens for the underconfident successful student. The possibility of supplementing the human capital provided by the chosen track with an ability higher than  $\hat{a}$  would imply that there is room for self-deception, i.e. that systematically underestimating one's ability might become an optimal solution, with a consequent bias in the choice of the track (a bias we want to avoid for the reasons outlined in Section 2.2). Of course, the effect of the mistake in evaluating ability does not need to be symmetric. In the simulation below we will assume that underconfidence has no effect ( $f(m) = 0$  when  $m > 0$ ), while overconfidence has a negative impact ( $f'(|m|) > 0$  when  $m < 0$ ). To complete the picture, we assume that a failure leaves the stock of human capital unchanged, i.e.  $k(\psi, \mu(a), a|f) = 0$ .<sup>17</sup>

Students are free to self-select into different tracks given the best estimate of their ability, trading off a lower human capital in case of success with a higher probability of acquiring it. If ability was known, the first-order conditions would imply<sup>18</sup>:

$$\psi^* = \frac{1}{2} \frac{1}{2 - \hat{a}}. \quad (5)$$

The presence of  $f(m)$  in Eq. (4) implies that it is always optimal to truthfully report one's unknown ability (i.e. to set the mistake  $\hat{\mu}(a) - a = 0$ ). Hence, the optimal track choice becomes an increasing function of confidence.

<sup>17</sup> This assumption is made without loss of generality as compared to the case in which the human capital accumulated in case of failure is positive but strictly lower:  $k(\psi, \mu(a), a|s) > k(\psi, \mu(a), a|f)$ .

<sup>18</sup> To analyze the role played by self-confidence in shaping the gap in educational attainments when agents are eventually learning their true level of ability we need to iterate this choice for several periods. In principle, we should compute the optimal track choice by maximizing a lifetime utility function. Since additional information about one's ability is valuable per se as long as it helps making better choices in the future, agents could be willing to pay a price to receive a more informative signal, by choosing a track slightly different than what would be optimal in a static framework. However, such an effect is of a second order magnitude in our framework and it does not determine appreciable changes in the results (see footnote 24 below), thereby not justifying the corresponding increase in the complication of the model. Hence, we assume that agents are myopic and that they maximize their expected utility period by period. Alternative theoretical approaches in which the information value of the educational or working choice is instead crucial have been proposed by Jovanovic (1979), Miller (1984), and more recently by Trachter (2011).

However, even removing any bias in the self-evaluation of ability,  $\hat{\mu}(a)$  and  $\hat{a}$  may still differ due to insufficient information. Eq. (5) therefore implies that both under- and overconfidence determine a suboptimal track choice and a loss of utility, as long as  $\hat{\mu}(a) \neq \hat{a}$ .

The effect of under- and overconfidence can differ as far as the accumulation of human capital is concerned. Rewriting confidence as the composition of optimal ability and the evaluation mistake  $\hat{\mu}(a) = \hat{a} + m$ , expected human capital is given by:

$$E(k) = -\frac{1}{4} \frac{\hat{a} + 2m - 3}{(\hat{a} + m - 2)^2(1 + f(m))}. \quad (6)$$

The relationship between confidence and human capital can be summarized by means of the derivative of  $E(k)$  with respect to the mistake  $m$ :

$$\frac{\partial E(k)}{\partial m} = \frac{1}{2} \frac{m - 1}{(\hat{a} + m - 2)^3(1 + f(m))} + \frac{1}{4} \frac{(\hat{a} + 2m - 3)f'(m)}{(\hat{a} + m - 2)^2(1 + f(m))^2}. \quad (7)$$

As long as small ability mismatches have a negligible impact (i.e. as long as  $f'(0)$  is sufficiently small), the derivative is positive around  $m = 0$  for every value of  $a \in [0, 1]$ . This means that a small degree of overconfidence ( $m > 0$ ) increases the amount of expected human capital, although at a price of lower utility because the increase of human capital would be acquired overestimating the expected return on the additional effort.<sup>19</sup> As overconfidence increases, the sign of  $\partial E(k)/\partial m$  depends on the magnitude of the effect of the mismatch. In the limit case in which there is no effect, e.g. when  $f(m) = 0$  in Eq. (4), or in any case when such an effect is negligible, the human capital acquired would monotonically increase with overconfidence, since the positive effect of a higher human capital acquired in the case of success dominates the negative effect of a lower chance that this event happens. In contrast, if the effect of overevaluating one's ability increases substantially with the size of the mistake (e.g. if  $f(m) = m^2$ ), the relationship between expected human capital and overconfidence becomes bow-shaped. As far as underconfidence is concerned, the condition that ensures that there is no incentive to self-deception is also sufficient to grant that human capital decreases monotonically as underconfidence increases.

Agents update their beliefs given the signal (success or failure) received at the end of each period.<sup>20</sup> In order to characterize the learning process and to investigate

<sup>19</sup> The reason is that the probability of success depends on the true level of ability, and overconfidence would grant a higher level of human capital when successful, but a positive outcome is less likely to happen than what an overconfident agent expects.

<sup>20</sup> Note that if the agent received a perfectly informative signal (like the exact amount of human capital acquired when successful) he could invert  $k(\psi, \mu(a), a|s)$  deriving with certainty her true ability level. However, data suggest that uncertainty about ability survives a long string of signals, which must therefore be not perfectly informative (or, if they are perfectly informative, agents cannot fully exploit them). In what follows we assume that agents only observe the event success vs. failure. In other words,

the effect of self-confidence on educational attainments we need to specify how beliefs about one’s ability are shaped. The Beta distribution perfectly fits our assumption of a finite support of the ability distribution, necessary to ensure that the probability of success is linear in ability. At the same time the Beta distribution is sufficiently general to allow prior beliefs to represent different levels of confidence while keeping the whole domain of ability in their support, something necessary because with a Bayesian learning process agents can never assign a positive probability to events excluded by the prior.

The density function of the  $Beta[\alpha, \beta]$  distribution is:

$$\mu(a) = \frac{a^{\alpha-1}(1-a)^{\beta-1}}{\int_0^1 a^{\alpha-1}(1-a)^{\beta-1} da}, \tag{8}$$

while the mean is given by:

$$\hat{\mu}(a) = \int_0^1 a\mu(a)da = \frac{\alpha}{\alpha + \beta}. \tag{9}$$

When  $\alpha = \beta > 1$  the distribution is symmetric and bell-shaped. The distribution is skewed to the left when  $\alpha > \beta > 1$ , and to the right when  $\beta > \alpha > 1$ .<sup>21</sup> The higher  $\alpha$  and  $\beta$ , the lower the variance and therefore the more precise the beliefs. We assume that ability is distributed in the population following a  $Beta[2.5, 2.5]$ , and that the same distribution also characterizes the beliefs of the median student. This is equivalent to assume that the median student ( $a = 0.5$ ) holds correct beliefs about her ability, because when  $\mu(a) \sim Beta[2.5, 2.5]$  confidence is  $\hat{\mu}(a) = 0.5$ .

Before analyzing the effect of over- and underconfidence let us focus on the median student in order to describe in some details the learning process. After observing the outcome, the agent updates her beliefs using Bayes rule. In particular, her posterior beliefs after observing a success are:

$$\mu(a|s) = \frac{(\psi a + 1 - \psi)\mu(a)}{\int_0^1 (\psi a + 1 - \psi)\mu(a)da}. \tag{10}$$

By contrast, if a failure was observed:

$$\mu(a|f) = \frac{(\psi - \psi a)\mu(a)}{\int_0^1 (\psi - \psi a)\mu(a)da}. \tag{11}$$

The mass of probability is reallocated according to the realization of the signal, towards the upper bound if successful (see Fig. 2, right curve) and toward the lower bound if not (see Fig. 2, left curve), keeping constant the support of the density. Notice that the bad event has a stronger effect when updating beliefs.<sup>22</sup>

agents only know the *potential* amount of human capital  $\psi$ , but not the *actual* amount corrected for the ability mismatch  $1 + f(m)$ . An intermediate situation in which additional information can be extracted from a noisy signal of the level of human capital actually acquired (in other words when different degrees of success are observable) could be formalized at the price of a significant increase in the complication of the model, but without appreciable additional insights. Hence, we prefer to stick to the simplest version of the information structure.

<sup>21</sup> The uniform is a special case of the Beta distribution when both parameters are equal to 1.

<sup>22</sup> The reason is that a failure is far less likely given the specification of the model. In fact, the student with correct prior beliefs will revise her

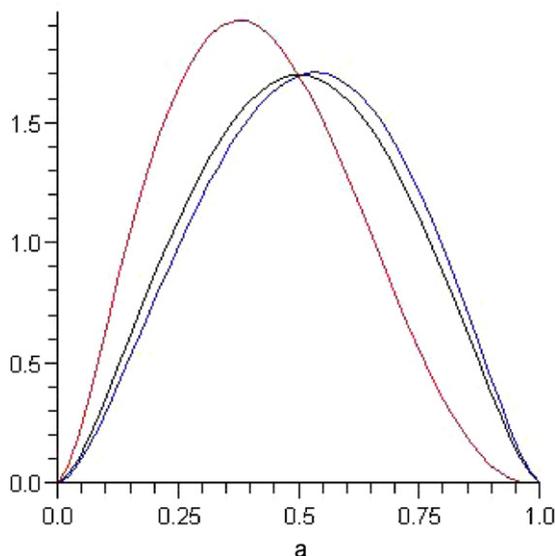


Fig. 2. Beliefs updating of the median student after the first signal.

The agent will then choose again the optimal track given posterior beliefs, that will be further revised after observing the outcome in the second period, and so on and so forth. The bottom line is that, within the support of initial beliefs, the distribution of beliefs changes according to the history of signals observed. Subsequent updates bring beliefs closer and closer to the true ability level as long as the agent receives informative signals.

#### 4. Simulation

In order to study the effect of self-confidence, we compare the choices made and the human capital accumulated by an agent whose ability is  $a = 0.5$  when she holds correct prior beliefs on average ( $\mu(a) \sim Beta[2.5, 2.5]$ ) against the counterfactuals in which she is underconfident and overconfident, respectively. In other words, we simulate the model picking up the median student and looking at the effect on her educational attainments of a wrong confidence in both directions. In fact, the higher human capital accumulated when the student is not too overconfident (i.e. when the mismatch effect does not prevail) and successful can be compensated by a probability of achieving it that is lower for two reasons. First, because the track is more difficult and therefore the same person is more likely to fail. Second, because true ability is lower than confidence. In the utility maximization only the former is correctly internalized, and the student will therefore be successful less often than she expects. This is the engine that eventually drives her confidence towards the true level of ability.

confidence upward a fraction  $1 - 0.5\psi$  of the times, while she will revise her confidence downward in the other  $0.5\psi$  times. While her expected posterior confidence is always unchanged at 0.5, the upward and downward revisions would be symmetric only when  $\psi = 1$ , i.e. when the two events are equally likely.

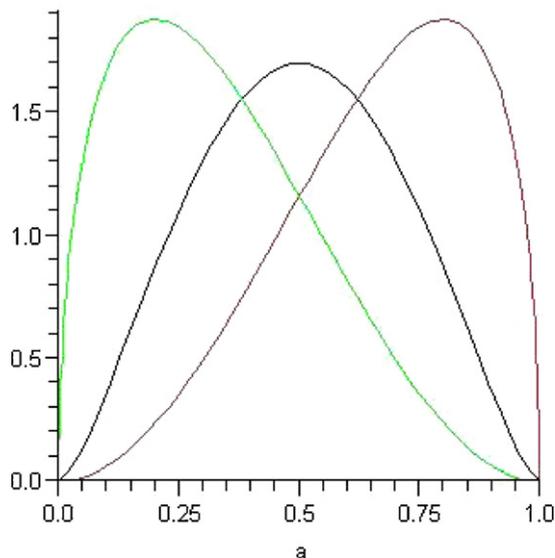


Fig. 3. Prior beliefs given the different levels of confidence.

We represent underconfidence with a distribution of prior beliefs

$$\mu(a) \sim \text{Beta}[1.5, 3] \quad (12)$$

skewed to the right. This implies a level of confidence  $\hat{\mu}(a) = 1/3$ , corresponding to the 24th percentile in the true distribution.

Similarly, overconfidence is summarized by a distribution of prior beliefs

$$\mu(a) \sim \text{Beta}[3, 1.5] \quad (13)$$

skewed to the left, which implies a level of confidence  $\hat{\mu}(a) = 2/3$ , corresponding to the 77th percentile in the true distribution. These parameters also imply that the three distributions have roughly the same variance, and therefore that over- and underconfidence are perfectly symmetric.<sup>23</sup> Prior beliefs of the three different types of student are summarized in Fig. 3. As far as the ability mismatch described in Eq. (4) is concerned, we choose no correction in case of underconfidence ( $f(m) = 0$  if  $m > 0$ ) and a quadratic term  $f(m) = 3m^2$  if  $m < 0$ , implying a discount of about 7.5% in the human capital acquired in the first period by a successful overconfident student.

We analyze what happens to the human capital accumulated by the three types of agents while the learning process takes place, iterating the updating of beliefs 45 times. Since the single realization of human capital relies upon a random component, we replicate the procedure 200 times.

Results show that the value of confidence slowly converges towards the true ability level for those starting with

<sup>23</sup> Although the probability of success does not depend on the variance of beliefs, the latter could still affect the updating process, since the more precise the beliefs, the lower the change of confidence induced by the same signal received. We do not want the learning pattern to be affected by a different precision of beliefs, and therefore we assume the same variance in the prior distributions.

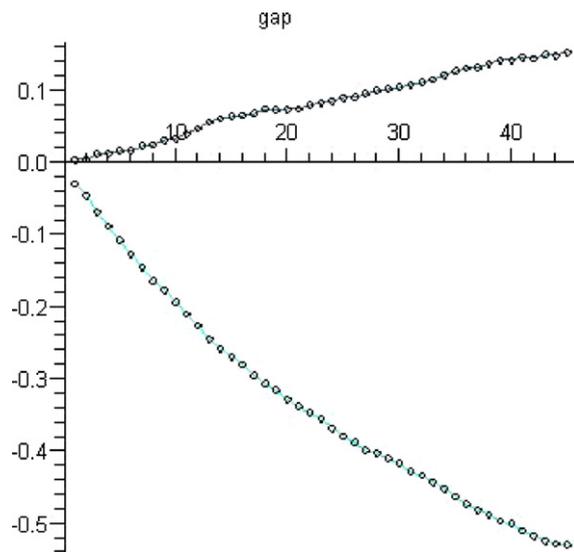


Fig. 4. Gap in the accumulation of human capital.

a wrong prior, but also that the learning process is far from being completed. In fact, at the end of the 45th iteration confidence is about .425 for the underconfident and .558 for the overconfident, in both cases significantly different than .5 ( $p < 0.001$ ).<sup>24</sup>

Fig. 4 displays the average gap, period by period, across repetitions, in the accumulation of human capital of the agents who start with wrong priors as compared to those starting with correct beliefs. The human capital accumulated by the underconfident is significantly lower than the human capital acquired by the student holding correct beliefs ( $p < 0.001$ ), while the opposite happens for the overconfident type ( $p < 0.001$ ), although the magnitude is different in absolute terms because of the cost of the mismatch  $f(m)$ . Notice that at the beginning, when overconfidence is larger (and so the cost of the mismatch), the human capital accumulated is not much higher, while it increases as compared to the student with correct beliefs as long as confidence converges towards the true type and the cost of the mismatch decreases. Given the chosen

<sup>24</sup> The speed of convergence of the two types differs a little bit. In fact, the mistake in confidence becomes significantly smaller for the overconfident ( $p = 0.038$ ). The reason is that the higher the track chosen, the more balanced the probability of success given the true level of ability  $a = 0.5$ , the more informative the signal. At first glance this seems to imply that the choice of track and the educational outcomes could have been different had we internalized the different informativeness of the signals by means of dynamic optimization. In fact, there seems to be an additional incentive to choose a higher track, thereby reducing the effect of underconfidence while increasing that of overconfidence. This is not the case, however, because such an argument holds only when the probability of success is computed knowing the true value of ability. When choosing  $\psi$ , in contrast, agents use the best estimate of their ability  $\mu(a)$ . Notice that the perceived probability of success is increasing in  $\mu(a)$ . Hence, internalizing the different informativeness of the signal would imply a lower revision of the optimal choice at low levels of ability. In any case, maximizing utility period by period implies choices that marginally differ in terms of magnitude and therefore a negligible mistake particularly at low levels of ability.

specification of the model, the gap between the overconfident and the underconfident turns out to be about 6%.

To summarize, self-confidence can determine significant differences in the outcomes observed. When the learning process reaches the fixed point implied by discovering the true level of ability, the three types in the simulation will start making the same choices and from that moment onwards they will be observationally equivalent. However, the level of human capital acquired is and will remain significantly different. Wrong beliefs about one's ability do not need to be self-confirming to explain unequal outcomes if they lead to significantly different choices during the learning process. As long as family background shapes children's beliefs about their ability, confidence can be a transmission mechanism that increases the intergenerational persistence of outcomes.

Notice that in the model the probability of success increases with innate ability only, while the human capital accumulated plays no role. As already noticed in Section 2.3, this simplifying assumption downplays the role of nurture, since achievements are also determined by the whole history of intermediate outcomes, in turn also driven by self-confidence. Therefore, what found by the model is once more a lower bound of the role of self-confidence, since the cumulative effect of the gap in the human capital accumulated during the learning process of one's ability is not taken into account. The role of nurture therefore implies that tests meant to measure students' ability are instead capturing also the gap in human capital accumulated up to that point because of a different family background. For instance, a centralized test administered at age 15 in order to select students into different tracks would probably classify as different two students characterized by the same innate ability but with a different background, thereby helping to perpetuate intergenerational inequalities. A policy implication arising from this consideration is therefore that cognitive tests should take place as early as possible in order to endow parents with measures of the innate level of ability of the children that are not confounded with the role that the family background can play through self-confidence.

## 5. Conclusions

In line with some recent contributions, we claim that the socio-economic background affects not only the actual stock of cognitive skills possessed by a child (innate ability) but also the beliefs about such (unobserved) cognitive skills, i.e. self confidence. There is indeed a vast literature supporting the hypothesis that people have imperfect knowledge of their ability and that many personality traits related to the concept of self-confidence are influenced by the family background in which a child grows up.

We provide further suggestive evidence about the link between confidence and family background using the PISA dataset, which is a representative cross-national survey of 15-year old pupils. We show that the link between background and confidence is strong and survives the inclusion of good proxies for unobserved and observed ability. Our proxies of ability are likely to bias downward the estimated link between confidence and background, since they

capture not only innate ability, but also the gap in human capital that has been accumulated up to that point.

We then propose a model in which fully rational agents, who maximize the expected acquisition of human capital, choose tasks according to their perceived ability. True ability and the difficulty of the chosen track affect the probability of success. After observing whether they succeed or not, students update their beliefs, fully exploiting the available information, following Bayes' rule. We simulate the model with a bootstrapping procedure and we show that choices distorted by over- and underconfidence lead to a significant gap in the accumulation of human capital during the process in which agents eventually learn their true level of ability.

In our model agents do not derive additional utility by holding a good self-image; the consequence of this assumption is that if a perfectly informed and benevolent planner could force individuals to choose the "right" task, the effect of wrong confidence would disappear. Nevertheless, even in a setting in which agents are fully rational and have standard preferences, a moderate degree of overconfidence can be beneficial in terms of the accumulation of human capital over the life course, although at a price of a lower utility (since overconfident and underconfident agents do not make, by construction, utility-maximizing choices). Underconfidence, on the other hand, is suboptimal in terms of both utility maximization and human capital accumulation.

The intergenerational transmission of beliefs can thus constitute a further channel through which socio-economic differences perpetuate from one generation to the next. Even if two individuals had the same innate cognitive ability, differences in beliefs would lead them to make different choices in terms of investment in education. The results of our analysis suggest that policy interventions aimed at providing early and precise feedback about the cognitive skills of children from disadvantaged backgrounds can be beneficial in helping to narrow the gaps in educational attainments, by avoiding that equally talented people make different choices only because they have inherited different beliefs about their potential.

## Acknowledgements

This research was supported by EC Grant, FP7-244592. We would like to thank Carlo Devillanova, Marcel Jansen, Marco Leonardi, Michele Pellizzari, Jan Van Ours, and seminar participants at Bocconi University, Universitat Autònoma de Madrid, Tilburg University, ASSET, EALE, ESSLE and ESPE conferences for useful comments and suggestions. All remaining errors are ours. The views expressed in this paper are those of the authors and do not involve the responsibility of the Bank of Italy.

## References

- Attanasio, O., & Kaufmann, K. (2009). *Educational choices, subjective expectations, and credit constraints*. NBER Working Papers 15087 National Bureau of Economic Research, Inc.
- Belzil, C. (2007). *Subjective beliefs and schooling decisions*. Discussion Paper Series 2820 IZA.

- Belzil, C., & Leonardi, M. (2007). Can risk aversion explain schooling attainments? evidence from Italy. *Labour Economics*, 14, 957–970.
- Bénabou, R. J. M., & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117, 871–915.
- Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, XXXIX, 1137–1176.
- Brunello, G., & Schlotter, M. (2011). *Non cognitive skills and personality traits: Labour market relevance and their development in education and training systems*. IZA Discussion Papers 5743 Institute for the Study of Labor (IZA).
- Cesarini, D., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Heritability of Overconfidence. *Journal of the European Economic Association*, 7, 617–627.
- Chevalier, A., Gibbons, S., Thorpe, A., Snell, M., & Hoskins, S. (2009). Students' academic self-perception. *Economics of Education Review*, 28, 716–727.
- Chowdry, H., Crawford, C., & Goodman, A. (2011). The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16. *Longitudinal and Life Course Studies*, 2, 5–76.
- Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31–47.
- Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43, 738–782.
- Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78, 883–931.
- Deke, J., & Haimson, J. (2006). *Valuing student competencies: Which ones predict postsecondary educational attainment and earnings, and for whom?* Technical Report Mathematica Policy Research Inc.
- Delaney, L., Harmon, C., & Redmond, C. (2011). Parental education, grade attainment and earnings expectations among university students. *Economics of Education Review*, 30, 1136–1152.
- Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment. *Psychological Science in the Public Interest*, 5, 69–106.
- Falk, A., Huffman, D., & Sunde, U. (2006). *Self-confidence and search*. Discussion Paper Series 2525 IZA.
- Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19, 499–505.
- Goodman, A., & Gregg, P. (Eds.) (2010). *Poorer children's educational attainment: How important are attitudes and behaviour?* Joseph Rowntree Foundation Report. London. Joseph Rowntree Foundation.
- Gregg, P., & Washbrook, E. (2011). The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 11. *Longitudinal and Life Course Studies*, 2, 41–58.
- Hansen, L. P., & Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, 91, 60–66.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24, 411–482.
- Hvide, H. K. (2002). Pragmatic beliefs and overconfidence. *Journal of Economic Behavior and Organization*, 48, 15–28.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics*, 125, 515–548.
- Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy*, 87, 972–990.
- Kaufmann, K. M. (2010). *Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns*. Working Paper Series 362 IGIER, Bocconi University.
- Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105, 473–522.
- Köszegi, B. (2006). Ego utility, overconfidence and task choice. *Journal of the European Economic Association*, 4, 673–707.
- Lazear, E. P. (1977). Education: Consumption or production? *Journal of Political Economy*, 85, 569–597.
- Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy*, 92, 1086–1120.
- Nguyen, T. (2008). *Information, role models and perceived returns to education: Experimental evidence from Madagascar*. Working Paper MIT.
- OECD (2009). *Top of the class. High performers in science in PISA 2006 programme*. OECD Publishing.
- Sjögren, A., & Sällström, S. (2004). *Trapped, delayed and handicapped*. Working Paper Series 613 Research Institute of Industrial Economics.
- Trachter, N. (2011). *Option value and transitions in a model of postsecondary education*. EIEF Working Papers Series 1103 Einaudi Institute for Economic and Finance (EIEF).
- Weinberg, B. A. (2009). A model of overconfidence. *Pacific Economic Review*, 14, 502–515.