Pathways to Poverty:

Theoretical and empirical analyses

Iain Peter Embrey*

Lancaster University

September 2019

A thesis toward the degree of Doctor of Philosophy,

supervised by
Steve Bradley and Kim Kaivanto

*MMath; PGCE; MEd; MSc Economics pending. i.embrey@lancaster.ac.uk.
Lancaster University Management School, Lancaster, LA1 4YX.

Abstract

The prevalence of poverty in advanced economies represents a challenge, both to economic theory and to society. We know that poverty is perpetuated by low levels of educational investment amongst disadvantaged children, but we have no credible theoretical explanation for the observed degree of that apparent underinvestment, and we have not yet developed sufficient policy tools to break the intergenerational cycle of deprivation. In response, this thesis undertakes theoretical and empirical analyses of the pathways that perpetuate poverty. I demonstrate that divergently low educational investment could arise as an equilibrium response to a grades-focussed educational system; I develop the existing state-of-the-art technique in econometric estimation of the educational production function; and I apply that technique to find strong empirical support for my theoretical model. In addition my results show that the average child's propensity to think analytically has a substantial influence over their developmental pathway, which suggests that models of educational investment should adopt a generalisation of Expected Utility Theory that allows agents to maximise one of two possible objective functions.

JEL: I24; D91; C73; C61; J24; I21; C38; D01; D81; I31; J64; B41

Keywords: Education; Human Capital; Poverty; Social Exclusion; Unemployment; Cognitive Development; Noncognitive Skills; Dynamic Factor Analysis; Participation Game; Decision Theory; Dual-Self.

Acknowledgements: I am deeply and sincerely grateful to Kim and to Steve for their supervision. I would also like to thank Colin Green and Ian Walker for their supervisory input, and the whole economics department at Lancaster for their kind support and keen insight. Worthy of particular mention are: Caren Wareing, Emma Fitchett, Craig Becker, Emanuele Bracco, Kwok Tong Soo, Geraint Johnes, David Kang, Mike Tsionas, Vincent O'Sullivan, David Rietzke, Kwok Tong Soo, Olivier Cardi, and Efthymios Pavlidis. This work was supported by the Economic and Social Research Council under studentship grant number ES/P000665/1; there are no further interests to declare.

Declaration

1. I confirm that I am the sole author of all work contained in this thesis, and that no part of it has previously been submitted toward any academic qualification.

2. Further, the work contained in this thesis is original, save in the following respects:

   (a) The penultimate and antepenultimate paragraphs of the introduction closely mirror observations published in Embrey (2019e);

   (b) In addition, the derivation of an individual's propensity to think analytically in Chapter 4 builds upon the generalised decision theory presented in Embrey (2019e), as is acknowledged there.

   (c) Chapter 2 is a development of previous work undertaken in Embrey (2019b). This is published as a departmental working paper, it was presented as a departmental seminar, and it was reviewed by several referees for the Journal of Political Economy. The present version simplifies and streamlines the presentation of that previous work; it was presented at the European Economic Association conference 2019, and it has been under review at the Journal of Economic Behavior and Organization for some months.

   (d) Chapter 3 is original work that has been presented at the North West Social Sciences Doctoral Training Centre Conference 2019, and at a departmental seminar. This work has been under review at the Journal of Human Resources for some months.

   (e) Chapter 4 is original work that has yet to be submitted for publication.

   (f) Chapters 3 and 4 make use of the analysis of CES production functions that was undertaken in Embrey (2019c). This is published as a departmental working paper, and it has been under review at the Journal of Economics for some months.

   (g) The Holistic Introduction and Conclusion refer to work that was undertaken in Embrey (2019a). This is published as a departmental working paper, it was presented at the Neoclassical Repair Shop conference 2019, and it has been under review at the Journal of Economic Surveys for some months.

   (h) Word Count: 53,928.

Contents

List of Figures

List of Tables

Chapter 1: **An Holistic Introduction**

"The ultimate goal of economic science is to improve
the living conditions of people in their everyday lives."

Economics, Samuelson & Nordhaus (2009, p.7)

The living conditions of people in their everyday lives are not all in equal need of improvement. Moreover, it is well known that people who are deprived in one aspect their everyday life tend to face deprivation across a cluster of life outcomes. As such, impoverished individuals are often characterised by the co-incidence of poor outcomes in education, employment, earnings, mental health, physical health, criminality, and family life (Hills, Le Grand & Piachaud 2002). In this thesis I pursue the goal of Samuelson and Nordhaus (see epigraph) by developing our understanding of the pathways that lead individuals into multi-dimensional poverty.

A large quantity of research has sought to address this question. Within economics, the majority of this research is empirical, and the typical approach is to use a large- or small-scale policy initiative to identify a causal effect on specific dimensions of socio-economic disadvantage. Such studies can illuminate the pathways that perpetuate poverty from parents to their children, because they can establish the extent to which exogenous changes in measurable characteristics affect those pathways. This approach provides an insight into *what* conditions might lead to poverty, but it says little about *how* those conditions might lead to poverty, and it cannot propose new avenues for policy intervention.

A complementary strand of empirical research seeks to estimate the technology by which socio-economic outcomes are produced. This literature builds upon the founding contributions of Cunha & Heckman (2008) and Cunha, Heckman & Schennach (2010), which investigate the simultaneous development of cognitive and noncognitive skills. From this literature we learn that early intervention in cognitive skill formation is vital, because a child's ongoing cognitive development pathway has largely been determined by the time they reach compulsory schooling age. This approach could be applied to test the effect of any measurable factor on the development of any socio-economic outcome, but hitherto its use has typically been limited to estimating the technologies of cognitive and noncognitive skill formation as functions of parental investment and skill levels.[1]

In order to investigate *how* various inputs could affect the production of socio-economic outcomes we therefore turn to economic theory. If we can model the mechanism by which an outcome is produced, then we can predict which factors are likely to affect its production, and we can suggest new avenues through which pubic policy might influence that production. Existing economic theory typically considers each dimension of poverty in isolation, and so there is a diverse literature of specific theories that are built to explain specific phenomena. For example, Akerlof & Kranton (2002) propose that a payoff to 'fitting in' could explain high school drop out, whilst Pissarides (1990, 2000) propose that individuals with greater wealth should place a higher relative value on leisure and so remain in unemployment for longer. I survey the existing theoretical explanations for very low educational investment in Chapter 2, and I find that we do not yet have an adequate explanation for the observed extent of this empirical phenomenon. Moreover, because existing theories target one specific socio-economic outcome, they rarely generalise to provide an holistic insight across multiple dimensions of poverty, and they do not explain why those multiple dimensions of poverty are highly correlated within individuals.

The first contribution of this thesis is therefore to propose and analyse a single economic model that is readily applicable to multiple dimensions of poverty. To achieve this we need to reconsider the mechanism by which outcomes such as education, employment, wage progression, health, criminality, and interpersonal relationships are produced. Traditional microeconomic models describe the production of these outcomes as a high-level decision over how much costly effort and/or financial investment to supply. I propose that new insight might be gained by disaggregating those high-level decisions into a series of elemental participation decisions such as: whether to attempt the current classwork, whether to respond to the current job advert, whether to search for new promotion opportunities this evening, whether to exercise this morning, whether to hang out with the street gang tonight, and whether to wash the dishes today. Each of these elemental decisions can be represented by a simple expected utility model, with participation payoffs that include long-term benefits, immediate costs, and the possibility of improving one's skill-level for similar tasks in the future. An improved skill-level might reduce future participation costs, or increase the likelihood that

[1] Schooling inputs have also been studied, and so has the role of childhood physical health in developing countries – see the literature review in Chapter 3.

participation would produce a positive feeling of success, but in either case the implications of dynamic skill development are likely to be increasing returns to participation, multiple (weak) equilibria, and therefore potentially inefficient outcomes (as explored by Arthur 1989). In the context of educational participation decisions, I show in Chapter 2 that these inefficient outcomes could include divergently low educational investment amongst disadvantaged children. In work published elsewhere, I sketch applications to other empirical anomalies including chronic unemployment and the crowding-out effect (Embrey 2019e).

The model presented in Chapter 2 applies this theoretical paradigm to educational investment decisions. It is particularly important to understand the mechanisms that lead to very low levels of educational investment amongst disadvantaged children, because this apparent underinvestment is an important first step on the pathway to multi-dimensional poverty in adulthood (Heckman 2006). I show in Chapter 2 that the model outlined above generates an Education Trap, under which children with more than a critical threshold level of initial disadvantage should optimally invest very little into their education thereafter. In addition, the model proposed in Chapter 2 operationalises a tangible mechanism of educational production, and so its parameters represent concrete aspects of educational opportunities. I am therefore able to derive practical policy recommendations from the model, each of which should reduce the proportion of children who select a low-investment pathway at equilibrium.

The fact that the model presented in Chapter 2 describes tangible educational opportunities also enables direct econometric testing of its validity. A stringent series of tests would require that heterogeneity in each of the model's proposed decision payoffs should have a statistically significant effect on the production of cognitive skills. In order to carry out those statistical tests, I devote the first part of Chapter 3 to a survey of the existing econometric approaches toward estimating the technologies of cognitive and noncognitive skill production. I conclude that the seminal contribution of Cunha, Heckman & Schennach (2010) still represents the state-of-the-art in that field, but I also identify several shortcomings in both the approach and the implementation of that work.

The second contribution of this thesis is therefore to build upon the pioneering work of Cunha, Heckman & Schennach (2010) to advance the frontier in estimating the technologies of simultaneous skill production. My contribution here includes: replacing the misleading anchoring technique with an explicit normalisation of the production function, reformulating the dynamic measurement system to remove over-identifying restrictions, nesting the production function to identify dynamic complementarity in investment, and correcting several other more minor coding errors. The first of these developments required substantial research into the properties of CES production functions; that research is now separately under review as Embrey (2019c), which may be found as a departmental working paper at https://eprints.lancs.ac.uk/id/eprint/131589/1/LancasterWP2019_004.pdf.

With the econometric developments of Chapter 3 in hand, I am able to test the theoretical predictions of the model presented in Chapter 2. Accordingly, the first part of Chapter 4 formally determines the implications of individual heterogeneity in each of the model's parameters, and discusses the interpretation of each parameter as an aspect of an individual's noncognitive skill. The second part of Chapter 4 operationalises those noncognitive skills using data from the UK Millennium Cohort Study, to find that measurable heterogeneity in each of the model's four parameters has statistically significant positive effects on ongoing cognitive skill formation. What is more, I find that the effects of an individual's drive to succeed and their idiosyncratic cost of exerting effort are both substantially larger than the effects of the benchmark proxy for noncognitive skills suggested by Cunha, Heckman & Schennach (2010). These results corroborate the model developed in Chapter 2, and they suggest specific aspects of noncognitive skill that could fruitfully be targeted in interventions that aim to support ongoing cognitive skill formation.

In addition to testing the model that was developed in Chapter 2, I also extend that model in Chapter 4 by using the generalised decision theory proposed in Embrey (2019a). Under that generalisation of Expected Utility Theory, an agent will either analyse each decision problem as per *Homo Economicus*, or they will act as if they were maximising only their immediately salient payoffs. The generalised decision theory therefore relaxes the assumption that a single functional form for utility can adequately describe the behaviour of all individuals on all occasions, by instead allowing each individual on each occasion to act so as to maximise one of two possible objective functions. The probability with which an individual will act according to the analytic preferences of *Homo Economicus* is denoted by ρ, and we refer to this parameter as the individual's idiosyncratic propensity to think analytically. As such, the statistical significance of the operationalised parameter ρ provides an econometric test of the relevance of the generalised decision theory to the present context, and in Chapter 4 I find that this parameter is a highly significant determinant of an individual's cognitive development pathway. My results show that, once a child reaches compulsory schooling age, the input share of their propensity to think analytically is larger than any other candidate noncognitive skill, and it accounts for 14% of their period-on-period cognitive development.

The final contribution of this thesis is therefore to provide initial evidence that an individual's propensity to think analytically could have a pervasive influence over their socio-economic outcomes. We demonstrate that this is the case for educational development, but we have already noted that the model presented in Chapter 2 is readily transferable to describe the incremental decisions that determine many other dimensions of socio-economic (dis)advantage. This suggests that the generalised decision theory could provide important new insights in each of those areas. For example, many health decisions are intuitively characterised by a conflict between impulsive preferences – such as the desire to watch a T.V. show right now – and analytic preferences – such as the long-term health benefits of going for a run instead. It is not clear that decision-makers necessarily act as if they project these qualitatively distinct dimensions of utility onto an hypothetical single-dimensional Expected Utility scale; instead they may follow the generalised decision theory and act as if they maximise just one of these two natural dimensions of utility on any given occasion. An initial investigation of this possibility is sketched in Embrey (2019e), which applies the generalised decision theory to provide an unified explanation for several important aspects of poverty. A thorough comparison of the foundational assumptions of Expected Utility Theory and its generalisation is provided in Embrey (2019a), and so this introduction will now provide only a brief overview of that discussion.

The main advantage of the normative assumption that there exists a single representative agent is that it typically affords a mathematically elegant analysis. However that mathematical elegance should not be mistaken for parsimony. Expected Utility Theory requires three layers of assumption: firstly, the set of relevant motivations is postulated; secondly, a functional form for each motivation is prescribed; and finally, the functional form of a single-valued utility function is also prescribed, whereby those disparate motivations are assumed to be traded-off against each other. The generalised approach typically also requires the first two layers of assumption, but it does not impose any homogeneous rule by which disparate motivations must be traded-off. Thus, ceteris paribus, the law of parsimony would favour the generalised theory (Ockham ca. 1323); a conclusion which would hold a fortiori if that generalised theory were to provide an unified explanation for multiple dimensions of poverty.

It is rare for modern decision theory to explicitly consider the validity of the above ‘single-self’ assumption set. This is because the revealed-preference paradigm of Samuelson (1938), and its formalisation by Savage (1954), demonstrate that an expected-utility representation must exist whenever a number of postulates are satisfied. The generalised theory proposed in Chapter 4 is fully compatible with those seminal observations. Its contribution is to expand the applicability of Expected Utility theory from situations in which the Savage postulates apply globally, to situations in which they apply conditional upon the decision-maker’s *state of mind*. Thus we do not require preferences to be complete, transitive, and consistent-through-time, but only that these characteristics separately describe both the agent's analytic and their impulsive thought-processes. This substantially more tenable assumption relaxes the normative assumption in much the same way that conditional independence relaxes an econometric independence assumption.

This thesis now proceeds by presenting the aforementioned theoretical, econometric, and empirical contributions in Chapters 2, 3, and 4 respectively. Mathematical derivations and ancillary estimation results are appended to each of these chapters in sequence, whilst supplementary materials such as additional robustness exercises are appended after the concluding chapter. That concluding chapter synthesises the key findings of the thesis, and it discusses the potential implications of these findings for our understanding of the pathways that lead to multi-dimensional poverty.

Chapter References

Akerlof, George A & Rachel E Kranton (2002). "Identity and schooling: Some Lessons from the Economics of Education". In: *Journal of Economic Literature* 40.4, pp. 1167–1201.

Arthur, W. Brian (1989). "Competing Technologies, Increasing Returns, and Lock-In by Historical Events". In: *The Economic Journal* 99.394, pp. 116–131.

Cunha, Flávio & James J Heckman (2008). "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Journal of Human Resources* 43.4, pp. 738–782.

Cunha, Flávio, James J Heckman & Susanne M Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Econometrica* 78.3, pp. 883–931.

Embrey, Iain Peter (2019a). "A Review of Economic Dual-Self Modelling". Lancaster.

— (2019b). "Education as a Participation Game: An Explanation for Educational Under- Investment". Lancaster.

— (2019c). "On the Benefits of Normalization in Production Functions". Lancaster.

— (2019e). "States of nature and states of mind: a generalized theory of decision-making". In: *Theory and Decision*.

Heckman, James J (2006). "Skill Formation and the Economics of Investing in Disadvantaged Children". In: *Science* 312.5782, pp. 2005–2007.

Hills, J., J. Le Grand & D. Piachaud (2002). *Understanding Social Exclusion*. Oxford: Oxford University Press.

Ockham, William of (1323). “Summa Totius Logicae”. In: *Summa Logicae, 1974*. Ed. by Ph Boehner, G Gal & S Brown. Olean: St. Bonaventure University.

Pissarides, Christopher A (1990). *Equilibrium Unemployment Theory*. 1st. Oxford: Blackwell Publishers Ltd.

— (2000). *Equilibrium Unemployment Theory*. 2nd. Cambridge: MIT press.

Samuelson, Paul A. (1938). “A Note on the Pure Theory of Consumers’ Behaviour”. In: *Economica* 5.17, pp. 61–71.

Samuelson, Paul Anthony & William Dawbney Nordhaus (2009). *Economics*. 19th ed. New York: McGraw-Hill Irwin, p. 702.

Savage, Leonard J (1954). *The Foundations of Statistics*. New York: John Wiley & Sons Inc.

Chapter 2

The Education Trap:

Could a grades-focussed educational system be perpetuating poverty in advanced economies?

Iain P. Embrey,

*Lancaster University**

Abstract:

Existing economic theory does not adequately explain the very low educational investment of disadvantaged children in advanced economies. We propose a new model of educational investment that endogenously separates disadvantaged children into a divergent low-investment equilibrium, and we observe that this education trap could transmit poverty between generations. We show that the education trap arises when a grades-focussed educational system induces children to care about their performance in each of the incremental educational investment decisions that they face on a daily basis, and we respond by deriving tangible policy recommendations that could establish an alternative, learning-focussed system.

JEL Codes: I24; D91; C73; C61; J24.

Keywords: Education; Human Capital; Participation Game; Poverty; Social Exclusion.

*i.embrey@lancaster.ac.uk; Lancaster University Management School, Lancaster, LA1 4YX.
I would like to thank Kim Kaivanto, Steve Bradley, Ian Walker, Colin Green, Geraint Johnes, Emanuele Bracco, David Rietzke, Kwok Tong Soo, and Roy Zilberman for their comments. Any remaining errors are entirely my own.

1 Introduction

Poverty is undesirable. Given the choice between an impoverished existence and earning a comfortable living, very few people would prefer the former. It is therefore puzzling that a substantial minority of children in advanced economies reveal precisely that preference by making minimal investment in their education. Disadvantaged children are disproportionately likely to choose very low levels of educational investment, and in making that choice they frequently become trapped within an intergenerational cycle of poverty (Sparkes & Glennerster 2002; Conti, Heckman & Urzua 2010; Lavecchia, Liu & Oreopoulos 2016). However, despite significant research interest, the question as to why disadvantaged children in advanced economies do not invest more into their education remains open.

The existing literature has investigated two related aspects of that question. The first asks whether disadvantaged children could be investing optimally given their constraints, and the second asks whether behavioural factors could distort normatively optimal investment levels to the observed degree. In Section 2 we survey the existing literature to conclude that neither normative constraints nor behavioural adaptations have yet been shown to provide a viable explanation for observed investment levels. For example, credit constraints rarely bind school attendance in the developed world; there is little evidence that impoverished families underestimate the benefits of schooling; and, though present-bias and behavioural payoffs seem relevant, their magnitude cannot feasibly account for the degree of observed under-investment, which Cunha & Heckman (2008) estimate as equivalent to an unobserved cost in the order of $500,000 for U.S. College attendance. This chapter proposes a new theory of educational investment, in which divergent educational pathways arise as an optimal response to small variations in early-years opportunities.

Our model of a grades-focussed educational system generates a stark bifurcation between high- and low-investment pathways at equilibrium. This result provides the final link in the following outline mechanism for the intergenerational cycle of poverty in advanced economies: i) at an early age, a child's educational opportunities are determined by her parents; ii) disadvantaged parents are likely to provide less frequent educational activities;[1] and iii) the education trap: our results show that children who develop less than a critical threshold level of ability

[1]The literature suggests many potential reasons for this. Disadvantaged parents may have: reduced time per child due to a higher incidence of single-parenthood and higher fertility rates; reduced capability due to the debilitating urgency of poverty; or credit constraints that restrict the provision of educational materials. Our model additionally suggests that disadvantaged parents may have reduced esteem for and knowledge of the educational process, because for them it was rarely optimal to participate in educational tasks given their own early disadvantage.

during their early years should optimally invest very little into their education thereafter. We refer to this result as the education trap because it implies that a grades-focussed educational system will trap disadvantaged children into an intergenerational cycle of poverty.

The education trap arises as a result of two key innovations in our model. Our first innovation is to disaggregate the canonical one-shot investment decision of Becker (1962, 1964) into the series of minor participation decisions that parents and children make on a daily basis. At the earliest ages, such decisions include whether to: talk to the child, play with the child, read with the child, and so forth. Then, as the child develops, she begins to take decisions such as whether to: engage in group activities, attempt classwork tasks, and study for tests. We demonstrate that the equilibrium outcome from this series of incremental investment decisions is meaningfully different to the equilibrium outcome when the same decisions are considered in aggregate. Intuitively, this is because the traditional aggregate approach does not discount the series of immediate costs and benefits which accrue to educational participation.

Our second innovation is to consider the implications of an educational system that places a strong emphasis on its pupils' grades. Recent decades have seen an increasing emphasis on school accountability and inter-school competition in the UK, the US, and elsewhere (Figlio & Rouse 2006; West & Peterson 2006; Bradley Forthcoming), however I know of no existing theoretical analysis of the implications that such grades-focussed education could have on childrens' incentives to study. We model grades-focussed education by assuming that children are taught to care about their performance in each incremental educational opportunity. In our model, children receive a positive payoff for success, and a negative payoff for failure, where these states are determined by a draw from their current cognitive ability level. This models a situation where children evaluate their success based upon any of: informal comparison against peers, formal assessment against grade criteria, or frequent enumeration of their working-at-grade.[2] This contrasts with a learning-focussed model in which success would be based upon whether new skills have been learned in the present period. The essence of a grades-focussed educational system is therefore that it rewards cumulative ability levels rather than valuing learning per se.

Our results suggest that both of these innovations are necessary for the existence of the education trap. That is, for a wide range of parameter values there is a unique high-investment equilibrium whenever agents either: perceive success as

[2]One consequence of the grades-focussed educational system in is that it has become common practice to gradate tasks in schools, at least in the UK: for example as a 'grade B question', or a 'level 3c skill'.

the development of their current ability level through educational participation, or perceive education as a single period with $T+1$ possible participation levels (rather than T periods of binary participation decisions). In contrast, we show that the unique equilibrium of our model of a disaggregated and grades-focussed educational system separates agents based upon their early-years ability development. Specifically, an arbitrarily small change in early development could, at the margin, precipitate a polar reversal in equilibrium investment thereafter. These findings provide the first viable explanation for educational investment levels that fall drastically below their canonically optimal level.

Our contribution builds upon several insights that are already established within the literature. Cunha & Heckman (2007) synthesise the empirical literature to identify six key stylised facts of educational development, and five of these empirical facts are explicable as consequences of divergent developmental pathways. Those authors go on to demonstrate that smoothly divergent pathways can emerge from a model that partitions childhood into more than one developmental period, provided that investments are assumed to be self-productive. We develop their insights by: i) showing that divergent educational investment can arise as a child's rational response to early disadvantage, rather than solely through parental investment decisions; ii) showing that this divergence can be discontinuous, rather than smoothly deviating around the optimal outcome; iii) deriving self-productivity, rather than requiring it as a primitive assumption; and iv) proposing a mechanism for the process of educational investment, rather than treating it as a black box.

The main policy recommendation of existing work is that intervention will be most effective in a child's early years. However, studies that treat educational production as a black box can offer no insight as to the form that such intervention should take. Because we model a mechanism of educational production, we provide a tangible interpretation for the otherwise abstract concept of educational investment. Our model implies that an effective intervention must support disadvantaged children and families to take-up educational opportunities, and that this will require them to both recognise those opportunities and to believe that they can be successful in those opportunities. However, these conditions would be sufficient only to ensure a contemporaneous increase in cognitive ability. For a lasting effect on future educational investment, the intervention would need to be maintained until its recipient reaches the threshold ability level for the high-investment pathway, otherwise low educational investment would remain their equilibrium response. Our results therefore additionally provide a detailed explanation of why interventions are most effective during a child's early years: the low-investment equilibrium diverges further from the threshold ability level as time progresses, such that eventually even sustained participation could no longer propel a disadvantaged child onto the high-investment pathway.

In addition to providing tangible recommendations for individual-level intervention, our model of the mechanism of educational production allows us to identify the systemic factors which generate divergent investment pathways as equilibria. We are therefore able to recommend specific and tangible interventions with the educational system, through which full participation could become the dominant strategy for all children of all ages. The educational system determines the set of parameters that describe each incremental educational opportunity that our agents encounter. In our framework, those parameters comprise: a future payoff to human-capital accumulation, a present cost of participation, a positive payoff from perceived success, and a negative payoff from perceived failure. Although those parameters are exogenous to the child, they represent concrete aspects of an educational opportunity, and so they can be manipulated by policy-makers, by educators, and by parents. For example, if effort and individual learning were valued over attainment then the psychic payoff to success could accrue instead to participation, and if the positive aspects of failure were emphasised[3] then its negative payoff could accrue instead to non-participation. Together, these changes would move our grades-focussed model towards a learning-focussed model, under which high-investment would be the equilibrium response for at least the vast majority of children. Although the concept of intervening with the system rather than the child represents a paradigm shift for some policy-makers, our results suggest that such intervention could meaningfully reduce the persistence of poverty in advanced societies.

Whilst the aforementioned policy implications are novel in both their scope and their specificity, one of the most important implications of our model is moral in nature. At any cross-section, a child's observable ability is endogenously co-determined by her innate ability endowment and by her sequence of educational participation to date. Because current ability influences future participation, those inputs not only interact inseparably within our educational production technology, but ex-post it would be impossible to disentangle their relative contributions due to an initial-conditions problem. However, children can influence neither their genetic endowment nor their early-years educational participation. Thus, since we have shown that apparent under-investment in adolescence could be an equilibrium response to early disadvantage, we must conclude that the victims of poverty may not be wholly responsible for their ostensibly poor educational investment decisions.

The chapter proceeds as follows. Section 2 reviews the existing educational

[3] Failure is positive in that it generates an opportunity for personal development, and also in that it demonstrates that the individual is pushing herself to take on challenging tasks.

investment literature, and finds that the observed extent of under-investment is not yet adequately explained. Section 3 then presents our model, whereafter Subsection 4.1 characterises its analytic solutions, and Subsection 4.2 illustrates those solutions numerically for a robust set of functional form assumptions. Section 5 analyses the origins of educational divergence within our model and then discusses the implications of our results. Section 6 concludes.

2 The Existing Literature

Educational investment decisions determine many individual outcomes. A large body of evidence suggests that the financial returns to education appreciably surpass market rates of return (Cahuc, Zylberberg & Carcillo 2014), that those returns may themselves be surpassed by the non-pecuniary benefits of education (Oreopoulos & Salvanes 2011), and that the social returns to education are probably of comparable magnitude to those personal benefits (McMahon 2004). It is therefore an important objective for economic theory to be able to explain the observation that a substantial minority of individuals drop out of education considerably before it would be optimal for them to do so (Oreopoulos 2007).

Most economic theories of educational investment are built upon the canonical investment model of Becker (1962, 1964). That model yields the elegant and intuitive result that individuals should optimally invest until the marginal cost of further education exceeds its marginal product. This implies that the apparent under-investment of many disadvantaged children could be an optimal response, if they either: possess a particularly low educational productivity, or experience a particularly high participation cost. We assess the evidence for each of these hypotheses in turn.

The first hypothesis lacks empirical support. It was shown as early as Griliches (1977) that the returns to education for observationally less able children are at least as great as those for their more able peers, and that conclusion is now supported by a large body of IV literature in which the LATE for individuals affected by exogenous increases in compulsory schooling often exceeds OLS estimates of the average returns to schooling (Harmon, Oosterbeek & Walker 2000). Thus it is not the case that those children who invest the least in their education do so because of lower productivity.

The second hypothesis has now also been refuted empirically. For an economically rational agent, educational participation costs arise due to credit constraints, however Carneiro & Heckman (2002) determine that such constraints are of minor importance in the developed world, and Jensen (2010) found that they affect only the poorest families in the developing world. These results suggest that the apparent educational under-investment of many disadvantaged individuals in the developed world does indeed represent a normatively suboptimal choice.[4] The challenge is therefore to understand the mechanism behind that choice.

Several economic theories attempt to explain suboptimal educational participation. One possible explanation is that disadvantaged children might under-invest because they underestimate their true returns to education. There is evidence that this may be an important factor in the Dominican Republic (Jensen 2010), but those authors believe that such ignorance is unlikely to be significant in the developed world, and Rouse (2004) finds firm evidence in support of that belief. Nevertheless Lavecchia, Liu & Oreopoulos (2016) survey a large number of nudge-based interventions to find that some succeed in increasing participation by expounding the benefits of post-compulsory eduction, which suggests that incomplete knowledge regarding the returns to education may contribute toward explaining under-investment.

Perhaps the most promising avenue toward explaining educational under-investment is the acknowledgement of behavioural aspects of decision-making. Lavecchia, Liu & Oreopoulos (2016) eloquently articulate the intuition that present-bias could lead to educational under-investment, and studies such as Shoda, Mischel & Peake (1990) have provided convincing experimental corroboration of that hypothesis. Nevertheless, Oreopoulos (2007) estimates the parameters of a standard investment model which incorporates present-bias to find that an implausibly large degree of bias would be necessary to completely explain observed under-investment.

A complementary approach could be to incorporate additional behavioural motivations into the model. For example, Wang & Yang (2003) and Köszegi (2006) include a payoff to self-worth within their agents' objective function, which induces a psychic cost of failure within educational participation decisions and therefore reduces participation. Analogously, Akerlof & Kranton (2002) include a payoff to social identity, and thereby suggest that poorly endowed agents might choose to reduce their educational effort in order to fit in with the 'burnouts'. These approaches each provide useful insights, but once again they seem unlikely to explain the magnitude of observed under-investment, which Cunha & Heckman (2008) estimate to be equivalent to an unobserved cost in the order of $500,000 for U.S. college attendance.

The model presented in Section 3 does predict severe under-investment in education by a subset of individuals. The separating equilibrium of our model is sus-

[4]If this conclusion is correct then one implication is that many early-leavers of education should be expected to later regret that decision. Bridgeland, Dilulio & Morison (2006) corroborate that implication, by establishing that 74% of U.S. high school drop-outs later regret dropping out.

tained by the existence of small psychic payoffs to percieved success and failure, which become consequential when the canonical one-shot educational investment decision is disaggregated into incremental participation decisions. This is because, although the life-cycle returns to educational investment are overwhelming in aggregate, the returns to participation in each incremental educational opportunity are less overwhelming, and when those returns are discounted to the perspective of a young child, that child's psychic payoffs can become instrumental. In particular, those payoffs drive a self-productivity in cognitive ability that leads to path-dependence and divergent outcomes at equilibrium.

Despite the fact that educational investment is inherently incremental, there is surprisingly little economic theory that examines more than a handful of periods of investment decisions. Sjögren & Sällström (2004) and Filippin & Paccagnella (2012) both analyse the many-period case, but neither model incorporates dynamic skill-development. Those papers focus instead on the implications of over- or under-optimism regarding an agent's fixed ability endowment, to reveal that over-optimism leads to greater participation. Some of the most important insights in this area are therefore applications of more general results. For example, Thaler & Shefrin (1981) analyse the conflict between an agent's ex-ante preferences and his extemporary desires, and O'Donoghue & Rabin (1999) analyse the implications of present-bias, both for sophisticated agents who anticipate it, and for naïve agents who only experience it. Those papers derive important stylised facts of inter-temporal decision-making as the result of inconsistent choice criteria across time. By contrast, we show that divergent educational investment decisions can arise for economically consistent agents. Moreover, our qualitative implications are robust to the alternative assumptions of myopia (ex Thaler & Shefrin 1981) and naïvety (ex O'Donoghue & Rabin 1999).

3 The Model

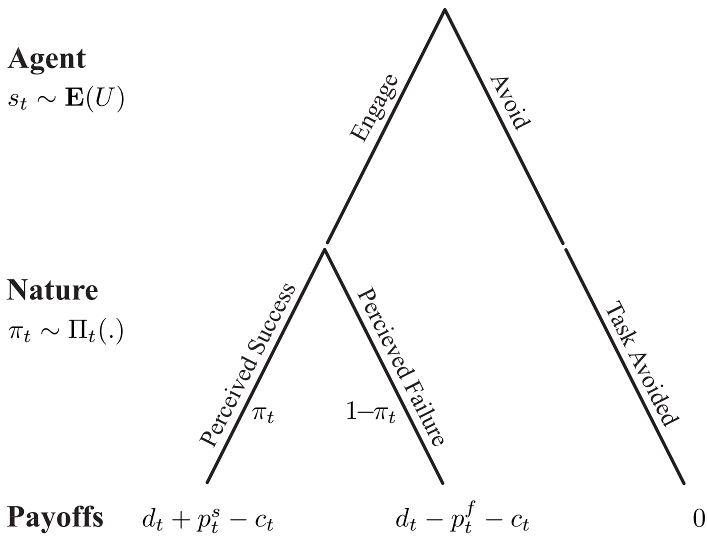
Agents face a series of T educational participation decisions. Their (potentially mixed) strategy space is therefore given by $S := \{s_1, s_2, ..., s_T\}$, where s_t is their chosen probability of participating in the period t opportunity.

Each individual decision is presented as an extensive form participation game in Figure 2.1. The decision utility payoffs that affect educational participation within a grades-focussed educational system are:

$\boldsymbol{d_t}$ the present value of the human capital developed by participation in the task,
$\boldsymbol{c_t}$ the direct and opportunity cost of effortful task participation,
$\boldsymbol{p_t^s}$ the psychic payoff to perceived success, and
$\boldsymbol{p_t^f}$ the psychic cost of perceived failure,

where the subscript $t \in \{1, 2, ..., T\}$ denotes period-specific, or equivalently, task-specific variation. We shall refer to the first two items as the material components of the payoff function, and the final two items as its psychic components. Since these payoffs are formally defined up to affine transformation, we may normalise the payoff of task avoidance to be 0, without loss of generality. It is then uncontentious to further assume that $p_t^s, p_t^f > 0 \ \ \forall t$ – that is: success is pleasant, and failure is unpleasant, ceteris paribus (see, for example: Bénabou & Tirole 2002; Wang & Yang 2003). Although we initially analyse the implications of the model for one representative agent, it will already be evident that individual outcomes must be substantially determined by individual heterogeneity in decision utilities. The implications of such individual heterogeneity are discussed in Section 5.

Figure 2.1: *A Representative Agent's Participation Decision*



As can be seen in Figure 2.1, the agent's probability of achieving perceived success at time t is denoted by π_t. π_t is considered to be a draw from Π_t, the distribution of the agent's probability of success across all possible tasks at time t. Π_t therefore captures the agent's relative cognitive ability in time t. As such, Π_t may be affected by a spectrum of individual and familial characteristics, but it will also be developed through the agents' educational participation in all periods $\tau < t$. In particular, if the agent engages with the present period opportunity then she will develop her cognitive ability such that in the next period her probability of achieving perceived success will increase in expectation, whereas if she avoids that task then her ability level will decrease relative to that of her peers and relative to expected standards. We capture these effects by assuming that Π_{t+1} will stochastically dominate Π_t if the agent engaged with task t, and that Π_{t+1} will be stochastically dominated by Π_t if the agent avoided task t. In this chapter we make the additional simplifying assumption that $\Pi_t(n)$ is uniquely determined by the period, t, and the number of educational tasks completed to date, n.[5] This assumption yields the intuitive and useful lemma that $\mathbf{E}\left(\Pi_t(n)\,|\,n\right)$ is a strictly increasing function of past engagement n, and a strictly decreasing function of t, because past avoidance is given by $t-n-1$. A formal proof of this and all subsequent propositions is provided in the mathematical appendix.

An agent's probability of success distribution $\Pi_t(n)$ therefore parametrises her stock of relative cognitive ability at age t, and so its evolution through time maps out the development of her human-capital. The developmental value d_t of participating in task t is therefore partly derived from its positive effect on the future likelihood of success Π_{t+1}, and partly from its contribution towards the final discounted value in period $T+1$ of having achieved educational attainment level n. This life-cycle payoff, which we denote by $V(n)$, will represent the sum of: expected future remuneration, expected non-pecuniary benefits of education, and the opportunity value of whichever further and higher educational opportunities are accessible to an agent of educational level n. Without loss of generality, we normalise $V(0) := 0$, recognising that in absolute terms $V(0)$ will be affected by factors such as social security policy. $d_t(n)$ is therefore necessarily a function of both the period t, and the current educational level n. We will impose the standard concavity assumptions that life-cycle returns to human-capital are positive but diminishing, that is $V'(n)>0$ and $V''(n)<0$, where $V'(n):=V(n+1)-V(n)$ denotes the first difference of V at n, and $V''(n):=V'(n+1)-V'(n)$ denotes the second difference of V at n. Symmetric concavity conditions are assumed for the expected probability of success, that is: $\mathbf{E}\left(\Pi_t(n)\,|\,n\right)''<0$ and $\mathbf{E}\left(\Pi_t(n)\,|\,n\right)^{\bullet\bullet}>0$, where $\mathbf{E}\left(\Pi_t(n)\,|\,n\right)^{\bullet}$ is the first difference of $\mathbf{E}\left(\Pi_t(n)\,|\,n\right)$ with respect to t at n.

The remaining payoffs are the direct and opportunity cost of effortful task participation c_t, and the psychic payoffs to perceived success or failure p_t^s and p_t^f. These payoffs will be determined by an individual's circumstances and by her psychological traits. In principle, each of these payoffs could therefore also evolve over time, however we will simplify our analysis by treating them as time-invariant characteristics. This approach allows us to isolate the novel implications of our model of grades-focussed educational production, and also to identify cleanly the effects of heterogeneity in those characteristics.

Our model of educational production consists of the supergame formed by T iterations of the stage game described above. To analyse the implications of that supergame we will derive its Bayesian Nash Equilibria, which O'Donoghue & Rabin (1999) refer to as 'perception-perfect' equilibria. This solution concept is

[5]This amounts to an assumption that educational tasks are perfect substitutes, which greatly improves tractability but costs little in generality, since its relaxation would only enhance the path-dependence of our model. In reality this assumption will be true to the extent that teachers and parents are able to differentiate educational tasks to match the current needs of each child.

strong, since it allows agents to choose their entire strategy S without restriction. It is therefore unsurprising that, with probability 1, there is a unique equilibrium for any given set of parameters, information structures, and commitment constraints (see Proposition 1). A weaker solution concept which requires only that each period's strategy s_t should be a best response, holding all other participation decisions s_{-t} constant, would generally produce two distinct weak equilibria, representing high- and low-participation pathways respectively. Proposition 3 exposes the fact that an arbitrarily fine change in initial conditions could determine which of these two divergent pathways will be the unique (Bayesian Nash) equilibrium outcome.

In our benchmark analysis we maintain standard rationality assumptions. In particular: our agents discount future payoffs exponentially; they possess both perfect and complete information about the full supergame (including their current ability distribution Π_t); and they possess no additional information as to the realised state of nature π_t. Given our context of educational decision-making, this assumption set could be considered restrictive. Several possible relaxations are therefore considered by the supplementary materials, where we find that outcomes are qualitatively unaffected by: present bias, under which agents use quasi-hyperbolic discounting (ex Laibson 1997); naïevity, under which agents fail to account for future psychic payoffs (ex O'Donoghue & Rabin 1999); or myopia, under which agents fail to consider any future-period decisions (ex Thaler & Shefrin 1981).[6] It is also reasonable to question whether agents may, in fact, have some prior information regarding the realisation π_t. One might contend, for example, that a child will learn something of the content of each educational task prior to choosing whether to attempt it. Accordingly, Section 2 analyses the case where agents receive a private signal as to their realised probability of success π_t, and thereby confirms that the effect of this adjustment is to introduce a stochastic tremor around the benchmark participation decision.

4 Analyses

In this section we analyse the consequences of a grades-focussed educational system. Subsection 1 presents our analytic results, which begin by establishing that participation in each period will be optimal if and almost only if the agent's ex-

[6] These alternative assumption sets are studied in greater detail by Embrey (2019b), where it is shown that their quantitative effects on the threshold level of cognitive ability that separates low- and high- participation equilibria will be important at the margin. That work also uncovers some interesting interaction effects, not least that increasing an agent's forward-planning ability will only increase her participation if she also possesses sophisticated knowledge of her future psychic payoffs.

pected probability of success exceeds some determinate critical value. This result yields three important corollaries: first, that equilibria will be unique, second, that equilibria will be in pure strategies, and third, that cognitive ability will be self-productive. We then show that this self-productivity leads to a profound path-dependence whereby equilibrium pathways separate dramatically and discontinuously around a well-defined threshold ability level (Proposition 3). Finally, we characterise the separating equilibrium of our model by showing: i) that there exists a sufficient criterion for full participation in all remaining periods, and ii) that if participation is delayed at all then it will be delayed to the maximum possible extent.

Subsection 2 illustrates our results numerically. This approach is useful for visualising our agents' best responses for each possible situation, and it is essential for exploring the stochastic element of our agents' educational development when they have private information.

1 Analytic Results

Our aim is to characterise the full equilibrium solution to our participation supergame. Accordingly, we are interested in the set of strategic best responses $S^* := \{\{\, s_t^*(n)\,\}_{n=0}^{t-1}\,\}_{t=1}^{T}$ for each feasible prior participation level n in each period t. Agents can derive their equilibrium strategy S^* through reverse induction, under the assumption that they will continue to act as fully rational and fully informed economic agents in future periods. We therefore proceed by establishing a recursive formulation for the expected utility payoff of any participation strategy $s_t(n) \in [0,1]$ for all feasible period×prior-participation pairs (t, n).

In period $T+1$ the agent's utility is given by $V(n)$, the discounted present value of the human capital accumulated through participation in n educational opportunities:

$$U_{T+1}(n) = V(n). \tag{2.1}$$

In periods $t = 1, ..., T$, we can therefore derive the expected utility of strategy $S_t^*(n) := \{\{\, s_\tau^*(n)\,\}_{n=0}^{\tau-1}\,\}_{\tau=t}^{T}$ by:

$$\begin{aligned}
\mathbf{E}U_t(n) :=&\ \mathbf{E}\Big(U_t(n, S_t)\,|\, n; S_t\Big) \\
&= (1 - s_t)\Big[\ \delta\mathbf{E}U_{t+1}(n) + 0\ \Big] + \\
&+ s_t\left[\mathbf{E}(\pi_t\,|\,n)\,(\delta\mathbf{E}U_{t+1}(n{+}1) + p^s - c) + (1{-}\mathbf{E}(\pi_t\,|\,n))\,\Big(\delta\mathbf{E}U_{t+1}(n{+}1) - p^f - c\Big)\right] \\
&= s_t\left[\delta\mathbf{E}U_{t+1}(n{+}1) {-} \delta\mathbf{E}U_{t+1}(n) - c {-} p^f + \mathbf{E}(\pi_t\,|\,n)(p^s{+}p^f)\right] + \delta\mathbf{E}U_{t+1}(n),
\end{aligned} \tag{2.2}$$

where δ is the agent's discount rate. From Equation (2.2) we can see that the present value of the human capital developed by participating in the period t task is given by $d_t(n) = \delta \mathbf{E}U_{t+1}(n+1) - \delta \mathbf{E}U_{t+1}(n)$. Given (2.2) we can define $U_t^*(n)$ as the maximum value of the agent's expected utility function for period t given prior educational level n:

$$U_t^*(n) := \max_{S_t} \ \mathbf{E}\Big(U_t(n, S_t) \,|\, n; S_t\Big), \tag{2.3}$$

which is well defined by the extreme value theorem.[7] In each period t the agent's best response $s_t^*(n)$ is therefore given by the value(s) of $s_t(n)$ at which $U_t^*(n)$ is realised, given that the best responses S_{t+1}^* will be played in all future periods.

$$\begin{aligned} s_t^* &= \underset{s_t}{\operatorname{argmax}} \ \ \mathbf{E}\Big(U_t(n, S_t) \,|\, n; s_t; S_{t+1}^*\Big) \\ &= \underset{s_t}{\operatorname{argmax}} \ \ s_t \left[\delta U_{t+1}^*(n+1) - \delta U_{t+1}^*(n) - c - p^f + \mathbf{E}(\pi_t \,|\, n)(p^s + p^f)\right]. \end{aligned} \tag{2.4}$$

The full equilibrium strategy S^* of our participation supergame can be derived by the iterative use of Equations (2.2)-(2.4), from period T back through to period 1. We begin to characterise the properties of this equilibrium in Proposition 1, which first confirms that it is indeed unique:

Proposition 1 *In the finitely repeated game with T periods, and with any non-degenerate probability-of-success distribution:*

1. *For any period t and prior attainment level n the best response $s_t^*(n)$ is unique with probability 1;*

2. *Any Bayesian Nash Equilibrium strategy $S^* \in \{0,1\}^{T(T-1)/2}$ with probability 1;*

3. *In any period t, $s_t^*(n) = 1$ if the agent's expected probability of success exceeds the determinate critical value*

$$\pi_t^*(n) := \frac{c + p^f - \delta\left[U_{t+1}^*(n+1) - U_{t+1}^*(n)\right]}{p^s + p^f},$$

and $s_t^*(n) = 0$ if $\mathbf{E}(\pi_t \,|\, n) < \pi_t^*(n)$.

The first two parts of Proposition 1 state that, with probability 1, the equilibrium response to any situation is a well-defined pure strategy. These results

[7] To see this consider $\mathbf{E}U_t(n)$ as a function of S_t: it is continuous on the closed and bounded domain of $[0,1]^{(T-t+1)(T+t-2)/2}$, and therefore attains its maximum provided that $\mathbf{E}U_{t+1}(n)$ is well-defined. However, we know that $\mathbf{E}U_{T+1}(n) \equiv V(n)$ is well-defined, and so by (2.2) and the principle of induction we have that $\mathbf{E}U_t(n)$ and hence $U_t^*(n)$ are well-defined for all t.

follow as corollaries of the third part, provided that the agent's probability-of-success distribution is continuous, because $\mathbf{E}U_t(n)$ is a monotonic function of s_t (see Equation 2.2) The third part of Proposition 1 derives a well-defined threshold level of ability, above which participation becomes optimal, and below which task avoidance is optimal. The intuition behind this result is clear: our agent will only exert costly effort if she is sufficiently confident that she will achieve a positive benefit from doing so.

The insight developed in Proposition 1 is intuitive, but it is also important. It suggests that, at the extensive margin, agents with higher current ability levels are more likely to participate in future educational opportunities. Thus, since the agent's stock of cognitive ability is itself developed through educational participation, we have established a mechanism by which cognitive ability will be self-productive. Moreover, since the probability that any given participation results in perceived success is itself determined by the agent's current ability level, our production technology for cognitive ability will also exhibit increasing returns to scale at the intensive margin. Proposition 2 formalises these results:

Proposition 2

1. *Cognitive ability is self-productive:*

$$n > m \Rightarrow \mathbf{E}(\pi_{t+1} \,|\, n) \geq \mathbf{E}(\pi_{t+1} \,|\, m) \text{ for all } t,$$

with strict inequality for all but at most one period t provided the marginal utility of present-period participation (as derived in Equation 2.4) satisfies the double crossing property in $(s_t; t)$.[8]

2. *Human-capital development exhibits increasing returns to scale:*

$$n > m \Rightarrow U_t^*(n) > U_t^*(m) \text{ for all } t.$$

The first result demonstrates that our model of the mechanism of educational production endogenously generates self-productivity in cognitive ability: that is higher levels of ability this period beget higher levels of ability next period. Although this is arguably the most fundamental stylised fact of educational production, it has hitherto lacked any theoretical basis. The second result of Proposition 2 is similar in spirit: it demonstrates that higher levels of ability generate higher levels of expected utility. Thus our technology of education production exhibits increasing returns (in the sense of Arthur 1989), which are known to produce: multiple weak equilibria, path-dependence, and inefficient outcomes. In the context

[8]The double crossing property is defined in Appendix A. It excludes pathological cases where $V'(n+T-t)$ and $\mathbf{E}(\pi_t \,|\, n)$ 'wiggle' across each other multiple times within the domain of t.

of educational production, these results provide a theoretical foundation for the importance of early intervention.

We explore the consequences of self-productivity and increasing returns to ability further in Proposition 3. Here we demonstrate that an arbitrarily fine difference in initial conditions could lead to a polar reversal in equilibrium educational investment. Although this property is chaotic in the mathematical sense, it is nevertheless intuitively accessible: I should invest fully if my chances of achieving success are sufficiently high, else I should avoid all costly participation. This result is always true when the life-cycle benefit from human-capital accumulation appears negligible, that is when $V'(n) \approx 0$, and so it provides a useful insight to the decision-making of younger children for whom the end of compulsory schooling could seem imponderably distant.

Proposition 3 *In the absence of life-cycle returns to education, i.e. with* $V'(n) \equiv 0$:

The equilibrium strategy in any period t will be full participation for all periods $\geq t$ if current ability exceeds some determinate threshold, and zero participation for all periods $\geq t$ if current ability is below that threshold.

The insight of Proposition 3 remains relevant when $V'(n) > 0$. This is because any child who embarks upon the high-participation pathway in their early years will remain on that pathway in equilibrium, whilst the period $T+1$ consequences of (non-)participation will only affect the low-participation equilibrium towards the end of compulsory schooling. To see this, consider the effect of exponentially discounting of a payoff $V(n)$ over many years of schooling: even a reasonable annual discount factor of 0.5 would attenuate more than 99.9% from the original value over 10 years.[9] The stark bifurcation that exists between equilibrium pathways for the limiting case of $V'(n) = 0$ is therefore likely to provide a good description of equilibrium behaviour for at least the initial stages of childhood development. Proposition 4 formalises this intuition by establishing that participation will be maximally postponed under any equilibrium of the full model.

Proposition 4

1. *Full-participation equilibria could exist:*
   $\mathbf{E}(\pi_t \,|\, n)(p^s + p^f) - (c + p^f) > 0 \quad \Rightarrow \quad s^*_\tau = 1$ for all periods $\tau \geq t$.

2. *A characterisation of Low-participation equilibria:*
   If ever $s^*_t(n) = 0$, then participation will be postponed so far as possible,

[9]For discussions of empirically determined discount rates see Frederick, Loewenstein & O'Donoghue (2002) and Benhabib, Bisin & Schotter (2010).

that is:

$$s_k^*(n) = 1 \text{ for any } \; k > t \quad \Rightarrow \quad s_\tau^*(n + \tau - k) = 1 \;\text{ for all periods } \tau \geq k.$$

Proposition 4.1 provides a sufficient criterion for full-participation. There are, however, two reasons why this criterion might not be necessary. The first of these is the recently-discussed fact that participation will be made more attractive by the existence of life-cycle payoffs $V(n)$. Although the magnitude of this effect will depend upon the specific functional-form and parameter value assumptions that an agent uses to calculate her expected utility, Section 2 confirms the intuition developed above that it is likely to be qualitatively insignificant except during the final years of compulsory schooling. In those final years, the consequences of (non-)participation become more immediate for our agent, and so she will become increasingly likely to invest in her remaining educational opportunities. The second part of Proposition 4 confirms that the low-participation equilibria that arise from these effects are indeed characterised by a block of consistent task avoidance followed by a block of consistent participation.

The second reason that condition 4.1 might not be necessary is that agents are forward-looking. Thus, if criterion 4.1 does not yet hold, but if it could nevertheless become true given a small amount of costly educational participation, then it might be beneficial for an agent to pay that short-term cost in order to realise a long-term gain. As above, the magnitude of this effect will be determined by the agent's discount rate, but our simulations suggest that it will be qualitatively small. This observation illuminates the important distinction between our model of sequential educational opportunities and the canonical one-shot educational investment decision. When T binary participation decisions are aggregated into a single period with $T{+}1$ possible investment levels, the immediate costs and payoffs of each investment level occur at the same time point. Thus a much larger 'short-term cost' could be sustained because its 'long-term' benefit will be realised at the same time as that cost, rather than discounted over many future periods. Section 1 shows that this distinction has substantial qualitative importance for a robust set of parameter values.

This subsection has analysed the implications of our model of grades-focussed educational production. We have established that its equilibria separate into divergent high- and low-participation pathways, where the former is attained by agents who possess at least a critical threshold level of cognitive ability. Because that prior ability level is itself developed through the early-years educational investment decisions that are made by a child's parents, those parental decisions determine the child's final educational attainment level at equilibrium. Thus our grades-focussed educational system traps disadvantaged children into a low-investment pathway.

We can contrast these results for a grades-focussed educational system with those that would arise under an alternative learning-focussed model. Under that alternative model, any learning that develops a child's current ability level would be recognised as a success, and any avoidance of an educational opportunity would be considered a failure to learn. To see the benefits of such a model, consider the recursive utility formulation that they generate:

$$\mathbf{E}\tilde{U}_t(n) = \; s_t \left[\delta \mathbf{E}\tilde{U}_{t+1}(n+1) - \delta \mathbf{E}\tilde{U}_{t+1}(n) + (p^s + p^f) - c\right] + \delta \mathbf{E}\tilde{U}_{t+1}(n) - p^f. \tag{2.5}$$

Equation (2.5) implies that a sufficient criterion for full participation in all periods would be:

$$(p^f + p^s) - c > 0. \tag{2.6}$$

Intuitively, this criterion states that participation would be guaranteed whenever the psychic payoffs due to engaging with an educational opportunity and learning from it exceed the cost of participation.

Criterion (2.6) is strictly weaker than that of Proposition 4.1, and it is also independent of π_t. Together, these results imply that: if full participation were optimal for the most able child under a grades-focussed model of education, then full-participation would be optimal for all children of all ability levels under our idealised learning-focussed educational model. We can also see that a learning-focussed educational system cannot sustain an eduction trap: low prior participation cannot induce low future investment, because equation (2.5) is also independent of current ability. Of course, even under this idealised model, poverty could still be transmitted between generations through, for example, heterogeneity in participation costs c.

2 A Quantitative Illustration of the Results

We now provide a numerical illustration of our results. Although we necessarily sacrifice generality to do this, the supplementary materials demonstrate that the findings presented here are remarkably robust to a comprehensive set of alternative specifications. Table 2.1 details our preferred specification.

The specification detailed in Table 2.1 provides a tractable model of 'reality'. To achieve this, we assume that agents encounter $T = 1,000$ educational opportunities between birth and the end of compulsory schooling, which is considerably more computationally viable than the 'true' number that might be two orders of magnitude greater. However, the supplementary materials demonstrate that our results are qualitatively indistinguishable from those that arise with either $T = 10,000$ or $T = 100$. The supplement also demonstrates that our results are ro-

Table 2.1: *The Parametric Assumptions for the Model Solved in this Subsection*

Parameter	Assumption	Notes
Number of periods	$T = 1,000$	Robust to parameter variation.
Initial ability distribution	$\Pi_1 \sim Beta[2.5, 2.5]$	As used by Filippin & Paccagnella 2012; Robust to truncated normal; Robust to parameter variation.
Π_t update magnitude	$\iota = 0.05$	Robust to parameter variation.
Value of education	$V(n) = V(T)\left[1-\left(\frac{99}{100}\right)^n\right]$	Robust to parameter variation; Robust to linear form.
Maximum participation benefit	$V(T) = 10,000$	Robust to parameter variation.
Psychic payoffs	$p^s = p^f = 5$	Robust to parameter variation; Robust to asymmetric values.
Participation cost	$c = 1$	Robust to parameter variation.
Discount rate	$\delta = 0.99$	Reasonable cf. † and ‡; Robust to parameter variation.

A detailed discussion and robustness checks are provided in Appendix 2.
† Benhabib, Bisin & Schotter (2010); ‡ Frederick, Loewenstein & O'Donoghue (2002).

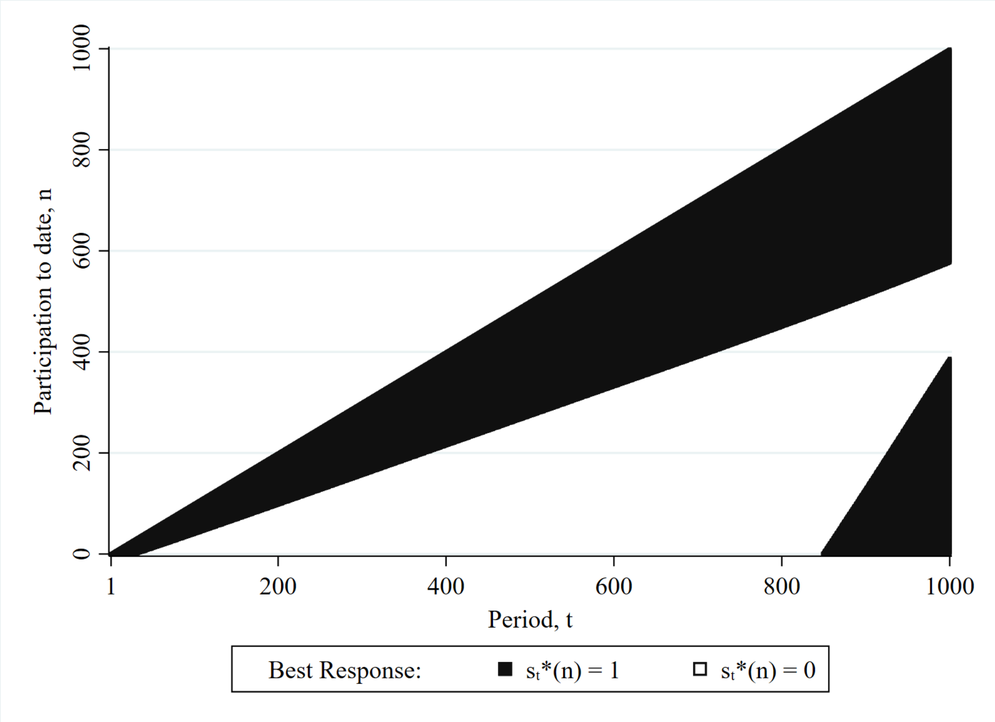
bust to comparable degrees of variation around our other parametric assumptions, and it provides a full rationale for our preferred specification. In summary, our preferred specification implements: exponential discounting with $\delta = 0.99$, for an empirically reasonable annualised discount factor of 0.53; Beta-distributed ability, as is natural for a distribution of probabilities;[10] and a maximum schooling benefit $V(T)$ which is approximately ten times greater than the total material cost $c\times T$ of full educational participation, to reflect the common empirical finding that the net benefit of compulsory education far outweighs its cost.

We now calculate the equilibrium strategy S^* of our model of grades-focussed educational production. S^* specifies the full set of best responses $s_t^*(n)$ to all possible period×prior-participation events $\{(t,n) : n < t\}$. As stipulated by Proposition 1, we find that these best responses are all pure strategies, and so Figure 2.2 illustrates S^* by shading every event (t,n) for which present-period participation is optimal.

Using Figure 2.2 we can trace out the equilibrium path of an agent who is exogenously placed at any event (t,n). If that event is shaded, then the agent will participate in that period, and so traverse North-East to begin the next period with prior participation $n+1$. If that event is not shaded, then the agent will

[10] This is bell-shaped on the support of $(0,1)$ and updates intuitively with positive but diminishing benefits of participation, and symmetric costs of task avoidance.

Figure 2.2: *The Equilibrium Strategy*



The equilibrium strategy S^* for the specification detailed in Table 2.1. All possible events are represented by the set of period×prior-participation pairs (t, n) such that $n < t$. Events that induce equilibrium participation are shaded.

traverse due East to begin the next period with prior participation n. Thus the unique equilibrium path from event (t, n) traverses through $T - t + 1$ line-segments, heading diagonally upward through any shaded regions and horizontally across any unshaded regions. Figure 2.2 therefore illustrates the results from Propositon 4, that participation will either be full or maximally postponed, and the result from Proposition 3, that there is a stark bifurcation between high- and low-participation pathways such that an arbitrarily small difference in initial conditions could lead to a life-changing difference in educational participation.

Let us now consider these findings within the context of a child's development. During the child's early years her educational participation decisions are taken by her parents, so let us suppose for definiteness that the child makes her own investment decisions starting from period 301. The initial event for each child would therefore be $(301, n)$, where the ability endowment $n \in [0, 300]$ is exogenous from her perspective. We can see that each child would optimally respond with full educational participation if she were endowed with sufficient educational opportunity during her early years, and conversely that the optimal response to an insufficient educational endowment would be to avoid all educational opportunities until towards the end of her compulsory schooling. Thus our model of grades-focussed education separates otherwise identical children onto divergent high- and low-participation equilibrium pathways, as a consequence of their early-years educational (dis)advantage.

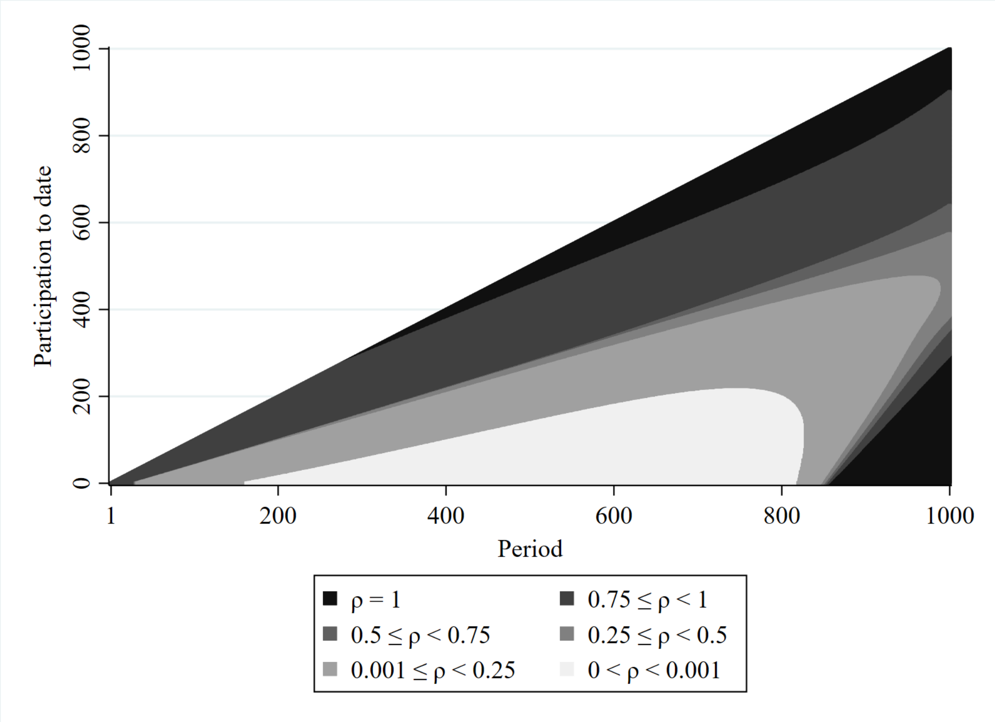
In reality, other forms of heterogeneity are possible. For example, agents may differ in their psychological traits p^s, p^f, their participation costs c, or their innate ability endowments Π_1, but our model shows that none of these differences are necessary to generate divergent equilibrium investment decisions. Thus we cannot reliably infer any conclusions regarding the character or innate intelligence level of any individual merely by observing her level of educational investment. Moreover, if we were to measure Π_t, the current ability of a child during period t, it would be impossible to disentangle the relative contributions of her early-years opportunities (nurture), from her innate ability (nature). Instead we conclude that, whatever the cause of a child's initial disadvantage, its effect is likely to be greatly exacerbated by our model of a grades-focussed educational system.

Figure 2.2 has illustrated the consequences of our analytic results. However, we noted in Section 3 that those analyses were predicated on the strong assumption that agents possess no private information as to their realised probability of success $\pi_t(n)$ for the present-period educational task. In reality, children are likely to gain at least some information as to their likelihood of achieving perceived success from the subject area of any academic task. Since the effect of this private information is bounded below by our benchmark analysis, we now establish an upper bound on its effect, by calculating the equilibrium strategy for agents who can perfectly condition their period t participation upon their realised ability $\pi_t(n)$. To visualise this conditional equilibrium strategy, Figure 2.3 shades each event according to the probability that the realisation $\pi_t(n)$ will be sufficiently high to induce equilibrium participation. We denote this probability by ρ.

From Figure 2.3 we can see that the effect of private information is to induce stochastic variation around the deterministic equilibrium of the benchmark case. Indeed, we can read-off the benchmark equilibrium from Figure 2.3 as its dichotomisation around the participation probability $\rho = 0.5$. However we can also see that the qualitative effect of private information is quite small, in that the stochastic variation around the benchmark equilibrium remains narrow for the majority of the supergame. For example, in period 301 we can see that participation probabilities jump from less than 0.25 to more than 0.75 as a result of a very small change in early-years development. This suggests that, even when private information is complete, its effect will be minimal except close to the margin between high- and low-participation.

To confirm this conclusion numerically, consider the case of an agent who begins making her own decisions from the event $(301, 157)$. Under the benchmark

Figure 2.3: *The Equilibrium Strategy with Private Information*



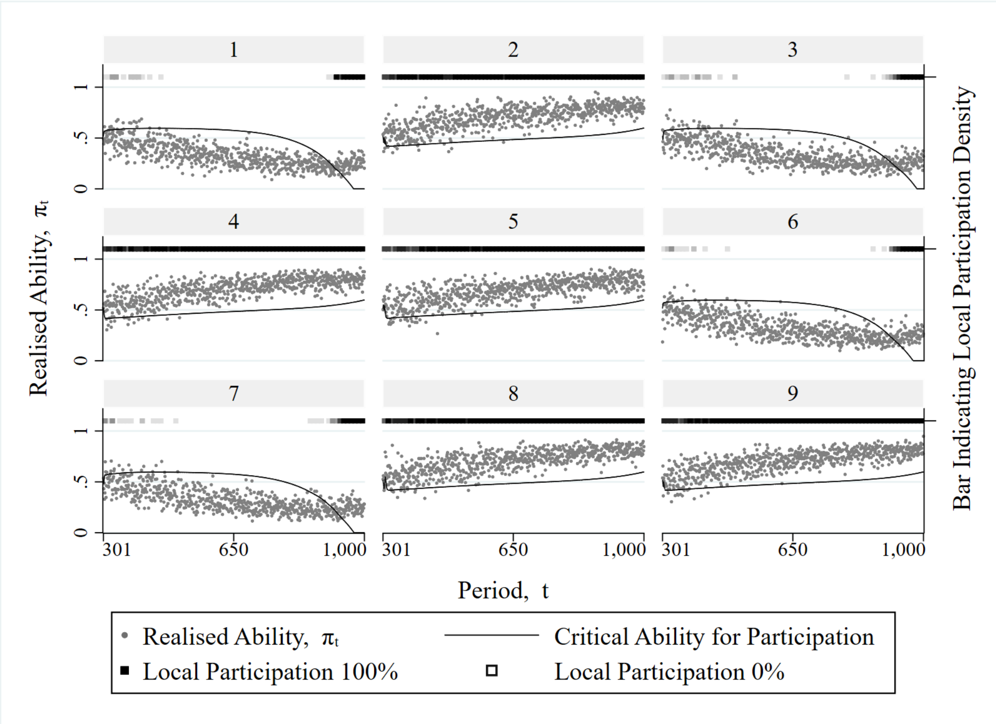
The equilibrium strategy S^* for the specification detailed in Table 2.1, when participation may be conditioned upon the realisation $\pi_t(n)$. Events are shaded according to the probability $\rho_t(n)$ that the agent's realised ability $\pi_t(n)$ would be large enough to induce equilibrium participation.

model, this agent would marginally surpass the threshold for full participation, but with private information the probability that she will participate in period 301 is $\rho_{301}(157) = 0.501$. Nevertheless, the consequences of that participation decision are profound. If the agent does participate in period 301, then she will participate in the next period with probability $\rho_{302}(158) = 0.617$, whereas, if she does not participate in period 301, she will participate in the next period with probability $\rho_{302}(157) = 0.381$. Thus, at the margin between high- and low-participation pathways, an agent who misses a small number of educational opportunities could swiftly diverge from the high-participation equilibrium. By contrast, a short distance away from that boundary the agent's position would be reasonably secure in that the effect of missing any given educational opportunity would be minimal, for example $\rho_{302}(168) - \rho_{302}(167) = 0.005$.

The implications of these numerical findings are illustrated by Figure 2.4, which simulates counterfactual human-capital development processes for 9 identical agents who each depart from the initial event $(301, 157)$. For each period $t \geq 301$, Figure 2.4 plots the realised cognitive ability draws π_t of each agent as grey dots, and it indicates the local participation density of each agent by a shaded bar across the top of her graph. We can therefore see that the relative ability of agents

2,4,5,8,&9 is progressively developed through their frequent educational participation, whilst the relative ability of agents 1,3,6,&7 is eroded by their increasingly sporadic participation in educational opportunities. We conclude from these findings that a grades-focussed educational system is likely to generate divergence in educational outcomes regardless of the degree of private information that children possess.

Figure 2.4: *Simulated Ability Development for 9 Identical Agents*



Simulated relative ability development for 9 representative agents, all of whom are endowed with $n = 157$ in period $t = 301$. Variation occurs because each agent receives a private signal as to her realised probability of success in each period. The bar above each panel is shaded to indicate the local participation density.

Figure 2.4 also provides an insight into the source of the divergence between our simulated agents. To see this, first note that each realised ability draw $\pi_t(n)$ will induce participation if and (almost) only if it is above the critical ability level $\pi_t^*(n)$ that was identified in Proposition 1. Figure 2.4 plots that critical ability threshold as a line, and thereby demonstrates that it generally evolves smoothly over time. In particular, the combined effect of diminishing returns to education and time-consistent discounting is to gradually increase the critical threshold for agents who participate frequently, whilst for agents who participate rarely we can see that these effects generate certain participation as the period $T+1$ consequences of non-participation loom large towards the end of compulsory schooling. But we can also see a marked jump in the critical ability level for all agents in the periods immediately following $t = 301$. This is because those periods are the only time that any agent remains in the neighbourhood of the high-low participation boundary, and in this region the critical threshold is highly sensitive to participation decisions. Thus it is the implications that these marginal agents' first few participation decisions have for their future payoffs that cause them to diverge towards either high- or low-participation pathways.

We have therefore established that private information induces stochastic variation around the benchmark equilibrium outcome, but that both high- and low-participation pathways are strongly attractive. Thus the boundary region wherein an agent's educational development could be qualitatively affected by chance will be relatively narrow, and so a child's early-years development is likely to remain the key determinant of her later developmental pathway. The results presented in Figure 2.4 for agents on that boundary could nevertheless have important implications for pedagogy and for intervention design. They suggest that any educational programme, whether provided by a parent, an educator, or an intervention, is likely to successfully engage a child in learning if and only if she perceives that she is successful during its initial stages.

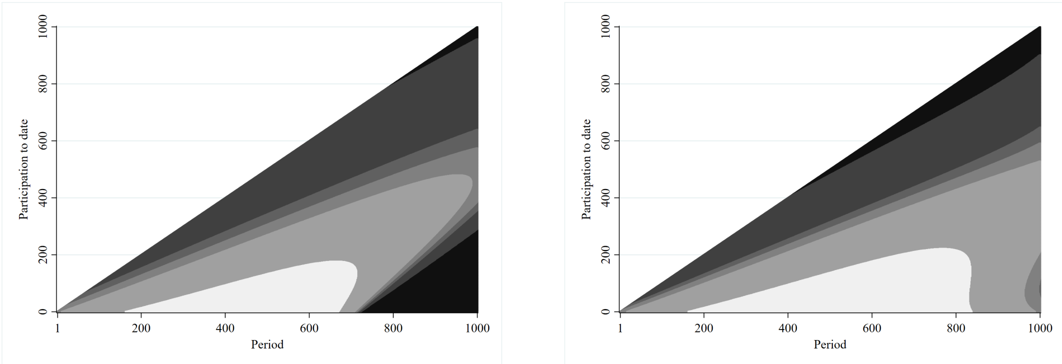
5 Discussion

1 The Origins of Educational Divergence

Our main discussion in Subsection 2 will focus on the implications of our results for policy and for educational practice. However, to ensure that those implications are valid, it is important to first identify the origins of the divergent equilibrium pathways that emerge from our model of educational production. This subsection addresses that need. We confirm here that divergence is generated by the combination of i) disaggregated educational investment decisions, with ii) payoffs that depend on the level, rather than the increment, of human capital.

We begin by ruling out possible alternative drivers for our findings. First, the supplementary materials confirm that our numerical results are not sensitive to changes in the parametric assumptions of Table 2.1, which is unsurprising given their close correspondence with our analytic results. Second, we consider the warning of O'Donoghue & Rabin (1999) that theoretical predictions for intertemporal behaviour may be driven by a sophistication effect that relies heavily on the demanding assumption of perfect information in these circumstances. To investigate this possibility, panel A of Figure 2.5 computes the participation probabilities $\rho_t(n)$ for myopic agents who do not perceive the existence of any future decision periods: by comparing the result to Figure 2.3 we can see that our conclusions are equally valid for myopic agents.[11]

Figure 2.5: *Equilibrium Strategies under Private Information*



(A) *Myopic agents* (B) *No learning if grade bad*

$\rho = 1$	$0.75 \leq \rho < 1$
$0.5 \leq \rho < 0.75$	$0.25 \leq \rho < 0.5$
$0.001 \leq \rho < 0.25$	$0 \leq \rho < 0.001$

These figures can be compared with the the preferred specification in Figure 2.3. They show the equilibrium strategy S^* when participation may be conditioned upon the realisation $\pi_t(n)$, for two alternative cases. In panel A, agents do not perceive the existence of any future decision periods; in panel B agents' relative ability increases if and only if they both participate and obtain a good grade for that participation. Events are shaded according to the probability $\rho_t(n)$ that the agent's realised ability $\pi_t(n)$ would be large enough to induce equilibrium participation.

As an additional robustness check, panel B of Figure 2.5 considers a conceptually different model of educational investment, under which agents might not necessarily be able to learn from educational opportunities. Some economic authors have developed models of learning under which agents may attempt an educational task, but fail to learn from so doing (examples include: Sjögren & Sällström 2004; Filippin & Paccagnella 2012). Such an assumption does not reflect the modern educational setting, in which teachers provide differentiated tasks and ensure that pupils progress, however as an exercise in robustness we also calculate results for a model in which agents learn if and only if they perceive that they have been successful. We can see that this model generates the unrealistic implication that agents with sufficient early-years disadvantage should never exert the effort to participate because they would be unlikely to learn anything even if they tried, but, that apart, we can see that a remarkably similar education trap would also exist under this alternative model.

The foregoing results suggest that properties i) and ii) may be sufficient to induce divergence in educational investment. For the converse, we can be certain that they are both necessary conditions. The necessity of payoffs that depend on grades rather than on learning was established analytically in Section 1, where we

[11]Embrey (2019b) conducts a more detailed investigation into sophistication effects within this model.

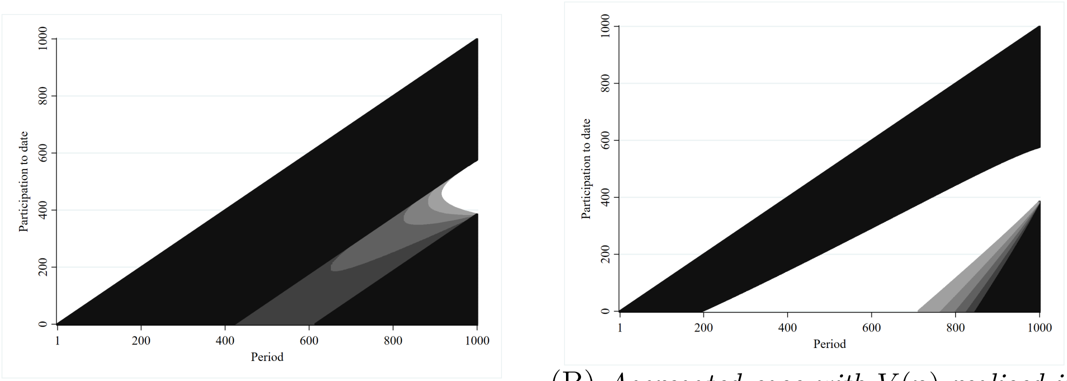
noted that the participation criterion for a learning-focussed system is independent of current ability level. Numerically, every possible event (t, n) would induce equilibrium participation if our benchmark model were adjusted to be learning-focussed, that is if p^s and p^f were to accrue to ability development rather than to grades that are drawn from the agent's current stock of ability.

It therefore remains to investigate whether disaggregation is necessary for divergence in our model. Under an aggregated investment model, an agent at the event (t, n) would perceive her $T-t+1$ remaining participation decisions as a single decision with $T-t+2$ possible investment levels. Hence aggregation has two effects: first, the immediate payoffs of future investment decisions should no-longer be discounted relative to the immediate payoffs of the present investment decision, and second, the period $T+1$ payoffs $V(n)$ should be less discounted relative to (most of) those immediate payoffs. The importance of disaggregation is therefore a numerical question since it depends upon the discount factor δ, although we are able to put a sign on its impact analytically. For the first effect, we noted in Section 1 that time-consistent discounting acts to reduce the likelihood that a disadvantaged agent would accept short-term expected losses to realise long-term expected gains, and for the second effect we note that increasing $V(n)$ unambiguously increases aggregate participation benefit. Thus both effects of aggregation act to increase equilibrium participation.

Panel A of Figure 2.6 establishes that, for our preferred specification, the total numerical effect of disaggregation is substantial. The shading here indicates the equilibrium investment rate r, that is the proportion of the available investment that should optimally be taken-up. We can see that the best response to the majority of initial events (t, n) would be full investment, and that, in particular, all events $(301, n)$ would induce full investment thereafter. Thus, and as expected, aggregation recovers the canonical result that educational investment is overwhelmingly beneficial. It is also possible to evaluate the relative importance of the two aggregation effects, which is achieved in Panel B of Figure 2.6. Panel B maintains the untenable assumption that all immediate payoffs are realised at the start of the aggregated period, to identify the pure effect of removing the relative discount factor between those immediate payoffs. It shows that this first effect of aggregation accounts for around half of the expansion in the high-participation region.

This subsection has therefore established that divergence in educational investment is generated by the combination of disaggregated educational investment decisions with payoffs that depend on the level, rather than the increment, of human capital. Both of these properties are necessary for the education trap to exist, and there is substantial corroborating evidence that their combination may also

Figure 2.6: *Participation Percentage under Aggregation*



(A) *Aggregated case*

(B) *Aggregated case with $V(n)$ realised in period $2T-t+1$*

The high-participation region of these figures can be compared with that of the benchmark specification in Figure 2.2. These figures show the proportion r of available investment that will be chosen in equilibrium, given that all future decision periods are aggregated into one and that the agent makes their decision at the event (t,n).

be sufficient. We now discuss our model's implications for the ways in which policy and practice could increase the educational investment of disadvantaged children.

2 Implications for Intervention Design

The previous section has established that, under the proposed model, educational outcomes are dichotomised into high- or low-participation pathways. Of these, the high-participation pathway is always optimal from the point of view of society, and from the point of view of the child who ex-post enjoys the benefits of high education at the expense of only sunk costs. Since educational investment is inherently incremental, there are two ways in which public policy could intervene to increase educational participation. Intervention could either seek to improve the cognitive ability of disadvantaged children, or it could support all children by manipulating the exogenous parameters of the model.

The most direct design of intervention would aim to improve the cognitive ability of disadvantaged children. This could be effective, because we have established that there generally exists some threshold level of ability above which high participation would become self-sustaining. However we have also established that this threshold ability level swiftly diverges away from the current ability of children on the low-participation pathway (see Figure 2.2 or 2.4), and so if such an intervention is not undertaken very early in the life-course it is likely to prove ineffective. Indeed, such an intervention could even prove counter-productive due to the psychic cost of trying, but failing, to catch up with peers whose ability is also steadily improving.

Our results therefore suggest that an indirect intervention design could be more effective. Although the model's environmental parameters are exogenous for the child, it will generally be possible for parents and teachers to manipulate them. For example: $-c$ could be made positive by the use of sufficiently fun and engaging tasks, or at worst by the imposition of credible sanctions on the outside option. p^s could be increased by agreeing appropriately challenging goals, and by the judicious use of praise and rewards. $-p^f$ could be made less negative, or possibly even positive, by both explicitly teaching and implicitly modelling that failure is positive: because it shows that you are taking on challenges and because it generates learning. Finally, the distant positive payoff of d_t could be made more immediate by emphasising the intrinsic value of developing one's abilities, and the sophisticated extrinsic value that present learning will render future tasks more accessible and therefore more enjoyable. Table 2.2 maps these specific implications onto existing pedagogical practices, thereby demonstrating that these implications articulate the insights of experienced teaching professionals.

Table 2.2: *Mapping the Model's Implications onto Existing Pedagogy*

p^s **accrues to participation**	Praise effort rather than intelligence (e.g. Mueller & Dweck 1998);
	Appropriately challenging goals (e.g. Bandura & Schunk 1981).
$-p^f$ **more positive**	Foster grit and resilience (e.g. Duckworth et al. 2007).
$-c$ **more positive**	Tasks should be engaging (e.g. Christenson, Reschly & Wylie 2012);
	Effective use of sanctions (e.g. Emmer, Everston & Anderson 1980).
d_t **more immediate**	Emphasise the formative use of assessment (e.g. Black & Wiliam 1998);
	Foster growth mindsets (e.g. Blackwell, Trzesniewski & Dweck 2007).

The implications discussed in this section support the conclusion of Heckman (2006) that early intervention is vital if the educational pathway of disadvantaged children is to be altered. However they also caveat that conclusion with the observation that it applies only to traditional, child-focussed interventions. Because we show that low levels of educational investment could arise as an equilibrium response to initial disadvantage, we must conclude that it may not be disadvantaged children who require intervention from policy-makers, but rather the educational system. Accordingly, we have identified specific and tangible ways in which a grades-focussed educational system could be altered to support the educational participation of all children. Our results suggest that these systemic interventions could be substantially more effective than a traditional, child-focussed intervention, and our belief is that they are also likely to be substantially less expensive to implement.

6 Conclusion

This chapter has responded to the increasing emphasis on educational accountability by modelling a grades-focussed educational system. Under this system, children care about their performance in the series of incremental educational opportunities that they encounter on a daily basis. We find that a grades-focussed system generates an education trap, whereby children with low levels of prior educational participation optimally invest very little into their future educational opportunities.

We have shown that the education trap has several important properties. First, it is discontinuous: life-changing differences in equilibrium investment can arise from arbitrarily small changes in initial conditions. Thus our model provides the first viable explanation for the empirical observation that many disadvantaged children make educational investment decisions which are drastically below their normatively optimal level.

Second, children are separated into divergent high- and low-investment equilibria even when they differ only in their early-years educational development. This result closes the cycle of intergenerational poverty transmission, because it implies that early disadvantage will substantially reduce later educational investment at equilibrium. This result also suggests that low levels of investment may represent a child's optimal response to their social environment, and so the observation of low educational investment does not necessarily imply any deficiency in a child's character or innate ability.

Third, the education trap cannot exist under a learning-focussed educational system. This suggests that the policy response to educational divergence should be to intervene with the system, rather than with the children who may be optimally responding to that system. Because our model provides an explicit mechanism for educational production, we have been able to recommend tangible interventions that could move a grades-focussed educational system towards a learning-focussed system. Our results suggest that these systemic interventions could contribute toward a meaningful reduction in the persistence of economic inequality in advanced economies.

A Mathematical Appendix

1 Proof of Lemma 0: $\mathbf{E}(\Pi_t(n))$ is a strictly increasing function of n, and a strictly decreasing function of t.

Since we have the simplifying assumption that $\Pi_t(n)$ is well-defined, we may, without loss of generality, choose the order in which the implied n periods of participation appear within $t-1$ prior periods. For the first result, let us therefore compare $\Pi_t(n)$ with $\Pi_t(m)$, where $m < n < t$, by assigning the first $t-n-1$ periods of both participation sequences to non-participation, and the following m periods to participation. The final $n-m$ periods were therefore either periods of participation to reach $\Pi_t(n)$, or periods of non-participation to reach $\Pi_t(m)$. Thus $\Pi_t(n)$ stochastically dominates $\Pi_t(m)$ by the transitivity of stochastic dominance, hence $\mathbf{E}(\Pi_t(n)) > \mathbf{E}(\Pi_t(m))$.

For the second result compare $\Pi_{t+1}(n)$ with $\Pi_t(n)$ by assigning the first n periods to participation in both sequences, and the remaining periods to non-participation. Thus $\Pi_{t+1}(n)$ is stochastically dominated by $\Pi_t(n)$, hence $\mathbf{E}(\Pi_t(n)) > \mathbf{E}(\Pi_{t+1}(n))$. □

2 Proof of Proposition 1

In any given period $t \leq T$, the equilibrium strategy s_t^* that maximises the agent's expected utility is given by (2.4). We proved in footnote 7 that this is well-defined.

Consider three cases. Firstly, it is possible that the expected benefit of participation, $\left[\delta U^*_{t+1}(n+1) - \delta U^*_{t+1}(n) - c - p^f + \mathbf{E}(\pi_t \,|\, n)(p^s + p^f)\right]$ could be precisely 0, in which case all values for the decision variable provide identical expected utility. However, provided that $\mathbf{E}(\pi_t \,|\, n)$ is continuously distributed (and since we assume $p^s, p^f > 0$), this case occurs with probability 0, and so we do not analyse it further. Otherwise, if $\left[\delta U^*_{t+1}(n+1) - \delta U^*_{t+1}(n) - c - p^f + \mathbf{E}(\pi_t \,|\, n)(p^s + p^f)\right] > 0$, then the optimal strategy is to set $s_t = 1$, and conversely, if $\left[\delta U^*_{t+1}(n+1) - \delta U^*_{t+1}(n) - c - p^f + \mathbf{E}(\pi_t \,|\, n)(p^s + p^f)\right] < 0$, then the optimal strategy is to set $s_t = 0$. Thus, with probability 1, the decision problem in period t has a unique equilibrium response of $s_t \in \{0, 1\}$. (result 1)

For definiteness we may therefore declare, with almost no loss of generality, that agents will participate in a given educational task if and only if:

$$\left[\delta U^*_{t+1}(n+1) - \delta U^*_{t+1}(n) - c - p^f + \mathbf{E}(\pi_t \,|\, n)(p^s + p^f)\right] > 0$$

$$\frac{c + p^f - \delta\left[U^*_{t+1}(n+1) - U^*_{t+1}(n)\right]}{p^s + p^f} < \mathbf{E}(\pi_t \,|\, n). \quad \text{(result 2)}$$

Finally, note that $s_t \in \{0,1\}$ with probability 1 for each of finitely many periods t. The conjunction of these events is that $S^* \in \{0,1\}^T$, and this therefore occurs with probability $\prod_{t=1}^{T} 1 = 1$. (result 3) □

3 Proof of Proposition 2

The first result follows immediately from Lemma 0 if $n > m+1$. If $n = m+1$ we need to consider four cases:

	$s^*_t(m) = 1$	$s^*_t(m) = 0$
$s^*_t(n) = 1$	case i	case ii
$s^*_t(n) = 0$	case iii	case iv

In cases i,ii, and iv the result again follows immediately from Lemma 0, because $n + s^*_t(n) > m + s^*_t(m)$. In case iii $n + s^*_t(n) = m + s^*_t(m)$, and so the basic result holds at least with equality. However, case iii can occur for at most one t if the marginal utility of present-period participation (2.4) satisfies the double crossing property in $(s_t; n)$. To see this, consider the (expected) marginal utility of present-period participation as a function of (s_t, n):

$$\begin{aligned} MU_t(s_t, n) :=& \mathbf{E}\Big(U_t(n, S_t) \,|\, n; s_t; S^*_{t+1}\Big) - \delta\mathbf{E}\Big(U_{t+1}(n, S_{t+1}) \,|\, n; S^*_{t+1}\Big) \\ =& s_t\left[\delta U^*_{t+1}(n+1) - \delta U^*_{t+1}(n) - c - p^f + \mathbf{E}(\pi_t \,|\, n)(p^s + p^f)\right], \end{aligned} \tag{2.7}$$

then define the double crossing property as the natural extension of the single crossing property that is discussed in Edlin & Shannon (1998): For partially ordered sets X and Y, a function $f : X \times Y \to \mathbb{R}$ exhibits the double crossing property in $(x; y)$, if, for all $x_1 > x_0$:

Definition 1 *1.* $f(x_1, y_0) < f(x_0, y_0)$ and $f(x_1, y_1) > f(x_0, y_1)$ for $y_1 > y_0 \Rightarrow$
$f(x_1, y_2) > f(x_0, y_2)$ for all $y_2 > y_1$; and

2. $f(x_1, y_0) \leq f(x_0, y_0)$ and $f(x_1, y_1) \geq f(x_0, y_1)$ for $y_1 > y_0 \Rightarrow$
$f(x_1, y_2) \geq f(x_0, y_2)$ for all $y_2 > y_1$.

Now the participation condition for $s_t^*(m) = 1$ is thus precisely that $MU_t(1, m) > 0$. Hence, if we have the double crossing property, then if there exists any $t_0 < t_1$ for which $MU_{t_0}(1, m) < 0$ and $MU_{t_1}(1, m) > 0$, then there cannot have been any $t_2 > t_1$ for which $MU_{t_2}(1, n_0) < 0$. But proposition 4.2 establishes that once participation becomes optimal it will remain optimal, and so if $MU_t(1, m) > 0$ for all $t \geq t_1$, then $MU_t(1, m + 1) > 0$ for all periods $t > t_1$, and so period t_1 is the unique period in which case iii can occur.

It is sensible to briefly discuss the intuition and viability of the double crossing properties in this context. Intuitively, this property would allow the possibility that $s_t^*(n)$ could cross from 1 to 0 as t increased due to $\mathbf{E}(\pi_t \mid n)$ no-longer exceeding the threshold level given in Proposition 1, and then to cross back from 0 to 1 as t increased further, due, in effect, to the discounted life-cycle payoff $\delta^{T-t+1}V'(n + T - t)$ becoming more consequential than the present participation cost $\mathbf{E}(\pi_t \mid n)(p^s + p^f) - (c + p^f)$. The double crossing property would, however, require that once $\delta^{T-t+1}V'(n + T - t)$ becomes more consequential than $\mathbf{E}(\pi_t \mid n)(p^s + p^f) - (c + p^f)$ it should remain so for all $\tau > t$. This is highly plausible, as its contravention would require $MU_t(1, m)$ to have higher-order roots (i.e. to evolve in a non-smooth manner). Equivalently, the double crossing property requires that $MU_t^\bullet(s_t, n)$ should satisfy the single crossing property in $(s_t; t)$.

For the second result of Proposition 2 we proceed inductively. In period $T + 1$ the result is trivially true, since positive returns to education $V'(n) > 0$ are assumed. For the induction step we again consider the cases i-iv from the proof of the first part of this Proposition.

case iv Here the induction step is trivial, because $U_t^*(n) = \delta U_{t+1}^*(n)$ and $U_t^*(m) = \delta U_{t+1}^*(m)$.

case ii The result follows in this case because $U_t^*(n) \geq \delta U_{t+1}^*(n) > \delta U_{t+1}^*(m) = U_t^*(m)$, where the first inequality follows from $s_t(n)^* = 1$, and the second follows from the induction assumption.

case i In this case,

$$\begin{aligned} U_t^*(n) =& \delta U_{t+1}^*(n + 1) + \mathbf{E}(\pi_t \mid n)(p^s + p^f) - (c + p^f), \text{ and} \\ U_t^*(m) =& \delta U_{t+1}^*(m + 1) + \mathbf{E}(\pi_t \mid m)(p^s + p^f) - (c + p^f). \end{aligned}$$

Here: the first term is larger in the upper line by the induction assumption, the second term is larger by Lemma 0, and the third term is identical.

case iii In this case,

$$\begin{aligned} U_t^*(n) =& \delta U_{t+1}^*(n), \text{ and} \\ U_t^*(m) =& \delta U_{t+1}^*(m+1) + \underbrace{\mathbf{E}(\pi_t \,|\, m)(p^s{+}p^f) - (c{+}p^f)}_{A}. \end{aligned}$$

Here the first term is weakly larger in the upper line by the induction assumption (since $n \geq m+1$), and the term A is negative. To see the latter, note that

$$\begin{aligned} \mathbf{E}(\pi_t \,|\, m)(p^s{+}p^f) - (c{+}p^f) <& \ \mathbf{E}(\pi_t \,|\, n)(p^s{+}p^f) - (c{+}p^f) \\ <& \left[\delta U_{t+1}^*(n+1) {-} \delta U_{t+1}^*(n)\right] + \mathbf{E}(\pi_t \,|\, n)(p^s{+}p^f) - (c{+}p^f) \\ <& \ 0, \end{aligned}$$

where the final inequality is due to $s_t^*(n) = 0$, the penultimate is a consequence of the induction assumption, and the first is true by Lemma 0. □

4 Proof of Proposition 3

Consider again the present-period marginal utility of participation as defined in Equation (2.7):

$$MU_t(s_t, n) = s_t \Big[\underbrace{\delta U_{t+1}^*(n+1) {-} \delta U_{t+1}^*(n)}_{A} \underbrace{-c{-}p^f + \mathbf{E}(\pi_t \,|\, n)(p^s + p^f)}_{B} \Big].$$

Expression A represents the forward-looking component of utility, and is positive by Proposition 2.2. Expression B is the immediate component of utility, and may either be positive or negative. If $B_t \geq 0$ then $B_\tau > 0 \forall \tau \geq t$, because $\mathbf{E}(\pi_{t+k} \,|\, n+k) > \mathbf{E}(\pi_t \,|\, n) \forall k \in \mathbb{N}$ by the transitivity of stochastic dominance. This implies that, if $B_t \geq 0$ then it will be optimal to participate in period t and in all periods thereafter.

Suppose conversely that $B_t < 0$. Then $U_t(n) = V$ can be realised by participating in no future periods (recall that for this Proposition we address the simplified case where $V(n) \equiv V$ because $V'(n) \equiv 0$). Alternative values of $U_t(n)$ could be realised by some combination of participation and non-participation, and it is our goal here to show that any such combination would be strictly dominated by a full-participation strategy. Suppose for contradiction that the optimal strategy $S_t^*(n)$ includes at least one period of non-participation on the equilibrium path (off-path deviation from full-participation is dealt with because this proof does not restrict the initial values of (t, n), and so any (t, n) is trivially on the equilibrium path for the initial conditions that start at that location).

The total utility $U_t^*(n \,|\, S_t^*(n))$ is made up of a series of immediate utility components $\sum_{\tau=t}^{T}\delta^{\tau-1}B_\tau$ plus the constant term V, and in the present situation the first of the $B_t < 0$. In order for the utility $U_t^*(n)$ to exceed that of complete non-participation, we must therefore have that at least one of the $B_\tau > 0$. Note also that, by the first paragraph of this proof, if ever $B_\tau > 0$ then $B_{\tau+k} > 0 \forall k \in \mathbb{N}$. Thus we may split our total utility into a series of negative (or zero) immediate payoffs followed by a series of positive immediate payoffs. The first paragraph also implies that the non-participation period assumed for contradiction must occur during the series of negative (or zero) payoffs. The presence of that period of non-participation therefore reduces the present value of the series of positive payoffs by more than a factor of δ. This is because it not only delays the first period of positive payoffs by (at least) one period (since $\mathbf{E}(\Pi_t(n)) > \mathbf{E}(\Pi_{t+1}(n))$ by Lemma 0), but in doing so it also reduces those payoffs (for the reason just mentioned) and removes the final (most positive) term from that series. The same period of non-participation reduces the magnitude of the series of negative payoffs by a factor of weakly less than δ, because it delays some subset of that series of payoffs, but in doing so it also makes them more negative because $\mathbf{E}(\Pi_t(n)) > \mathbf{E}(\Pi_{t+1}(n))$. Now, since the equilibrium pathway of $S_t^*(n)$ was preferable to non-participation, its original series of positive payoffs must have exceeded in (discounted present) value its series of negative payoffs, and so the insertion of the non-participation period has reduced the value of the former series by an amount strictly larger in absolute value than the amount by which it has increased the value of the latter series. Thus the inclusion of any period of non-participation cannot be optimal unless $S_t^*(n)$ includes no periods of participation at all. $\square$

5 Proof of Proposition 4

The first part of this proposition was proved in the first paragraph of the proof for Proposition 3 – it was only thereafter that the simplification $V'(n) \equiv 0$ was applied.

For the second part of this proposition, we first need to establish the Lemma

$$0 < \mathbf{E}(\pi_t \,|\, n) - \delta\mathbf{E}(\pi_{t+1} \,|\, n) < \mathbf{E}(\pi_{t-1} \,|\, n) - \delta\mathbf{E}(\pi_t \,|\, n). \tag{2.8}$$

To see that this is true, first note that $\mathbf{E}(\pi_t \,|\, n) > \mathbf{E}(\pi_{t+1} \,|\, n)$ by Lemma 0, and so $\mathbf{E}(\pi_t \,|\, n) > \delta\mathbf{E}(\pi_{t+1} \,|\, n)$ a fortiori, since $\mathbf{E}(\pi_{t+1} \,|\, n) > 0$ and $0 < \delta < 1$. Then, for the second inequality, note that

$$\begin{aligned}\mathbf{E}(\pi_t \,|\, n) - \mathbf{E}(\pi_{t+1} \,|\, n) &< \mathbf{E}(\pi_{t-1} \,|\, n) - \mathbf{E}(\pi_t \,|\, n)\\ &< \mathbf{E}(\pi_{t-1} \,|\, n) - \mathbf{E}(\pi_t \,|\, n) + (1-\delta)\Big[\mathbf{E}(\pi_t \,|\, n) - \mathbf{E}(\pi_{t+1} \,|\, n)\Big],\end{aligned}$$

where the first line is true by the concavity assumption $\mathbf{E}\left(\Pi_t(n)\,|\,n\right)^{\bullet\bullet} > 0$, where the second line is true by Lemma 0, and where the second line rearranges to give (2.8).

With the Lemma (2.8) in hand, we now seek to derive a contradiction from an equilibrium subpath of:

$$S_t^* = 0, \underbrace{1, ..., 1}_{k}, 0, ...$$

To do this, consider two permutations of this subpath:

$$S_t^a : \underbrace{1, ..., 1}_{k}, 0, 0, ... \qquad \& \qquad S_t^b : 0, 0, \underbrace{1, ..., 1}_{k}, ...$$

Each subpath results in the situation $(t + k + 2, n + k)$, and so each yields an identical utility over any future participations from period $t + k + 2$ onward. We may therefore compare the total immediate utility contribution $\tilde{U}$ of each subpath accross periods $t - t + k + 2$, and, if $S_t^*(n)$ is indeed the equilibrium outcome, then it should produce the greatest utility over those periods. Now

$$\begin{aligned}
\tilde{U}_t(n, s_T^*) &= \sum_{i=1}^{k+1} \delta^i \Big[\mathbf{E}(\pi_{t+i}\,|\,n{+}i{-}1)(p^s + p^f) - (c + p^f)\Big] \\
\tilde{U}_t(n, s_T^a) &= \sum_{i=1}^{k+1} \delta^{i-1} \Big[\mathbf{E}(\pi_{t+i-1}\,|\,n{+}i{-}1)(p^s + p^f) - (c + p^f)\Big] \\
\tilde{U}_t(n, s_T^b) &= \sum_{i=1}^{k+1} \delta^{i+1} \Big[\mathbf{E}(\pi_{t+i+1}\,|\,n{+}i{-}1)(p^s + p^f) - (c + p^f)\Big],
\end{aligned}$$

and we need

$$0 \leq \tilde{U}_t(n, s_T^*) - \tilde{U}_t(n, s_T^a) \ \ ,$$

which implies that

$$\begin{aligned}
0 &\leq \sum_{i=1}^{k+1} \delta^{i-1} \Big[\delta\mathbf{E}(\pi_{t+i}\,|\,n{+}i{-}1) - \mathbf{E}(\pi_{t+i-1}\,|\,n{+}i{-}1)\Big](p^s + p^f) \ + \delta^k(1{-}\delta)(c + p^f) \\
0 &\leq \sum_{i=1}^{k+1} \delta^{i} \Big[\delta\mathbf{E}(\pi_{t+i}\,|\,n{+}i{-}1) - \mathbf{E}(\pi_{t+i-1}\,|\,n{+}i{-}1)\Big](p^s + p^f) \ + \delta^{k+1}(1{-}\delta)(c + p^f) \\
&< \sum_{i=1}^{k+1} \delta^{i} \Big[\delta\mathbf{E}(\pi_{t+i+1}\,|\,n{+}i{-}1) - \mathbf{E}(\pi_{t+i}\,|\,n{+}i{-}1)\Big](p^s + p^f) \ + \delta^{k+1}(1{-}\delta)(c + p^f) \\
&= \tilde{U}_t(n, s_T^b) - \tilde{U}_t(n, s_T^*).
\end{aligned}$$

Here the second line is true because $\delta > 0$, the third line is true by the Lemma given in (2.8). Thus if S_t^* is indeed preferred to S_t^a, then, in turn, S_t^b must be preferred to S_t^*, and so S_t^* can never be an optimal subpath. $\square$

Chapter References

Akerlof, George A & Rachel E Kranton (2002). "Identity and schooling: Some Lessons from the Economics of Education". In: *Journal of Economic Literature* 40.4, pp. 1167–1201.

Arthur, W. Brian (1989). "Competing Technologies, Increasing Returns, and Lock-In by Historical Events". In: *The Economic Journal* 99.394, pp. 116–131.

Bandura, Albert & Dale H. Schunk (1981). "Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation". In: *Journal of Personality and Social Psychology* 41.3, pp. 586–598.

Becker, Gary S (1962). "Investment in Human Capital: A Theoretical Analysis". In: *Journal of Political Economy* 70.5, pp. 9–49.

— (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* New York: National Bureau of Economic Research.

Bénabou, Roland & Jean Tirole (2002). "Self-Confidence and Personal Motivation". In: *The Quarterly Journal of Economics* 117.3, pp. 871–915.

Benhabib, Jess, Alberto Bisin & Andrew Schotter (2010). "Present-bias, quasi-hyperbolic discounting, and fixed costs". In: *Games and Economic Behavior* 69.2, pp. 205–223.

Black, Paul & Dylan Wiliam (1998). "Inside the Black Box: Raising Standards Through Classroom Assessment". In: *Phi Delta Kappan* 80.2, pp. 139–148.

Blackwell, Lisa S, Kali H Trzesniewski & Carol S Dweck (2007). "Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention in Child". In: *Child Development* 78.1, pp. 246–263.

Bradley, Steve (Forthcoming). "The effects of Quasi-Markets in Education". In: *Economics of Education,* ed. by Steve Bradley & Colin Green. 2nd. Amsterdam: Elsevier.

Bridgeland, John M, John J Jr Dilulio & Karen Burke Morison (2006). *The Silent Epidemic: Perspectives of High School Dropouts.* Tech. rep. Washington: Civic Enterprises & Peter D. Hart Research Associates.

Cahuc, Pierre, André Zylberberg & Stéphane Carcillo (2014). *Labour Economics.* 2nd. Cambridge: MIT press.

Carneiro, Pedro & James J Heckman (2002). "The Evidence on Credit Constraints in Post- Secondary Schooling". In: *The Economic Journal* 112.April 2001, pp. 705–734.

Christenson, Sandra L, Amy L Reschly & Cathy Wylie (2012). *Handbook of research on student engagement.* New York: Springer.

Conti, Gabriella, James J Heckman & Sergio Urzua (2010). "The Education-Health Gradient". In: *American Economic Review* 100.2, pp. 234–238.

Cunha, Flávio & James J Heckman (2007). "The Technology of Skill Formation". In: *The American Economic Review* 97.2, pp. 31–47.

— (2008). "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Journal of Human Resources* 43.4, pp. 738–782.

Duckworth, Angela L et al. (2007). "Grit: Perseverance and passion for long-term goals." In: *Journal of Personality and Social Psychology* 92.6, pp. 1087–1101.

Edlin, Aaron S & Chris Shannon (1998). "Strict Single Crossing and the Strict Spence-Mirrlees Condition: A Comment on Monotone Comparative Statics". In: *Econometrica* 66.6, pp. 1417–1425.

Embrey, Iain Peter (2019b). "Education as a Participation Game: An Explanation for Educational Under- Investment". Lancaster.

Emmer, Edmund T, Carolyn M Everston & Linda M Anderson (1980). "Effective Classroom Management at the Beginning of the School Year". In: *The Elementary School Journal* 80.5, pp. 219–231.

Figlio, David N. & Cecilia Elena Rouse (2006). "Do accountability and voucher threats improve low-performing schools?" In: *Journal of Public Economics* 90.1-2, pp. 239–255.

Filippin, Antonio & Marco Paccagnella (2012). "Family background, self-confidence and economic outcomes". In: *Economics of Education Review* 31.5, pp. 824–834.

Frederick, Shane William, George Loewenstein & Ted O'Donoghue (2002). "Time Discounting and Time Preference: A Critical Review". In: *Journal of Economic Literature* XL.June, pp. 351–401.

Griliches, Zvi (1977). "Estimating the Returns to Schooling: Some Econometric Problems". In: *Econometrica* 45.1, pp. 1–22.

Harmon, Colm, Hessel Oosterbeek & Ian Walker (2000). *The Returns to Education: A Review of Evidence, Issues and Deficiencies in the Literature*. December. London: Centre for the Economics of Education, pp. 1–51.

Heckman, James J (2006). "Skill Formation and the Economics of Investing in Disadvantaged Children". In: *Science* 312.5782, pp. 2005–2007.

Jensen, Robert (2010). "The (Percieved) Returns to Education and the Demand for Schooling". In: *The Quarterly Journal of Economics* 125.2, pp. 515–548.

Köszegi, Botond (2006). "Ego Utility , Overconfidence , and Task Choice". In: *Journal of the European Economic Association* 4.4, pp. 673–707.

Laibson, David (1997). "Golden Eggs and Hyperbolic Discounting". In: *The Quarterly Journal of Economics* 112.2, pp. 443–477.

Lavecchia, Adam M, Heidi Liu & Philip Oreopoulos (2016). "Behavioral Economics of Education: Progress and Possibilities". In: *The Handbook of the Economics of Education,* ed. by Eric A Hanushek, Stephen Machin & Ludger Woessmann. Vol. 5. Amsterdam: North-Holland. Chap. 1, pp. 1–74.

McMahon, Walter W (2004). "The Social and External Benefits of Education". In: *International Handbook on the Economics of Education.* Ed. by Geraint Johnes & Jill Johnes. Cheltenham: Edward Elgar Publishing Ltd. Chap. 6, pp. 211–259.

Mueller, Claudia M & Carol S Dweck (1998). "Praise for Intelligence Can Undermine Children's Motivation and Performance". In: *Journal of Personality and Social Psychology* 75.1, pp. 33–52.

O'Donoghue, Ted & Matthew Rabin (1999). "Doing It Now or Later". In: *The American Economic Review* 89.1, pp. 103–124.

Oreopoulos, Philip (2007). "Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling". In: *Journal of Public Economics* 91.11-12, pp. 2213–2229.

Oreopoulos, Philip & Kjell G Salvanes (2011). "Priceless: The Nonpecuniary Benefits of Schooling". In: *Journal of Economic Perspectives* 25.1, pp. 159–184.

Rouse, Cecilia Elena (2004). "Low-income students and college attendance: An exploration of income expectations". In: *Social Science Quarterly* 85.5 SPEC. ISS. Pp. 1299–1317.

Shoda, Yuichi, Walter Mischel & Philip K. Peake (1990). "Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions". In: *Developmental Psychology* 26.6, pp. 978–986.

Sjögren, Anna & Susanna Sällström (2004). "Trapped, Delayed and Handicapped". In: *CEPR Discussion Paper Series.* CEPR Discussion Paper Series, No. 613. London: Centre for Economic Policy Research.

Sparkes, Jo & Howard Glennerster (2002). "Preventing Social Exclusion: Education's Contribution". In: *Understanding Social Exclusion.* Ed. by J. Hills, J. Le Grand & D. Piachaud. Oxford: Oxford University Press. Chap. 11, pp. 178–201.

Thaler, Richard H. & Hersh M. Shefrin (1981). "An Economic Theory of Self-Control". In: *Journal of Political Economy* 89.2, pp. 392–406.

Wang, X. Henry & Bill Z. Yang (2003). "Why Competition may Discourage Students from Learning? A Behavioral Economic Analysis". In: *Education Economics* 11.2, pp. 117–128.

West, Martin R. & Paul E. Peterson (2006). “The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments”. In: *Economic Journal* 116.510, pp. 46–62.

Chapter 3

Re-Estimating the Technology of Cognitive and Noncognitive Skill Formation

Iain P. Embrey,

*Lancaster University**

Abstract:

Cognitive and noncognitive skills determine many life outcomes, and so there has been significant research interest into the technology by which they are produced. We advance the state-of-the-art in this field by building upon the definitive contribution of Cunha, Heckman and Schennach (*Estimating the Technology of Cognitive and Noncognitive Skill Formation*; Econometrica, 2010). We develop that contribution by normalizing its production function, by nesting that production function, by relaxing over-identifying restrictions, and by amending several minor discrepancies within the code. These developments greatly improve the fit of the model, and they reverse some counter-intuitive findings that formerly emerged from it.

JEL Codes: I21;J24;C38

Keywords: Human Capital; Cognitive Skills; Noncognitive Sills; Education; Factor Model; CES Production Technology.

*i.embrey@lancaster.ac.uk; Lancaster University Management School, Lancaster, LA1 4YX.
I would like to thank would like to thank Kim Kaivanto, Steve Bradley, and economics department members at the University of Lancaster for their comments. Any remaining errors are entirely my own.

1 Introduction

The productivity of economic agents is determined by their stock of human capital. It is therefore a central goal of economic science to understand the technology by which that human capital is produced. Such knowledge would help children and adults to develop their productivity, earnings, and welfare; it would help policy makers to provide effective intervention where necessary; and it would help educators and parents to support the next generation. Nevertheless, the economic literature has only recently begun to model the technology of cognitive and noncognitive skill production through multiple stages of childhood development.

The current economic model of childhood skill formation is developed in a series of papers by James Heckman and Flávio Cunha. Cunha & Heckman (2007) demonstrate that a multi-stage model of cognitive and noncognitive skill formation can explain several important stylised facts of childhood development that are not consistent with the traditional single-period theory of educational investment. Cunha & Heckman (2008) contribute the first econometric estimation of that theoretical model, and Cunha, Heckman & Schennach (2010, hereafter CHS) develop that contribution by incorporating a CES skill production function to relax the assumption that educational inputs are perfectly substitutable across time and across input type. CHS also incorporate: a factor-analytic model of measurement error, a structural investment equation to correct for endogenous investment decisions, and an anchoring equation for latent factor scores, to address their inherently arbitrary location and scale.

In this chapter we develop the analyses of CHS in several important respects. First, we identify that the anchoring described above implicitly estimates the CES production function along an uninformative ray through the production space. We correct this by explicitly normalizing the CES function around sample mean input levels. By implementing this correction we find that early cognitive development is substantially more dependent upon parental investments than was previously thought, whilst early noncognitive ability is not so much determined by the investments that parents make as by their noncognitive and cognitive skill levels. These results suggest that future CES estimates of the human capital production function should explicitly consider the ray along which share parameters are computed.

Second, we relax the over-identifying restriction imposed by re-normalizing factor loadings in each period. This problem was identified by Agostinelli & Wiswall (2016b), who provide a full discussion of its implications. We also make a number of other minor improvements to the model and to the code used by CHS, and these are detailed in Appendix A. Appendix B provides evidence that these improvements are statistically important by comparing our model's auxiliary equation estimates to those of CHS. Overall, Table 3.3 shows that these developments produce a highly significant improvement in model fit.

Third, we implement a more flexible specification for the technologies of cognitive and noncognitive skill formation. By nesting CES production functions, we allow the elasticity of technical substitution to vary between different input pairs. Our results suggest that the more restrictive non-nested form meaningfully distorts the estimated elasticities of substitution between specific input types. This approach also allows us to contribute the first unbiased estimation of dynamic complementarity in inter-temporal investment, for which our results provide moderate support. These results therefore provide further evidence that intervention policies should target the early stages of childhood cognitive development.

The chapter proceeds as follows. Section 2 reviews the existing literature on skill formation, to establish that Cunha, Heckman & Schennach (2010) still represents the state of the art in that literature. Section 3 reprises their model and outlines the developments contributed by this chapter: Subsections 2, 3, and 4 discuss production function nesting, normalisation, and over-identification respectively. Section 4 then discusses the results of our re-estimation of the technology of skill formation, and Section 5 concludes.

2 The Literature on Skill Formation

An expanding literature now builds on the work of Cunha and Heckman to estimate the technology of skill formation. A central aim of that literature is to establish which of three key properties might characterise skill formation. Those properties are commonly referred to as self-productivity, cross-productivity, and dynamic complementarity, however there are inconsistencies in the usage of each term. We therefore proceed by first proposing a standardised definition for each property, before discussing the evidence for each property in the next subsection.

Self-productivity is the property whereby the next-period level of a skill is an increasing function of its present-period stocks. Analogously, cross-productivity is the property whereby the next-period level of one skill type is an increasing function of the present-period stocks of another skill type: for example next-period cognitive ability may be positively affected by present-period noncognitive ability. We say that there is dynamic complementarity of investment when the efficacy of present-period investment (in percentage terms) is an increasing function of past investment levels. This is true if and only if the elasticity of intertemporal substitution between investments is less than 1 (proof in Appendix C). In practice, dynamic complementarity could equivalently be defined as the property whereby present-period investment is a complement to present-period ability stocks, since the effect of investment is universally positive in the literature. Dynamic complementarity between investments will therefore be present if and only if the elasticity of technical substitution between investments and skills is less than 1.

Several existing papers use alternative definitions of dynamic complementarity. Some studies conflate the concept with cross-productivity, and some stipulate that the cross-partial derivative of production with respect to ability and investment be positive. The latter condition requires the marginal product of investment to be an increasing function of present ability stocks in absolute, rather than percentage, terms. The reasons why complementarity is best defined through elasticities are well known, but it is useful to illustrate them with an example. Consider a Cobb-Douglas production function. Cross-productivity is present, since the technology is an increasing function of all relevant inputs, and similarly the cross-partial derivative is positive, except for pathological parameter values. Nevertheless, we do not consider Cobb-Douglas inputs to be complements: intuitively, the percentage increase in output caused by an increase in any given input is independent of the level of all other inputs, and mathematically the elasticity of technical substitution between inputs is 1.

It may sometimes be useful to define complementarity between skill types with an analogous elasticity condition to that for dynamic complementarity between investments. If doing so it is important not to conflate those two properties. Whilst it is true that they necessarily coincide under a CES production function, that coincidence is an artificial restriction imposed by the assumption of a single complementarity parameter. Several alternative functional forms exist which relax that restriction. Functional forms for the technology of skill formation are compared in Section 2.2, whereafter Section 3.1 provides formal mathematical definitions for the properties described here. We now examine the existing empirical support for each of those properties.

1 Our Knowledge of the Technology of Skill Formation

Table 3.1 organises the literature on skill formation. It outlines the econometric specifications used by each study, and summarises their results regarding the aforementioned key properties of the technology of skill formation. Results are coded: ● where they support the relevant property, ⊛ where they provide limited support for it, ○ where they do not support it, and 'N/A' where their methodology cannot accommodate that property.

It is immediately apparent from Table 3.1 that self-productivity is a robust characteristic of cognitive and noncognitive skill formation. Moreover, each of the studies listed here finds that skill formation is increasingly characterised by

Table 3.1: *Studies that Estimate the Technology of Skill Formation*

	Production:			Results:‡		
	Linear	CES	Inputs:†	Self-Prod.	Cross-Prod.	Dynam. Comp.
Cunha & Heckman (2008)	●		Cog; Noncog; Invest; P.Cog; P.Noncog.	●	⊛	N/A
Cunha, Heckman & Schennach (2010)		●	Cog; Noncog; Invest; P.Cog; P.Noncog.	●	⊛	⊛
Helmers & Patnam (2011)	●		Cog; Noncog; Health; Invest; Schooling.	⊛	⊛	N/A
Reuß (2011)	●		Cog; Noncog×2; Invest×many.	●	●	N/A
Coneus, Laucht & Reuß (2012)	●		Cog; Noncog×2; Invest.	●	⊛	N/A
Nicoletti & Rabe (2014)	Linear-interacted		Cog; Invest; Schooling.	●	N/A	⊛
Fiorini & Keane (2014)	●		Time Use ×many; Current Skill.	●	N/A	N/A
Attanasio et al. (2015)		●	Cog; Noncog; P.Cog; P.Noncog; Invest×2.	●	●	○
Attanasio, Meghir & Nix (2015)		●	Cog; Health; Invest; P.Cog; P.Health.	●	⊛	⊛
Attanasio et al. (2017)	Nested CES		Cog; Health; Invest; P.Cog; P.Health.	●	●	⊛
Pavan (2016)	●		Cog; Invest; P.Cog.	●	N/A	N/A
Agostinelli & Wiswall (2016a)	Translog		Cog; Invest.	●	N/A	○
Hernández-Alava & Popli (2017)	●		Cog; Noncog; Invest×2.	●	⊛	N/A
Heredia (2017)	Cobb-Douglas		Cog; Invest×2.	●	N/A	N/A

†Possible inputs are: Cognitive Skills, Noncognitive Skills, Health, Parental Investment, Parental Cognitive Skills, Parental Noncognitive Skills, Parental Health, and Schooling. 'Noncog×2' indicates that two separate aspects of Noncognitive ability are included. Studies typically also control for some covariates.
‡ Results are coded as: ● where they support the relevant property, ⊛ where they provide limited support for it, ○ where they do not support it, and not applicable (N/A) where their methodology cannot accommodate that property.

self-productivity as children become older.

The evidence for cross-productivity between skill types is less consistent. Most studies find that cross-productivity is concentrated in the early stages of childhood development, although Reuß (2011) and Coneus, Laucht & Reuß (2012) find an opposite trend. This is probably because those studies split noncognitive skill into mental and emotional skills, where the former is rather more aligned with cognitive ability than the standard operationalisation of noncognitive skill. Those studies are also unique in finding consistent evidence for bi-directional cross-productivity between cognitive and noncognitive skills; in the remaining literature cross-productivity is generally uni-directional, although there is no clear trend as to which of cognitive and noncognitive skills begets the other.

The existing evidence regarding cross-productivity between health and cognitive skill in the developing world is clearer: health seems to be an important determinant of early cognitive development (Attanasio, Meghir & Nix 2015; Attanasio et al. 2017). Attanasio et al. (2017) is also the only existing study to separately identify complementarity between skill types: their point estimates suggest that health and cognition are strongly complementary in Ethiopia and strongly substitutable in Peru, although these estimates are very imprecise. The inter-country differences are probably attributable to the way in which their latent health factor is derived from height, weight, and self-reported health: the factor loadings of self-reported health are small in both cases, and generally opposite in sign between the two countries. It seems that being large could be a sign either of good health or of poor health, with the former signal dominant in the Peruvian sample and the latter dominant in the Ethiopian sample.

Table 3.1 reveals mixed findings regarding dynamic complementarity. It is most strongly supported by the results of Nicoletti & Rabe (2014), which demonstrate that school investment in teaching staff complements cognitive ability, but that other school investments substitute for cognitive ability. These contrasting findings hint at a possible cause for the inconsistent results elsewhere in the literature. A CES production technology imposes a single elasticity of substitution across all of its inputs. Since parental investment typically contributes a relatively small share of aggregate production, the underlying inter-temporal elasticity of substitution for investment may not be easily inferred from the global CES complementarity parameter. In Section 3.2 we show how this challenge can be overcome by nesting CES production functions, in an approach similar to that used by Attanasio et al. (2017) to identify complementarity between health and cognition.

Several of the papers listed in Table 3.1 make contributions to our knowledge of skill formation which go beyond the results tabulated there. For example Pavan (2016) extends the standard framework to establish that the well-known birth-order effect, whereby first-born children exhibit greater cognitive development than their siblings, is substantially explained by differential parental investment levels in early childhood. Agostinelli & Wiswall (2016a,b) make an important methodological contribution, first by demonstrating that commonly-imposed factor-loading assumptions are over-identifying, and second by proposing a novel estimation strategy that can identify returns to scale and total factor productivity in the technology of skill formation. These contributions are discussed further in Section 3.3, where we suggest that the underlying conditions for that estimation strategy may be untenable in practice.

Fiorini & Keane (2014) make a unique contribution by modelling skill formation as a function of childhood time use. Their results suggest that: educational activities with parents have a highly positive effect on both cognitive and non-cognitive development; educational activities with others are similarly beneficial for cognitive development but provide little benefit for noncognitive development; after-school activities are beneficial for reasoning skills but less beneficial for vocabulary knowledge; and conversely that media use is beneficial for vocabulary but not for reasoning. The policy implications of those findings are clear, however further research in this vein is needed to ensure that they are not driven by reverse causality. For example, are children who make poor progress more likely to be detained after school, and if a child's behaviour deteriorates are their parents more likely to send them to others for educational activities? Further research might also seek to embed detailed investment data from Fiorini & Keane (2014) into a more complete model of skill formation, in order to identify whether those investments are mutually and/or dynamically substitutable.

2 Functional Forms for the Technology of Skill Formation

A variety of functional forms for the technology of skill formation are used in the literature (see again Table 3.1). Linear forms are restrictive, because they assume perfect substitutability between inputs and because they assume constant marginal returns to inputs, both of which are robustly refuted by the literature. Those limitations could be overcome by including higher-order terms and their interactions, however even including second-order terms would require $k.(k{+}3)/2{-}1$ free parameters for a k-input technology,[1] which could soon over-fit most available data.

The Cobb-Douglas form ameliorates the restriction of constant marginal returns, but it replaces the restriction of perfect substitutability with one of perfect independence. As above, the latter restriction could be overcome by including interaction terms. Doing so would generate a Translog specification with $k.(k{+}1)/2{-}1$ parameters, a total which could feasibly be reduced further if some interactions were empirically unimportant. However, a major disadvantage of the Translog

[1]When unit total-factor productivity and returns-to-scale are assumed; see the discussion later in this Section.

form is that its parameters are not readily interpretable.

One practical advantage of the CES production function is that its parameters give directly the elasticity of substitution between inputs, and the production share of each input (provided that the function is appropriately normalised; see Section 3.3). The CES form also has the theoretical advantage that it nests Linear and Cobb-Douglas forms as special cases, and that the Translog form is a restriction of its second-order Taylor expansion around the normalisation point (Embrey 2019c). The CES function represents a parsimonious specification with k free parameters, however its main restriction is that it imposes the same elasticity of substitution parameter across inputs. That restriction can nevertheless be relaxed by nesting CES functions at a cost of one parameter per nest level. Our preferred specification is therefore a CES function, normalised to sample mean input levels and nesting investment to separately identify its level of dynamic complementarity with current skill levels. This specification is now formally set out.

3 The Model

1 A Framework for Skill Formation, and its Key Characteristics

This subsection begins with a brief reprise of the underlying framework of skill formation, as developed by Cunha & Heckman (2007) and Cunha & Heckman (2008). We denote the vector of child i's cognitive and noncognitive skills at time t as $\theta_{i,t} = (\theta_{i,t}^{C}, \theta_{i,t}^{NC})$. Similarly, we denote $\vartheta_i = (\vartheta_i^C, \vartheta_i^{NC})$ as a vector of parental cognitive and noncognitive skills, which we model as time-invariant. Our final input is $I_{i,t}$: the level of parental investment in child i at time t. The technology of skill formation is therefore given by

$$\theta_{t+1} = f_t(\theta_t, I_t, \vartheta) = f_t(\theta_t^C, \theta_t^{NC}, I_t, \vartheta^C, \vartheta^{NC})$$

Self-productivity in skill $k \in \{C, NC\}$ is therefore present whenever:

$$f^k_{\theta^k,t} = \frac{\partial \theta^k_{t+1}}{\partial \theta^k_t} > 0.$$

The corresponding condition for cross-productivity of skill $-k$ on skill k is therefore:

$$f^k_{\theta^{-k},t} = \frac{\partial \theta^k_{t+1}}{\partial \theta^{-k}_t} > 0.$$

Dynamic complementarity of investment in skill k is considered to be present whenever the elasticity of technical substitution between current ability and current parental investment is less than 1:

$$e^k_{\theta I,t} = \frac{\partial\left(\theta_t/I_t\right)}{\partial\left(f^k_{I,t}\Big/f^k_{\theta,t}\right)} \cdot \frac{\left(I_t/\theta_t\right)}{\left(f^k_{\theta,t}\Big/f^k_{I,t}\right)} < 1.$$

So long as the effect of investment is positive, the above condition is true if and only if the elasticity of inter-temporal substitution for investment is positive. Moreover, Appendix C shows that it is also equivalent to requiring the effect of investment (in percentage terms) to be an increasing function of current ability stocks.

2 The Econometric Model and Estimation Procedure

In principle, the technology f_t could vary between periods, between skills, and between individuals. However, our object of interest is a generalised technology across individuals, upon which we must impose some tractable functional form. We will estimate two model specifications: Model (3.1) reproduces the CES production function of CHS except that it makes the normalisation point Θ_t explicit (see Section 3.3), whilst our preferred specification (3.2) is the simplest nesting of CES functions that can identify dynamic complementarity of investment. For parsimony, we follow CHS in fixing f_s across each of two stages of development $s \in \{1,2\}$.

$$f^C_{i,t} = \left(\gamma^C_{s,1}\left(\frac{\theta^C_{i,t}}{\Theta^C_t}\right)^{\varphi^C_s} + \gamma^C_{s,2}\left(\frac{\theta^{NC}_{i,t}}{\Theta^{NC}_t}\right)^{\varphi^C_s} + \gamma^C_{s,3}\left(\frac{I_{i,t}}{\Theta^I_t}\right)^{\varphi^C_s} + \gamma^C_{s,4}\left(\frac{\vartheta^C_i}{\Theta^{PC}_t}\right)^{\varphi^C_s} + \gamma^C_{s,5}\left(\frac{\vartheta^{NC}_i}{\Theta^{PNC}_t}\right)^{\varphi^C_s}\right)^{\frac{1}{\varphi^C_s}} e^{\eta_{i,t,C}}$$

$$f^{NC}_{i,t} = \left(\gamma^{NC}_{s,1}\left(\frac{\theta^C_{i,t}}{\Theta^C_t}\right)^{\varphi^{NC}_s} + \gamma^{NC}_{s,2}\left(\frac{\theta^{NC}_{i,t}}{\Theta^{NC}_t}\right)^{\varphi^{NC}_s} + \gamma^{NC}_{s,3}\left(\frac{I_{i,t}}{\Theta^I_t}\right)^{\varphi^{NC}_s} + \gamma^{NC}_{s,4}\left(\frac{\vartheta^C_i}{\Theta^{PC}_t}\right)^{\varphi^{NC}_s} + \gamma^{NC}_{s,5}\left(\frac{\vartheta^{NC}_i}{\Theta^{PNC}_t}\right)^{\varphi^{NC}_s}\right)^{\frac{1}{\varphi^C_s}} e^{\eta_{i,t,NC}}$$

$$\text{where:} \quad \sum_{j=1}^{5}\gamma^k_{s,j}=1; \quad \varphi^k_s \in [-\infty, 1]; \ \ \forall s;\ k\in\{C, NC\}. \tag{3.1}$$

$$f^k_{i,t} = \left(\gamma^k_{s,7}\left[\gamma^k_{s,6}\left(\gamma^k_{s,1}\left(\frac{\theta^C_{i,t}}{\Theta^C_t}\right)^{\psi^k_s} + \gamma^k_{s,2}\left(\frac{\theta^{NC}_{i,t}}{\Theta^{NC}_t}\right)^{\psi^k_s}\right)^{\frac{\varphi^k_s}{\psi^k_s}} + \gamma^k_{s,3}\left(\frac{I_{i,t}}{\Theta^I_t}\right)^{\varphi^k_s}\right]^{\frac{\phi^k_s}{\varphi^k_s}} + \gamma^k_{s,4}\left(\frac{\vartheta^{PC}_i}{\Theta^{PC}_t}\right)^{\phi^k_s} + \gamma^k_{s,5}\left(\frac{\vartheta^{NC}_i}{\Theta^{PNC}_t}\right)^{\phi^k_s}\right)^{\frac{1}{\phi^k_s}} e^{\eta_{i,t,k}}$$

$$\text{where:} \quad \gamma^k_{s,1}+\gamma^k_{s,2}=1; \quad \gamma^k_{s,3}+\gamma^k_{s,6}=1; \quad \gamma^k_{s,4}+\gamma^k_{s,5}+\gamma^k_{s,7}=1; \quad \forall s;\ k\in\{C, NC\};$$
$$\text{and where:} \quad \psi^k_s, \varphi^k_s, \phi^k_s \in [-\infty, 1] \quad \forall s;\ k\in\{C, NC\}. \tag{3.2}$$

CES production functions are defined for non-negative inputs, that is for $(\theta^C_t, \theta^{NC}_t, I_t, \vartheta^C, \vartheta^{NC}) \in (\mathbb{R}^+)^5$. This is consistent with the theoretical domain of those inputs, however it has some practical consequences. The first of these is that the individual error terms $\eta_{i,t,k}$ used in the estimation of (3.1) and (3.2) are included multiplicatively and as exponentials. The second is that our auxiliary equations – for investment, for measurement, and for the evolution of income and investment – are written in terms of log inputs, so that their 'true' values will necessarily be $\in \mathbb{R}^+$.

The technologies (3.1) and (3.2) share several salient features. For example, they both assume unit total factor productivity and constant returns to scale; this is discussed in Section 3.4. In addition, they both normalise each latent input such that the post-normalisation input vector $(1, 1, 1, 1, 1)$ corresponds to a specified point $\Theta_t = (\Theta_t^C, \Theta_t^{NC}, \Theta_t^I, \Theta_t^{PC}, \Theta_t^{PNC})$ in the production space. The importance of CES normalisation is set out in Section 3.3, and discussed in detail by Embrey (2019c). In our estimation, we set $\mathbf{\Theta_t}$ at the sample mean level of each input at time t; this ensures that the $\gamma_{s,j}^k$ represent input shares along a policy-relevant ray through the production space.

The distinction between Models (3.1) and (3.2) is that, where the former assumes a single elasticity of substitution between all inputs, the latter allows three distinct elasticities of substitution between specific inputs. This flexibility is particularly useful in the present context, because it is of significant interest to identify whether parental investments are dynamic complements or dynamic substitutes. In order to do this, we require separate identification of the elasticity of substitution between investment and current skill levels (see the discussion that opens Section 2), which is given by φ in Model (3.2). Model (3.2) is the simplest CES production function which can identify that elasticity of substitution.

At this stage, it is valuable to make two observations concerning specifications (3.1) and (3.2). First, the technology described by (3.2) is a strict generalisation of (3.1), and second, the technology described by (3.1) is a strict generalisation of the functional form used in CHS. The first generalisation is given by relaxing the assumption that $\varphi_s^k = \phi_s^k = \psi_s^k$ for all s and $k \in \{C, NC\}$, and the second is given by relaxing the assumption that $\Theta_t = \mathbf{1}$, for all t.

Two main challenges arise in the estimation of (3.1) and (3.2). The first is that parental investment may be endogenous, either because parents systematically alter their investment $I_{i,t}$ in response to their child's current skill level $\theta_{i,t}$, or because parents systematically alter their investment in response to informative signals of their child's next-period skill level that they receive before our observation of $\theta_{i,t+1}$. We correct for the former type of endogeneity by simultaneously estimating a structural investment equation (3.3), and we correct for the latter by allowing residual investment $\pi_{i,t}$ – the error term from (3.3) – to be correlated with $\eta_{i,t,k}$. This approach was developed by CHS, and it is broadly equivalent to the control function method used by Attanasio et al. (2015), Attanasio, Meghir & Nix (2015) and Attanasio et al. (2017), except that their multi-stage estimation neglects the information regarding $\theta_{i,t}$ that measurements of $I_{i,t}$ implicitly provide through the investment equation.

$$\ln I_{i,t} = \delta_{t,1} \ln \theta^{C}_{i,t} + \delta_{t,2} \ln \theta^{NC}_{i,t} + \delta_{t,3} \ln \vartheta^{C}_{i} + \delta_{t,4} \ln \vartheta^{NC}_{i} + \delta_{t,5} \ln y_{i,t} + \delta_{t,6} + \pi_{i,t} \tag{3.3}$$

$$\eta_{i,t,k} = \gamma^{k}_{s,8} \pi_{i,t} + \nu_{i,t,k}, \quad \text{for } k \in \{C, NC\}, \tag{3.3.1}$$

where $y_{i,t}$ is the family income of individual i at time t, and where we assume that the residual error term $\nu_{i,t,k}$ is normally distributed with mean 0 for each skill k in each period t, and that it is independent across skills and periods.

The second main challenge is that child skills, parent skills, and parental investments are all inherently unobservable latent factors. Nevertheless, since we have two or more measurements $Z_{l,t,m}$ for each latent factor $l \in \{C, NC, I, PC, PNC\}$ at each time t, we can use a factor-analytic model to allow for error in each individual measurement. The basis of such a model is an assumption that the common variance between those measurements is precisely the latent factor of interest. CHS prove non-parametric identification of a factor-analytic measurement model (3.4) embedded within the present framework,[2] up to factor location and scale in each period, and Agostinelli & Wiswall (2016b) prove that, with all functional forms considered here, it is only necessary to fix the location and scale of the factors in one period. We therefore assume the following measurement error model:

$$\begin{aligned}
Z_{i,k,t,m} &= \mu_{k,t,m} + \alpha_{k,t,m} \ln \theta^{k}_{i,t} + \varepsilon_{i,k,t,m}, \quad k \in \{C, NC\},\\
Z_{i,I,t,m} &= \mu_{I,t,m} + \alpha_{I,t,m} \ln I_{i,t} + \varepsilon_{i,I,t,m},\\
Z_{i,Pk,m} &= \mu_{Pk,m} + \alpha_{Pk,m} \ln \vartheta^{k}_{i} + \varepsilon_{i,Pk,m}, \quad k \in \{C, NC\},
\end{aligned}$$

with $\alpha_{l,1,1} = 1 \;\; \forall\, l \in \{C, NC, I, PC, PNC\}$,
and with the initial mean of $\ln \theta^{k}_{i,1,}$ $\ln I_{i,1,}$ & $\ln \vartheta^{Pk}_{i}$ set to 0 for all i; $k \in \{C, NC\}$. (3.4)

Here we assume that the errors $\varepsilon_{i,l,t,m}$ are normally distributed with mean 0 for each measurement m of each latent factor l in each period t, and that they are uncorrelated across measurements for any given factor in any given time period. In principle a single measure could proxy multiple latent factors, provided that each factor has at least one unique measurement, however in practice that would reduce

[2] Given at least two measurements for each factor for each period, and, in the case of serially correlated measurement errors, one additional measurement. Serial correlation in measurement errors is likely in this context, since many of the measurement items recur over several periods, however the available data easily surpass both sets of requirements.

the clarity of interpretation, and so it is sensible to assign each measurement to at most one latent factor. In this chapter we follow the measurement assignments of CHS, because our objective is to identify the effect of our adaptations to their model.

To close the model, we must provide a structure for the evolution of parental skills, family income, and residual investments. We model parental cognitive and noncognitive skills as time-invariant, and we model income and residual investments according to an AR(1) process:

$$\begin{aligned}
\vartheta^k_{i,t+1} &= \vartheta^k_{i,t} = \vartheta^k_{i,}\ \forall k \in \{C, NC\}_{,} \\
\ln y_{i,t+1} &= \varrho_y \ln y_{i,t} + \rho_I + \nu_{i,t,y}, \\
\ln \pi_{i,t+1} &= \varrho_\pi \ln \pi_{i,t} + \rho_\pi + \nu_{i,t,\pi}.
\end{aligned} \tag{3.5}$$

where, as usual, the errors $\nu_{i,t,\pi}$ & $\nu_{i,t,y}$ are assumed to be normally distributed with mean 0. Note that the $\nu_{i,t,\pi}$ are uncorrelated with present and past inputs by construction, but that we also require both $\nu_{i,t,\pi}$ and $\nu_{i,t,y}$ to be uncorrelated with future error terms $\nu_{i,t,k}$ & $\nu_{i,t,y}$. This assumption is reasonable as one would expect that, if any shocks were to affect multiple residuals then they would do so contemporaneously. It is less certain that we may assume $\nu_{i,t,y}$ to be uncorrelated with all current skill levels $\theta^k_{i,t}$ & ϑ^k_i, since this would require: i. independence from current shocks $\nu_{i,t,k}$, ii. independence from all past skill levels, and iii. independence from past error terms $\nu_{i,t',y}$, $\nu_{i,t',\pi}$, and $\nu_{i,t',k}$ for all $t' < t$. Conditions ii. and iii. will be true if the effect of period 1 parent skills on family income, and the association between child skill formation and the complete history of family income over $t' \leq t$, are both captured by the first lag of family income. In practice this may be a simplifying assumption, however it appears to be valid as we find that specifying an AR(2) process for family income does not affect our results.[3] Condition i. may be restrictive, because certain events might influence both family income and a child's cognitive performance or noncognitive behaviours through their emotional connection with their parents; to the extent that any such association is positive our model will therefore over-estimate the importance of parental investment and parental income.

This chapter broadly follows the estimation procedure of CHS. The model parameters are jointly estimated by maximising a single likelihood function, which compares each observed measurement in each time period with its value as predicted by the technology of skill formation and the global factor loading for that

[3]If Model (3.2) is adapted to include an AR(2) term: the estimated factor shares all remain within 0.03 of their original values, the AR(2) coefficient is 29.7 times smaller than the AR(1) coefficient, and the likelihood ratio statistic is 0.96 against a critical value of 3.84 for one additional degree of freedom.

measurement. Numerical likelihood contributions are calculated by modelling each latent factor as normally distributed,[4] and by approximating their next-period distributions via the Unscented Kalman Filter. Five time-invariant individual covariates are included in each measurement equation: gender, age, and dummies to control for: teenage mothers, age at assessment, and non-linear age effects. The estimation procedure is described in detail by Cunha, Heckman & Schennach (2010) and by Cunha (2011), and so we do not reproduce those descriptions here. However, our supplementary materials do contribute a new guide to the code used to produce our results.

Our code is adapted from that developed by CHS. The following three subsections discuss the most substantial adaptations which we make to their code, whilst other, less complex, corrections and adaptations are described in Appendix A. To our knowledge, CHS represents the only existing paper to implement a single-step estimation procedure for a dynamic latent factor model of multi-dimensional skill formation. Although it is computationally expensive, this approach has a significant advantage in efficiency since each parameter is estimated taking account of its indirect implications across the complete system of equations and across all time periods. We therefore believe that our contributions build upon the best available estimation procedure for the technology of skill formation.

3 CES Normalisation

Consider a simple CES production function:

$$\theta_{t+1} = \left[\gamma_1 \theta_1^{\phi} + \gamma_2 \theta_2^{\phi}\right]^{\frac{1}{\phi}} \tag{3.6}$$

The share parameters γ_k are so called, because in the special (Cobb-Douglas) case where $\phi \to 0$ the share of total resource that should optimally be spent on θ_k is γ_k. This result is derived in Appendix D, by maximising equation (3.6) subject to the 'budget constraint' $p_1\theta_1 + p_2\theta_2 \leq R$, then solving for γ_1. In our context, resource R could include time, energy, and human capital, and p_i indicates the quantity of resource required to generate one unit of skill θ_i. At any optimal interior allocation we therefore have that:

$$\gamma_1 = \frac{p_1\theta_1(\theta_1)^{-\phi}}{p_1\theta_1(\theta_1)^{-\phi} + p_2\theta_2(\theta_2)^{-\phi}} \tag{3.7}$$

and so when $\phi \to 0$ the share interpretation of γ_1 is immediate. The derivation for γ_2 is symmetric.

[4] CHS do likewise, since they find no evidence to suggest that a mixture of normals would better fit empirical factor distributions.

In general, we can see from Equation (3.7) that the share of total resource that should optimally be spent on θ_k is a complex function that not only depends upon the γ_k, but also upon a chosen point (θ_1, θ_2) in the production space, and the complementarity parameter ϕ. Outside of the Cobb-Douglas case, the parameter γ_k cannot be interpreted as a general share parameter – it is correctly interpreted as share of θ_k only along the ray in the production space where $\theta_1 = \theta_2$. Thus in the present context where the θ_k are arbitrarily located latent factors, the parameters γ_k are particularly uninformative.

This problem can be overcome by normalizing the CES function. CES normalisation was first introduced by De La Grandville (1989), and a thorough discussion of its importance can be found in Embrey (2019c). In essence, CES normalisation rescales the units of the parameter space so that we may freely select the ray along which the γ_k represent share parameters. For the normalised CES function (3.8), the chosen ray is that which contains the specified point (Θ_1, Θ_2). For our main specification we set Θ_k as the sample mean value of $\{\theta_k\}_i$, so that our estimates of the γ_k represent input shares at a point relevant to the population in question.

The normalised CES function has the following form:

$$\theta_{t+1} = \left[\gamma_1 \left(\frac{\theta_1}{\Theta_1}\right)^{\phi} + \gamma_2 \left(\frac{\theta_2}{\Theta_2}\right)^{\phi}\right]^{\frac{1}{\phi}}, \tag{3.8}$$

from which we can determine that the share interpretation of the γ_k is valid whenever the inputs θ_1 and θ_2 are in the ratio of $\Theta_1 : \Theta_2$:

$$\gamma_1 = \frac{p_1 \theta_1 (\frac{\theta_1}{\Theta_1})^{-\phi}}{p_1 \theta_1 (\frac{\theta_1}{\Theta_1})^{-\phi} + p_2 \theta_2 (\frac{\theta_2}{\Theta_2})^{-\phi}} \tag{3.9}$$

It is important to note that the normalised CES form (3.8) requires no additional assumptions over the simplified form (3.6). Rather, it makes the implicit simplifying assumption that $\Theta_1 = 1$ and $\Theta_2 = 1$ both explicit and manipulable. This observation also makes plain the mechanism by which the procedure developed by CHS to anchor skills on adult outcomes affects parameter estimates. That procedure implicitly renormalises CES inputs so that parameters are estimated along the ray where input quantities are in inverse proportion to their approximate importance for adult outcomes. That normalisation is difficult to justify, and it produces estimates that are significantly different to those at the mean input levels for the population in question.

4 Factor Over-Identification

Agostinelli & Wiswall (2016b) observe that the practice of artificially fixing the location and scale of each latent factor in each period represents a restrictive assumption. Since their discussion of the matter is comprehensive, we provide only a brief summary here.

For a single-period model, the measurement system in (3.4) identifies latent factors only up to location and scale. That is, the model fits equally well for any linear transformation of each latent factor. It is therefore necessary to artificially fix the location and scale of each latent factor, which we (and CHS) achieve by fixing the initial mean value and one initial loading for each factor. However the situation is not the same for the dynamic model set out in Section 3.2. Once the locations and scales of the latent state variables are fixed for any one period, they will also be known in all other periods. This is because (all of) the CES functional forms used here satisfy the *known location and scale* property defined by Agostinelli & Wiswall (2016b): their output level at two or more points in the production space is independent of all estimated parameters.[5] The two unknown values of factor location and scale can therefore be identified using the two equations generated by those known points in the production space. In this chapter we therefore remove the over-identifying restrictions implied by fixing CES factor loadings in periods $t > 1$.

Analogously, it is important to allow factor loadings for residual investment to be freely estimated in periods $t > 1$. To do this we require the AR(1) process for residual investment to have known location and scale, and so we normalise ρ_π to zero and ϱ_π to unity in equation (3.5). CHS operationalise a different specification that fixes one factor loading for residual investment in each period, whilst allowing ϱ_π to be estimated freely. These two approaches are not equivalent. Because CHS assume that ϱ_π is constant for all periods t, their fixed factor loadings become arbitrary and restrictive assumptions from period $t = 3$ onwards. Appendix B shows that our specification substantially and significantly reduces the unexplained variance in estimated residual investment levels.

In contrast to the latent state variables, we maintain the assumption of CHS that family income is directly observable. Thus we normalise the factor loadings for measurements of family income to unity in all periods. We nevertheless allow the effect of the control variables to vary through time, which implies that the intercept terms in (3.4) must also be allowed to vary between measurements. In order to identify those intercept terms we therefore require the AR(1) process

[5] For example, in equations (3.1) and (3.2), the output level is known whenever input values are directly proportional to the normalisation point $(\Theta_t^C, \Theta_t^{NC}, \Theta_t^I, \Theta_t^{PC}, \Theta_t^{PNC})$.

for family income to have known location, but we also wish to allow income to evolve through time. We achieve this by normalising ρ_I to zero in equation (3.5) whilst allowing ϱ_y to take any positive value. This AR(1) evolution assumption is reasonable because real-terms income is measured on a defined scale that is stationary across periods, but we nevertheless test its validity by confirming that an AR(2) specification does not affect our results (see footnote 3).

Agostinelli & Wiswall (2016a) go beyond a discussion of factor over-identification, to propose a novel strategy that promises to identify more general production technologies with freely estimable locations and scales. One such technology would be a generalisation of the simple CES form (3.6) to include a total-factor-productivity term A_t, and a returns-to-scale term ρ_t:

$$\theta_{t+1} = A_t \left[\gamma_1 \theta_1^{\phi} + \gamma_2 \theta_2^{\phi} \right]^{\frac{\rho_t}{\phi}} \tag{3.10}$$

The identification of such technologies rests upon the existence of some *age invariant* measure for each dynamic input. As suggested by the definition reproduced below, that measure is then taken as the definitive scale of its associated latent factor – up to linear transformation through its measurement equation (3.4). In the case of latent skill variables this is problematic because their measurement scale is inherently arbitrary. For example, there is no reason to suppose that any individual's abstract 'cognitive ability level' should be exponentially related to their score on a Peabody Individual Achievement Test, as required by Agostinelli & Wiswall (2016a). That relationship could equally well be linear, logarithmic, S-shaped, or follow some highly non-linear monotonic functional form, because cognitive ability is an abstract construct which has no inherent cardinal quantification. It is therefore meaningless to estimate total-factor-productivity or returns-to-scale in the technology of latent skill formation.

> Agostinelli & Wiswall (2016a): "**Definition 2:** A pair of measures $Z_{t,m}$ and $Z_{t+1,m}$ is age-invariant if $E(Z_{t,m}|\theta_t = p) = E(Z_{t+1,m}|\theta_{t+1} = p)$ for all $p \in \mathbb{R}^{++}$."

5 The Data

Since we aim to build upon the methodological contribution of CHS, we base our empirical estimates upon the data that they use, and we also map the same set of measurements onto the same set of latent factors. As such, a full description of the data can be found in Cunha, Heckman & Schennach (2010) and their associated supplementary materials; we summarise only the most salient points here.

The data are taken from the U.S. National Longitudinal Study of Youth '79

Children and Young Adults cohort, and they are available from the National Longitudinal Surveys Program website at www.nlsinfo.org. The data are derived from biennial surveys of the firstborn children of the '79 cohort, from birth to age 14. We follow CHS in splitting these measures into two developmental stages: from 0-6 years, and from 5-14 years of age. Over that time, measures of cognitive ability transition from school readiness items to the Peabody Individual Achievement Test, whilst measures of noncognitive ability focus on prosocial and antisocial behaviours. Measures of parental investment include the provision of materials and experiences, as well as familial interactions. Measures of parental cognitive skills include the verbal, non-verbal, and mathematical abilities of the mother, whilst measures of parental noncognitive abilities focus on the self-worth and locus-of-control of the mother.

4 Results

Our estimates for the technologies of cognitive and noncognitive skill formation are presented in Tables 3.2 and 3.3 respectively. Both tables reproduce the results from Table V of CHS in their first column, and they identify those of our results that differ from CHS at the 95% significance level.[6] Tables 3.2 and 3.3 show substantial and statistically significant differences between our results and those of CHS. Where Model (3.1) differs from CHS this can be attributed to normalisation (together with the other minor coding corrections detailed in Appendix A), and where Model (3.2) differs from (3.1) this can be attributed to nesting CES production functions.

The bottom panel of Table 3.3 provides the log likelihood of each specification.[7] Although normalisation adds no parameters to the model, the full set of minor corrections listed in Appendix A adds 44 parameters to the model, of which

[6]Under the assumption that error terms are uncorrelated between the models, an appropriate test of statistical difference would be provided by Welch's unequal variances t-test with both tails. However, in our case the error terms will be highly correlated, because they are derived from the same sample using similar methods and identical measurements. Welch's test is therefore a strong criterion that provides a lower bound on the set of statistically different estimates. If the error terms were perfectly correlated and of equal variance, then any difference in estimates would be statistically significant. In comparison, student's t-test with two tails would provide accurate inference if the error variances were equal and if one quarter of that variance were common, and so it probably also represents a conservative test in this setting (note that Student's t statistic always exceeds Welch's). We indicate the outcomes of both tests in Tables 3.2 and 3.3.

[7] We report the log likelihood of the core model, which omits the additional idiosyncratic contributions of each specification: for example CHS have an additional likelihood contribution from their anchoring (CES normalisation) procedure, and we have an additional likelihood contribution from our investment equation shrinkage procedure. The 'core' log likelihood omits those additional idiosyncratic contributions so that it may be compared between the alternative specifications, which are then nested from left to right.

Table 3.2:
The Technology of Cognitive Skill Formation; Input Shares and Elasticities

	1st Stage (ages 0-6)		
Input	CHS Table V	Model (3.1)	Model (3.2)
Cognitive Skill $\gamma^C_{1,1} \cdot (\gamma^C_{1,6} \cdot \gamma^C_{1,7})$	.485	.342**	.352**
	(.031)	(.041)	(.042)
Noncognitive Skill $\gamma^C_{1,2} \cdot (\gamma^C_{1,6} \cdot \gamma^C_{1,7})$	.062	.001*	.001*
	(.026)	(.030)	(.030)
Parental Investments $\gamma^C_{1,3} \cdot (\gamma^C_{1,7})$	.261	.486**	.482
	(.026)	(.111)	(.113)
Parental Cognitive Skill $\gamma^C_{1,4}$	.035	.086	.082
	(.015)	(.027)	(.030)
Parental Noncognitive Skill $\gamma^C_{1,5}$	.157	.084	.083
	(.033)	(.058)	(.057)
e-sub: (Investment) $1/(1-\varphi^C_1)$	2.410	.893**	.983**
		(.136)	(.304)
e-sub: Child Skills $1/(1-\psi^C_1)$			—
			–
e-sub: Parent Skills $1/(1-\phi^C_1)$			.676**
			(.231)
	2nd Stage (ages 5-14)		
Cognitive Skill $\gamma^C_{2,1} \cdot (\gamma^C_{2,6} \cdot \gamma^C_{2,7})$	.884	.923**	.923**
	(.013)	(.015)	(.015)
Noncognitive Skill $\gamma^C_{2,2} \cdot (\gamma^C_{2,6} \cdot \gamma^C_{2,7})$	.011	.002	.002
	(.005)	(.002)	(.002)
Parental Investments $\gamma^C_{2,3} \cdot (\gamma^C_{2,7})$	.044	.000	.000
	(.011)	(.028)	(.028)
Parental Cognitive Skill $\gamma^C_{2,4}$	.051	.074	.073
	(.008)	(.019)	(.019)
Parental Noncognitive Skill $\gamma^C_{2,5}$	.011	.002*	.002*
	(.012)	(.026)	(.026)
e-sub: (Investment) $1/(1-\varphi^C_2)$	.450	.500	—
		(.043)	–
e-sub: Parent Skills $1/(1-\phi^C_2)$			.499
			(.044)

NOTE: Standard errors in parentheses; *Significantly different from CHS point estimate at the 95% confidence level (Student's t); **Significantly different from CHS at the 95% confidence level (Welch's unequal variances t).

Table 3.3:
The Technology of Noncognitive Skill Formation; Input Shares and Elasticities

Input	CHS Table V	Model (3.1)	Model (3.2)
	1st Stage (ages 0-6)		
Cognitive Skill $\gamma_{1,1}^{NC} \cdot (\gamma_{1,6}^{NC} \cdot \gamma_{1,7}^{NC})$	.000 (.028)	.001 (.039)	.001 (.039)
Noncognitive Skill $\gamma_{1,2}^{NC} \cdot (\gamma_{1,6}^{NC} \cdot \gamma_{1,7}^{NC})$	.602 (.034)	.693 (.056)	.691 (.056)
Parental Investments $\gamma_{1,3}^{NC} \cdot (\gamma_{1,7}^{NC})$	.209 (.031)	.000 (.129)	.000 (.129)
Parental Cognitive Skill $\gamma_{1,4}^{NC}$	.014 (.013)	.077** (.029)	.077** (.029)
Parental Noncognitive Skill $\gamma_{1,5}^{NC}$	.175 (.033)	.229 (.068)	.231 (.069)
e-sub: (Investment) $1/(1-\varphi_1^{NC})$	.683	.642 (.083)	— –
e-sub: Parent Skills $1/(1-\phi_1^{NC})$			.647 (.085)
	2nd Stage (ages 5-14)		
Cognitive Skill $\gamma_{2,1}^{NC} \cdot (\gamma_{2,6}^{NC} \cdot \gamma_{2,7}^{NC})$	.002 (.011)	.000 (.007)	.000 (.007)
Noncognitive Skill $\gamma_{2,2}^{NC} \cdot (\gamma_{2,6}^{NC} \cdot \gamma_{2,7}^{NC})$	.857 (.011)	.845 (.014)	.847 (.015)
Parental Investments $\gamma_{2,3}^{NC} \cdot (\gamma_{2,7}^{NC})$	.104 (.022)	.055 (.063)	.056 (.075)
Parental Cognitive Skill $\gamma_{2,4}^{NC}$	.000 (.008)	.000 (.017)	.000 (.017)
Parental Noncognitive Skill $\gamma_{2,5}^{NC}$	.037 (.021)	.100 (.056)	.097 (.071)
e-sub: (Investment) $1/(1-\varphi_2^{NC})$	.657	.664 (.054)	.703 (.355)
e-sub: Parent Skills $1/(1-\phi_2^{NC})$			.640 (.228)
	Model Fit		
Core Log-Likelihood†	205 421.73	205 108.87	205 108.27

NOTE: Standard errors in parentheses; **Significantly different from CHS at the 95% confidence level (Welch's unequal variances t);
† Omits anchoring contributions to maintain inter-model comparability (see footnote 7).

19 are freely estimable;[8] thus the χ^2 critical value of 60.5 for 44 degrees of freedom provides a conservative likelihood ratio test between CHS and Model 3.1. The test statistic between those specifications is 625.7, which therefore indicates unambiguously that CES normalisation and our minor corrections improve model fit. The two levels of CES nesting introduced by Model 3.2 require 8 additional parameters, although 6 elasticity parameters approach redundancy under this specification due to near-zero estimates on specific factor shares. The χ^2 critical value of 6.0 for 2 degrees of freedom therefore represents a generous likelihood ratio test of whether nesting improves model fit, and the test statistic of 1.2 does not approach that threshold. Thus we can conclude that our normalisation and minor corrections deliver a substantially and significantly better-fitting model than the benchmark of CHS Table V, but that nesting the CES production does not deliver a statistically significant improvement to model fit. We can also see from Tables 3.2 and 3.3 that nesting has little effect on our estimated input shares, although there is statistically insignificant evidence to suggest that elasticities of substitution may differ by input type.

We will now assess the implications of our results for each of the main conclusions that CHS derive from their estimates. CHS emphasise that their conclusions support five important stylised facts of skill formation: i) self-productivity increases with age for both skill types; ii) complementarity between investments and current cognitive skill levels increases with age; iii) the converse is true for noncognitive skills; iv) cognitive skills *stabilise*, that is later investment is ineffective; and v) noncognitive skills *flourish*, that is they respond particularly well to later investment. We discuss the implications of our results for each of these conclusions in turn.

The results of CHS demonstrate robustly that self-productivity increases with age for both cognitive and noncognitive skills. This can be seen by the substantial and statistically significant increases in the own-factor production shares for each skill type between developmental stages. Our results support the same high-level conclusion, although they suggest that the increase is larger than previously thought for cognitive skills, and smaller than previously thought for noncognitive skills. These findings are perhaps best considered as a consequence of the changes in the estimated factor shares of the other inputs, and so they are discussed when we consider stylised facts iv) and v) below.

CHS provide suggestive evidence that the complementarity between investments and current cognitive skill levels may increase with age. Whilst their pro-

[8] Five investment equation parameters $\times$ seven periods are shrunk to be approximately equal in 'real terms' across each of two periods; where 'real terms' compensate for non-uniform changes in the arbitrary variances of the latent state variables as they develop through time.

duction technology cannot separately identify this parameter, their estimates for a global complementarity parameter are sufficiently different and sufficiently precise to infer that this conclusion is likely. Model 1 replicates this finding with regard to the global elasticity of substitution in cognitive skill production, but it does so with two important caveats. First, we find no support for the surprising CHS result that early inputs to cognitive development are highly substitutable – instead we find them to be weakly complementary. Second, we note that the substantial and statistically significant increase in complementarity between current cognitive skills and other inputs as children age should not be interpreted as an increase in the dynamic complementarity of parental investment, because we find no evidence that later parental investment has any measurable impact on cognitive development.

CHS provide very little evidence that the complementarity between investments and current noncognitive skill levels decreases with age. We can see from Tables 3.2 and 3.3 that their main results suggest that, if anything, the opposite may be true. CHS therefore claim support for this stylised fact based upon the results of an alternative specification that does not correct for the endogeneity of investments: however, whilst those results have the opposite comparative sign, the difference between them is smaller and it remains far from statistical significance. Although we do find a small and statistically insignificant decrease in global complementarity in noncognitive skill production between stages, as above our estimates suggest that this should not be interpreted in terms of the dynamic complementarity of investments because earlier investments do not affect noncognitive skill production.

Arguably the most important stylised fact of human capital development is that early investment in cognitive abilities is vital, because those abilities are insensitive to later investments. This result is undoubtedly supported by CHS, not only because of the small input share of second stage investments, but also because of the low elasticity of substitution in the second stage of cognitive skill development. Our results support this stylised fact a fortiori. We find that the input share of early parental investment for cognitive development is substantially and significantly larger than indicated by CHS, whilst the input share of later parental investment is vanishingly small in our results.

The dual of the above stylised fact is that noncognitive abilities become more manipulable as children age. CHS declare support for this result on the basis of their ambiguous findings regarding the change across time of the elasticity of substitution in noncognitive skill formation that we discussed above. In reality, those ambiguous findings will be more than outweighed by the fact that they estimate a substantial reduction in the input share of investment between the two stages of noncognitive development. We find the opposite: when the model is normalised to an appropriate point within the production space, the share of parental investment in the technology of noncognitive skill production increases between developmental stages, albeit to a modest final level.

Outside of the main results discussed above, our findings are generally in line with those of CHS. For example, neither we nor CHS find much support for cross-productivity between skill types; indeed the weak positive effect of noncognitive skills on early cognitive development that was found by CHS vanishes in our results. However there are two further substantial differences between our estimates and those of CHS. It seems that the results from CHS Table V tend to misattribute the effect of parental skills on early noncognitive development to an effect of investments, and that they do the opposite for early cognitive skill development. The latter misattribution is particularly anomalous, since it leads to the counter-intuitive result that parental noncognitive skills are both substantially and significantly more important for their child's early cognitive development than are parental cognitive skills. These findings are not replicated in our results, and so we conclude that they be an artifact of the anchoring (CES normalisation) procedure in CHS, which estimates share parameters along a ray through the production space that is far from the sample mean input values. It is difficult to predict the effect of this inappropriate CES normalisation, because the econometric model fits a highly non-linear production function recursively to map out eight periods of skill development.

5 Discussion and Conclusion

In this chapter we build upon the contribution of Cunha, Heckman & Schennach (2010) in three important respects: we normalise its CES production function to estimate input shares at a policy-relevant point in the parameter space; we nest that CES production function to identify dynamic complementarity between investments; and we correct the over-identification within its factor-analytic measurement model. We also make several more minor improvements and corrections. We establish that these developments significantly improve the fit of the model, and we therefore also contribute a guide to the code in order to make the estimation procedure accessible for future research.

Our developments induce substantial and statistically significant changes in the estimated technologies of skill formation. For example, we establish that the share of parental investment in the production of noncognitive skills increases with the age of the child, which implies that parental investment can have a greater effect on the production of noncognitive skills as children age. CHS state that this is an important stylised fact of skill development, but their estimated elasticities are ambiguous with regard to this conclusion, and their estimated input shares are opposed to it. Our estimates for the technology of cognitive skill formation are broadly in-line with those of CHS, although we reverse their two most surprising findings. First, we find that early inputs to cognitive development are somewhat complementary, rather than highly substitutable, and second we find that parental cognitive skills are of comparable importance to parental noncognitive skills in fostering early childhood cognitive development, rather than being substantially less important. In addition to these results, we provide the first unbiased estimates of the degree of dynamic complementarity in parental investment, which suggest that it may be present to a moderate degree.

Our estimates for the technology of cognitive and noncognitive skill formation therefore enhance the empirical support for the main policy implications of Cunha, Heckman & Schennach (2010). We conclude that early intervention is vital for cognitive skill development, because later cognitive ability is almost entirely characterised by self-productivity, and because the elasticity of substitution between current skills and later parental inputs is very low. Conversely, noncognitive abilities remain malleable at much later ages. This presents an opportunity for successful later intervention, but it also presents the hazard that initially promising noncognitive skill development could be disrupted by negative experiences in later childhood.

A Minor Corrections and Adaptations to the Code of CHS

Each paragraph outlines a minor correction or adaptation to the code of CHS. None of these issues are discussed by CHS, either in their main paper or in their supplementary materials.

Non-zero coefficients in the parental investment equation (3.3) imply that our measures of investment also have non-zero factor loadings on the other state variables. For example if measurement $Z_{I,t,m}$ has loading $\alpha_{I,t,m}$ on $\ln I_t$, and $\ln \theta^C_{i,t}$ has coefficient $\delta_{t,1}$ in the investment equation, then measurement $Z_{I,t,m}$ should have factor loading $\alpha_{I,t,m} \times \delta_{t,1}$ on $\ln \theta^C_{i,t}$. A simultaneous estimation procedure such as that used here can exploit that implied relationship, and does so via the module `policy.f90` in the code of CHS. That module as written in CHS has no effect since the variable `t` is uninitialised; we initialise it to loop over periods.

CHS assume that the policy equation coefficients $\delta_{t,j}$ are invariant in absolute terms across time. However, since the relative locations and variances of the state variables vary substantially through time as a consequence of arbitrary factor-normalisation decisions, this assumption amounts to a strong and arbitrary restriction. We relax that restriction to imposing instead that the state-variable coefficients remain constant in real, rather than absolute, terms given the variances of the state variables – that is we impose: $\delta_{t,1} \times \mathrm{Var}(\theta_t^C) \times \mathrm{Var}^{-1}(I_t) = \delta_{t',j} \times \mathrm{Var}(\theta_{t'}^C) \times \mathrm{Var}^{-1}(I'_t)\ \forall\, t, t'$, and similarly for the other state variables. We then allow greater flexibility by imposing this restriction separately for each stage $s \in \{1, 2\}$. Since it is impossible to impose these restrictions contemporaneously within the estimation (the normalised values of each state variable would need to be defined before its nominal value could be known for any individual), the cross-equation restrictions are imposed by shrinkage (penalising the log likelihood for any inter-temporal deviations from the above equality). Appendix B shows that these adaptations substantially increase the significance of the investment equation coefficients.

CHS assume that the effect of residual investment $\pi_{i,t}$ on individual production shocks $\eta_{i,t,k}$ is fixed across stages $s \in \{0, 1\}$ and skill types $k \in \{C, NC\}$. We allow that effect ($\gamma_{s,8}^k$) to vary by stage and by skill type. Appendix B shows that there are statistically significant differences between these coefficients.

The code of CHS has a couple of additional typographical errors. First, the factor loading of $Z_{NC,t,19}$ was overlooked in the `mappings` module, and second, at one point in the anchoring section in `likeAUX.f90` the code referred to `KF` rather than `KFanch`. The latter would have significant repercussions, however since we now normalise our CES transitions (Section 3) that code section is no longer used. Similarly, the redundancy of CHS-style anchoring alleviates the problems that may be caused within the CES production function, where the anchoring process of CHS raised state variables were to a power rather scaling them linearly.

B Auxiliary Equation Estimates and Goodness-of-fit Measures

In this appendix we present the estimated coefficients of the auxiliary equations that sit behind our main results of tables 3.2 and 3.3, and we present our estimates for the auxiliary parameters that sit alongside those results. These comprise: the investment equation (3.3), the laws of motion for family income and residual investments (3.5), and the effects of heterogeneity within the technologies of skill formation $\nu_{i,t,k}$ and $\gamma_{s,8}^k$ for $k \in \{C, NC\}$. Estimates of the measurement equations are far too numerous to reproduce here, but are included within the supplementary material 'data and code'.

Table 3.B1 shows the estimated parameters of the investment equation for the

periods t =1-7. Period 8 parameters are uninformative, because there are no period 9 measurements with which to identify the left hand side of (3.3) through the technology function in period 8. The size of these parameters is difficult to interpret, because both the left- and right-hand sides of equation (3.3) are latent factors with arbitrary scales. However the sign and significance of estimates can be interpreted. We can see that, with the exception of family income, the investment function estimates of CHS are far from statistically significant. This suggests either that investment is not endogenous with respect to the included state variables, or that their investment function is poorly specified.

CHS assume that their investment equation coefficients are constant in absolute terms through all time periods. This assumption is inappropriate, because the variables on both sides of equation 3.3 are latent factors that evolve in a non-uniform fashion over time: for example parent skills are assumed to be constant, whereas child skills evolve according to a complex non-linear technology. Thus the relative scale of these factors will vary through time, which means that the absolute size of their coefficients should change inversely to that variation to map out a time-consistent investment function. One way to proceed would be to standardise variables before running the investment equation, but this is impossible because the investment equation is required in order to derive the state variables in each period. Accordingly, we proceed by scaling the investment coefficients post-estimation to cancel out factor variance, then by shrinking those scaled parameters towards consistent values for each stage of development by means of a likelihood penalty.

Table 3.B1: *Estimates of the Investment Equation Parameters*

		Model (3.1)						
	CHS Table V	Birth	Age 1-2	Age 3-4	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Cognitive Skill $\delta_{t,1}$	.003	.068	.219*	.193*	-.015	-.016	-.015	-.015
	(.154)	(.037)	(.030)	(.024)	(.021)	(.013)	(.013)	(.012)
Noncognitive Skill $\delta_{t,2}$	.047	.259*	.210*	.202*	.042*	.048*	.051*	.055*
	(.181)	(.049)	(.023)	(.020)	(.011)	(.009)	(.010)	(.010)
P. Cog. Skill $\delta_{t,3}$	.034	.052	.118*	.149*	.094*	.118*	.130*	.141*
	(.110)	(.041)	(.030)	(.035)	(.027)	(.023)	(.027)	(.029)
P. Noncog. Skill $\delta_{t,4}$	-.156	.002	.001	.003	.200*	.251*	.277*	.299*
	(.225)	(.091)	(.076)	(.090)	(.070)	(.063)	(.069)	(.069)
Family Income $\delta_{t,5}$	.142*	.077	.151*	.176*	.098*	.122*	.132*	.145*
	(.042)	(.056)	(.032)	(.035)	(.026)	(.021)	(.024)	(.024)
		Model (3.2)						
Cognitive Skill $\delta_{t,1}$		.068	.216*	.191*	-.015	-.016	-.014	-.014
		(.037)	(.030)	(.024)	(.021)	(.013)	(.013)	(.012)
Noncognitive Skill $\delta_{t,2}$		.259*	.210*	.202*	.042*	.048*	.051*	.055*
		(.049)	(.022)	(.020)	(.011)	(.009)	(.010)	(.010)
P. Cog. Skill $\delta_{t,3}$		.052	.118*	.149*	.095*	.119*	.131*	.141*
		(.041)	(.030)	(.035)	(.027)	(.023)	(.027)	(.029)
P. Noncog. Skill $\delta_{t,4}$		.000	-.003	-.002	.198*	.248*	.274*	.296*
		(.091)	(.075)	(.090)	(.070)	(.063)	(.070)	(.069)
Family Income $\delta_{t,5}$		.076	.149*	.175*	.098*	.122*	.132*	.144*
		(.056)	(.032)	(.035)	(.026)	(.021)	(.024)	(.024)

NOTE: Standard errors in parentheses; * Indicates significantly different from zero at the 95% confidence level.

Table 3.B1 shows highly significant estimated coefficients in the investment equation, with signs that match our expectations. We can also see some important differences between stages of development, with the effect of child cognitive skills on investment reducing markedly from age five, and the importance of parent noncognitive skills increasing markedly in that second stage of development. These results are consistent across our specifications.

Our specification for the laws of motion (3.5) of family income and of residual investment also varies from that of CHS. We allow the factor loadings for residual investment in periods $t>1$ to be freely estimated, and so we normalise the AR(1) coefficient on residual investment to unity. CHS estimate that coefficient, but fix one factor loading to unity in each period. These changes are not equivalent - by imposing the same AR(1) coefficient through for all periods t, the extra factor normalisations in CHS represent arbitrary and restrictive assumptions from period $t=3$ onwards. From Table 3.B2 we can see that our adaptation substantially and significantly reduces the unexplained variance in residual investment estimation.

Table 3.B2: *Estimates of Family Income and Residual Investment Equations*

	Log Family Income at time $t+1$		
	CHS Table V	Model (3.1)	Model (3.2)
Time t Log Income ϱ_y	.908	.901	.901
	(.005)	(.005)	(.005)
Variance of Shocks $\nu_{i,t,y}$	.115	.113	.113
	(.002)	(.002)	(.002)
	Log Residual Investment at time $t+1$		
Time t Log Residual Inv. ϱ_π	.831	—	—
	(.007)	–	–
Variance of Shocks $\nu_{i,t,\pi}$	.062	.019	.019
	(.003)	(.001)	(.001)

NOTE: Standard errors in parentheses.

Table 3.B3 shows the estimated degree of heterogeneity within the technology of skill formation under each specification. Individual heterogeneity is manifest through the error term of the technology of skill formation, which is further decomposed by Equation (3.3.1) into individual- and time-specific shocks for each skill type, and into the correlation between residual investment (that which cannot be predicted by the investment equation) and the error term in the production function. Again, the magnitude of these coefficients cannot be directly compared

due to the arbitrary scale of latent skill levels, but their sign and significance can be.

We can see that unexplained variation (shocks to production) is similarly significant across specifications, but that our estimates differ markedly from those of CHS regarding the effect of residual investment on skill production (they refer to this effect as that of unobserved heterogeneity). The results of CHS imply that unobserved heterogeneity which increases parental investment also systematically reduces both cognitive and noncognitive skill production, which is somewhat counter-intuitive. Our estimates suggest that our investment equation generally predicts almost the entire systematic effect of parental investment on skill production, except in the case of later cognitive skill production where unobserved heterogeneity that increases investment also increases cognitive skill production. These more intuitive findings are probably a consequence of the fact that our investment equation appears to be somewhat better specified than that of CHS (see Table 3.B1).

Table 3.B3: *Estimates of individual Heterogeneity*

	Technology of Cognitive Skill Formation		
	CHS Table V	Model (3.1)	Model (3.2)
Variance of Shocks			
Stage 1 $\mathrm{Var}(\nu_{i,t,C} : t \in \{1-3\})$	.165	.498	.502
	(.007)	(.016)	(.016)
Stage 2 $\mathrm{Var}(\nu_{i,t,C} : t \in \{4-7\})$	.098	.445	.449
	(.003)	(.013)	(.013)
Effect of Residual Investment (unobserved heterogeneity)			
Stage 1 $\gamma^{C}_{1,8}$	-.046	.000	.000
	(.017)	(.131)	(.130)
Stage 2 $\gamma^{C}_{2,8}$	-.046	.137	.138
	(.017)	(.044)	(.045)
	Technology of Noncognitive Skill Formation		
Variance of Shocks			
Stage 1 $\mathrm{Var}(\nu_{i,t,NC} : t \in \{1-3\})$	.203	.991	.996
	(.012)	(.057)	(.058)
Stage 2 $\mathrm{Var}(\nu_{i,t,NC} : t \in \{4-7\})$	.102	.488	.491
	(.003)	(.016)	(.017)
Effect of Residual Investment (unobserved heterogeneity)			
Stage 1 $\gamma^{NC}_{1,8}$	-.066	.000	.000
	(.029)	(.160)	(.161)
Stage 2 $\gamma^{NC}_{2,8}$	-.066	.036	.036
	(.029)	(.063)	(.075)

NOTE: Standard errors in parentheses.

C The Marginal Productivity of Input A (in percentage terms) is increasing in Input B if and only if their Elasticity of Substitution < 1

Proof of the above statement. Consider a simple production technology $Y = f(A, B)$, where all inputs and outputs are non-negative. Then the marginal productivity of A, in percentage terms at the point (A, B), is given by:

$$\frac{\partial f}{\partial A} \cdot \frac{A}{Y} =: f_A \cdot \frac{A}{Y}$$

This quantity is increasing in input B if and only if:

$$\begin{aligned}
\frac{\partial \left(f_A \cdot \frac{A}{Y} \right)}{\partial B} \cdot B &> 0 \\
\frac{AB}{Y} \cdot f_{AB} - \frac{AB \cdot f_A \cdot f_B}{Y^2} &> 0 \\
f_{AB} &> \frac{f_A \cdot f_B}{Y} \\
1 &> \frac{f_A \cdot f_B}{Y \cdot f_{AB}}
\end{aligned}$$

since inputs and outputs are non-negative. The right-hand side of the last inequality is precisely the original form of the elasticity of technical (or intertemporal) substitution between A and B, as derived by Hicks (1932).

D Derivation of the economic interpretation of the γ_i in a nested and normalised CES function

Consider a normalised n-factor CES production function

$$f = \left(\gamma_1 \left[\frac{x_1}{\Theta_1} \right]^\phi + \gamma_2 \left[\frac{x_2}{\Theta_2} \right]^\phi + ...\gamma_n \left[\frac{x_n}{\Theta_n} \right]^\phi \right)^{\frac{1}{\phi}}$$

$$\text{where: } \phi_s^k \in [-\infty, 1]\,; \quad \sum_{i=1}^{n} \gamma_i = 1; \quad \text{and: } \gamma_i \geq 0 \; \forall \, i. \tag{3.11}$$

We first prove Lemma 1:

Lemma 1 *Under monotonic preferences and with positive prices, the parameters γ_i in the normalised n-factor CES production function (3.11) represent the share of resources optimally dedicated to input x_i at any point along the ray in the production space that contains the normalisation point.*

Proof:

Maximise (3.11) subject to the resource constraint

$$\sum_{i=1}^{n} p_i x_i \leq R. \tag{3.12}$$

This holds with equality when preferences are monotonic and prices p_i are positive. The first order condition of this system under Lagrangian maximisation is therefore given by the set of n equations:

$$\frac{\gamma_i}{\phi\Theta_i}.\phi\left[\frac{x_i}{\Theta_i}\right]^{\phi-1}.f(\cdot)^{(1-\phi)} = \lambda p_i. \tag{3.13}$$

By forming the ratio of any two of these equations we obtain the well-known relationship between the optimal marginal rate of substitution between two goods and their price ratio, specifically:

$$\frac{\gamma_i\Theta_j.\left[x_i/\Theta_i\right]^{\phi-1}}{\gamma_j\Theta_i.\left[x_j/\Theta_j\right]^{\phi-1}} = \frac{p_i}{p_j}, \tag{3.14}$$

which implies that

$$\gamma_i.\frac{\left[x_i/\Theta_i\right]^{\phi}}{x_i p_i}.\frac{x_j p_j}{\left[x_j/\Theta_j\right]^{\phi}} = \gamma_j \;\forall\, j \not\equiv i.$$

We now sum these equations across all $j \not\equiv i$ to obtain

$$\gamma_i.\frac{\left[x_i/\Theta_i\right]^{\phi}}{x_i p_i}.\sum_{j\not\equiv i}\frac{x_j p_j}{\left[x_j/\Theta_j\right]^{\phi}} = \sum_{j\not\equiv i}\gamma_j = 1-\gamma_i$$

rearranging gives

$$\gamma_i.\frac{\left[x_i/\Theta_i\right]^{\phi}}{x_i p_i}.\sum_{j=1}^{n}\frac{x_j p_j}{\left[x_j/\Theta_j\right]^{\phi}} = 1$$

$$\gamma_i = \frac{x_i p_i\left[x_i/\Theta_i\right]^{-\phi}}{\sum x_j p_j\left[x_j/\Theta_j\right]^{-\phi}}. \tag{3.15}$$

Equation (3.15) is specialises to give (3.9) in the two-factor case, and specialises further to give (3.7) at the implicit normalisation of $\mathbf{\Theta} = 1$. Furthermore, note that, at any point $\boldsymbol{x}$ along the ray containing $\mathbf{\Theta}$ we will have that $\boldsymbol{x} = k.\mathbf{\Theta}$ for some scalar $k \in \mathbb{R}$. Hence, along that ray, equation (3.15) reduces to give

$$\gamma_i = \frac{x_i p_i}{\sum x_j p_j} = \frac{x_i p_i}{R}.$$

Thus we have demonstrated that γ_i can be interpreted along that ray as the share of total resources which should optimally be allocated to x_i. $\square$

For our main theorem we show that Lemma 1 generalises to an arbitrary nested CES production function.

Theorem 1 *Under monotonic preferences, with positive prices, and at any point along the ray in the production space that contains the normalisation point: the share of total resources optimally allocated to input x_i within an arbitrary nested and normalised CES production function is obtained by multiplying its coefficient γ_i with the γ coefficients on all nested levels that contain x_i.*

Proof:

We proceed by induction over the number of nesting levels in an arbitrary nested and normalised CES production function $g(\boldsymbol{y})$. Lemma 1 provides the result for nesting level 0, and so to prove 1 it suffices to show the induction step whereby an arbitrary normalised CES production function (3.11) is substituted for some (normalised) input y_i/Θ_i within an arbitrary nested and normalised CES production function $g(\boldsymbol{y})$.

We first fix some notation. Denote the composite, post-substitution, production function as $G := g \circ f$. Denote by m the number of factors in the initial production function g, and let $\boldsymbol{\Theta^g} \in \mathbb{R}^m$ be its normalisation point. Fix any point $\boldsymbol{y} = k.\boldsymbol{\Theta^g}$ along the ray in the production space containing $\boldsymbol{\Theta^g}$. Denote by $\boldsymbol{\Theta^f}$ the normalisation point of the nested CES function f that is to be substituted for y_i, and denote by $\boldsymbol{\Theta^G}$ the concatenation of the first $i-1$ elements of $\boldsymbol{\Theta^g}$, followed by the n elements of $\boldsymbol{\Theta^f}$, followed by the final $m-i$ elements of $\boldsymbol{\Theta^g}$. Similarly denote by $\boldsymbol{Y}$ the concatenation of the first $i-1$ elements of $\boldsymbol{y}$, followed by the n elements of $\boldsymbol{x}$, followed by the final $m-i$ elements of $\boldsymbol{y}$.

By the induction assumption, the share of total resource optimally dedicated to each input y_j within g is given by product of its coefficient γ_j with the γ coefficients for all nesting levels which contain y_j. Denote that product as $\prod_j$ for each input y_i. We must show that the share of resources optimally allocated to each input y_j, $j \not\equiv i$ is unchanged by the above substitution, and that the share of resources optimally allocated to each additional input x_k is given by $\gamma_k.\prod_i$.

For the former, the proof of Lemma 1 demonstrates that it would suffice to show that the marginal product $MP_j(y)$ of any input y_j, $j \not\equiv i$ at the point y of the production function g is equal to $MP_{\tilde{j}}(Y)$, the marginal product of the corresponding input $\tilde{j}$ at the point Y of the composite production function G, where $\tilde{j} := j$ if $j < i$, and $\tilde{j} := j + n$ otherwise. To see this, note that the term

$\gamma_i\left[y_i/\Theta_i\right]^{\varphi}$ in $MP_j(y)$ is equal to $\gamma_i\left[k\right]^{\varphi}$, and that the substituted term

$$\gamma_i\left(\gamma_{x_1}\left[\frac{x_1}{\Theta_1}\right]^{\phi}+\gamma_{x_2}\left[\frac{x_2}{\Theta_2}\right]^{\phi}+...\gamma_{x_n}\left[\frac{x_n}{\Theta_n}\right]^{\phi}\right)^{\frac{\varphi}{\phi}}=\gamma_i\left(\gamma_{x_1}\left[k\right]^{\phi}+\gamma_{x_2}\left[k\right]^{\phi}+...\gamma_{x_n}\left[k\right]^{\phi}\right)^{\frac{\varphi}{\phi}}$$
$$=\gamma_i\left[k\right]^{\varphi}$$

in $MP_{\tilde{j}}(Y)$ is equivalent because $\sum_{k=1}^{n}\gamma_{x_k}=1$.

Given that the share of resources optimally allocated to each input $y_j,\ j\not\equiv i$ is unchanged by the substitution of f for y_i/Θ_i, and since the γ_i sum to unity, we have that the share of the total resources R^G available for the composite function that is optimally allocated to the newly nested function f remains $\prod_i .R^G$. We may therefore apply Lemma 1 to f with the total resources $R := R^f = \prod_i .R^G$ to see immediately that

$$\gamma_{x_i}.\textstyle\prod_i = \frac{x_i p_i}{R^G},$$

namely that the product of the coefficient on each newly nested input x_i with all coefficients on its nesting levels $\prod_i$ gives the share of total resources that should optimally be allocated to that input at any point along the ray from the origin that contains Θ^G. □

Note that Theorem 1 could equivalently be rewritten as Corollary 1:

Corollary 1 *If the normalised CES function f represents the production of an intermediate good y_i/Θ_i which is an input to normalised CES production function g, and each function is normalised so that its share parameters are interpretable along a ray-of-interest within its respective production space, then the compound production function $G := g \circ f$, is normalised so that its share parameters are interpretable along the corresponding ray within its production space.*

Further analysis of arbitrarily normalised and nested production functions are provided in Embrey (2019c). In particular, I show there that standard results regarding the derivation of Linear, Leontief, and Cobb-Douglas production functions as limiting cases of the CES family extend to the arbitrarily normalised and nested case, and that the standard derivation of the Translog production function as a second order approximation around the Cobb-Douglas special case is extended to a new derivation of the Translog function as a second order approximation to a normalised CES production function around its normalisation point.

Chapter References

Agostinelli, Francesco & Matthew Wiswall (2016a). “Estimating the Technology of Children’s Skill Formation”. In: NBER Working Paper, No. 22442. Cambridge.

— (2016b). “Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development”. In: NBER Working paper, No. 22441. Cambridge.

Attanasio, Orazio, Costas Meghir & Emily Nix (2015). “Human Capital Development and Parental Investment in India”. In: Economic Growth Center Discussion Paper, No. 1052. New Haven.

Attanasio, Orazio et al. (2015). “Estimating the production function for human capital: Results from a randomized control trial in Colombia”. In: National Bureau of Economic Research working papers, no. 20965.

Attanasio, Orazio et al. (2017). “Human capital growth and poverty: Evidence from Ethiopia and Peru”. In: *Review of Economic Dynamics* 25, pp. 234–259.

Coneus, Katja, Manfred Laucht & Karsten Reuß (2012). “The role of parental investments for cognitive and noncognitive skill formation - Evidence for the first 11 years of life”. In: *Economics and Human Biology* 10.2, pp. 189–209.

Cunha, Flávio (2011). “Recent developments in the identification and estimation of production functions of skills”. In: *Fiscal Studies* 32.2, pp. 297–316.

Cunha, Flávio & James J Heckman (2007). “The Technology of Skill Formation”. In: *The American Economic Review* 97.2, pp. 31–47.

— (2008). “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Journal of Human Resources* 43.4, pp. 738–782.

Cunha, Flávio, James J Heckman & Susanne M Schennach (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Econometrica* 78.3, pp. 883–931.

De La Grandville, Olivier (1989). “In Quest of the Slutsky Diamond”. In: *The American Economic Review* 79.3, pp. 468–481.

Embrey, Iain Peter (2019c). “On the Benefits of Normalization in Production Functions”. Lancaster.

Fiorini, Mario & Michael P Keane (2014). “How the Allocation of Children’s Time Affects Cognitive and Noncognitive Development”. In: *Journal of Labor Economics* 32.4, pp. 787–836.

Helmers, Christian & Manasa Patnam (2011). “The formation and evolution of childhood skill acquisition: Evidence from India”. In: *Journal of Development Economics* 95.2, pp. 252–266.

Heredia, Diana Jimena Martinez (2017). “El rol de la inversión en la función de producción de habilidades cognitivas: evidencia de Colombia, Chile y Perú”. In: Serie Documentos Cede, No. 2017-12. Bogotá.

Hernández-Alava, Mónica & Gurleen Popli (2017). “Children’s Development and Parental Input: Evidence From the UK Millennium Cohort Study”. In: *Demography* 54.2, pp. 485–511.

Hicks, John R (1932). *The Theory of Wages*. London: MacMillan and Co. Ltd.

Nicoletti, Cheti & Birgitta Rabe (2014). “School inputs and skills: Complementarity and self-productivity”. In: IZA Discussion Paper Series, No. 8693. Bonn.

Pavan, Ronni (2016). “On the Production of Skills and the Birth-Order Effect”. In: *Journal of Human Resources* 51.3, pp. 699–726.

Reuß, Karsten (2011). “Determinants of personality and skill development in the Socio-emotional environment during childhood”. In: MPRA Paper No. 82818. Munich.

Chapter 4

Noncognitive Skills: Theory and Empirics

Iain P. Embrey,

*Lancaster University**

Abstract:

Noncognitive skills have provoked substantial research interest in recent years. However, existing evidence for their importance relies heavily upon reduced-form regressions that do not account for the simultaneous production of cognitive and noncognitive skills. Moreover, existing constructs of noncognitive skill are typically chosen for their convenience, or else for their explanatory power over psychological survey items. In response, we analyse a decision-theoretic model of educational development to derive five candidate noncognitive skills, and we test their importance by adapting an established longitudinal and structural econometric model of multidimensional skill formation. We find: first, that noncognitive skills matter for cognitive development; second, that different aspects of noncognitive skills matter to different degrees; and third, that once a child begins to take her own decisions, her propensity to think analytically becomes the most important noncognitive determinant of her ongoing cognitive development.

JEL Codes: I24; D91; C73; C61; J24; I21; C38

Keywords: Noncognitive Skills; Education; Human Capital; Cognitive Development; Dynamic Factor Analysis.

*i.embrey@lancaster.ac.uk; Lancaster University Management School, Lancaster, LA1 4YX.
I would like to thank Kim Kaivanto, Steve Bradley, and Byunghoon Kang for their comments. Any remaining errors are entirely my own.

1 Introduction

Noncognitive skills are increasingly acknowledged as important determinants of individual outcomes. It is now well established that childhood noncognitive skill levels predict adult outcomes in employment, wages, health, and criminality (Moffitt et al. 2011; Segal 2013; Heckman, Pinto & Savelyev 2013). However, research into noncognitive skills rarely goes beyond reduced-form empirics. Furthermore, there is little consensus as to what noncognitive skills are, and there is almost no theoretical basis for the mechanism(s) by which they might act.

Our first contribution is to derive a theoretical foundation for five candidate noncognitive skills. We build upon the model of childhood skill formation that was developed in Chapter 2 (Embrey 2019f), to identify five specific and tangible mechanisms through which an aspect of noncognitive skill could affect a child's cognitive development. We show precisely how and why a child's day-to-day educational investment decisions will be affected by: her academic confidence, her resilience to failure, her drive to succeed, her cost of effortful task participation, and her propensity to think analytically. Each of these potential mechanisms generates a testable hypothesis that its associated noncognitive skill should be an important determinant of childhood cognitive development.

Our second contribution is to test each of these hypotheses. Ordinary regressions cannot do this reliably, because cognitive and noncognitive skills develop simultaneously from the same set of inputs. To account for this simultaneity, we adapt the econometric model of cognitive and noncognitive skill formation that was pioneered by Cunha, Heckman & Schennach (2010) and developed further in Chapter 4 (Embrey 2019d). This recursive model of multidimensional skill formation utilises longitudinal data to disentangle the effects of self-productivity and cross-productivity in the technologies of cognitive and noncognitive skill production. Thus our hypotheses may be tested by the statistical significance of the cross-productivity parameter for the input share of present-period noncognitive skills in the production of next-period cognitive skills. Additional benefits of this estimation strategy are that it implements a factor-analytic measurement model to correct for measurement error, and that it incorporates a structural investment equation to correct for the potential endogeneity of parental investments.

We find that all five candidate noncognitive abilities are statistically significant determinants of cognitive development. Moreover, the estimated effect sizes also have economic significance. For the average child, we estimate that the input shares of her drive to succeed and her cost of effortful participation separately account for around 9% of her period-on-period cognitive development during her early years, whilst her propensity to think analytically accounts for 14% of her period-on-period cognitive development during later childhood. Our point estimates of the input share for each of these specific noncognitive skills are substantially and significantly larger than those of our benchmark model, which follows Cunha, Heckman & Schennach (2010) in assuming that prosocial and antisocial behaviours adequately proxy for generic noncognitive skill levels.

Our results complement the seminal findings of Cunha, Heckman & Schennach (2010) by providing micro-foundations for specific aspects of noncognitive skill, and by uncovering evidence that several of these explain a larger part of childhood cognitive development than generic noncognitive skill levels. Our results contrast with those of Cunha et al. in that we find some specific aspects of noncognitive skill to be increasingly important in later childhood, whereas they find noncognitive skill to be almost irrelevant after the age of five. Cunha et al. take this result to be a manifestation of the well-established fact that cognitive skills stabilise from early childhood, and this conclusion is supported by their estimates which show that cognitive skills are increasingly self-productive in later childhood. We also find that cognitive skills are increasingly self-productive in later childhood – except when the noncognitive skill in question is the child's propensity to think analytically. This countervailing result is commensurate with the observation that a child's own decision-making will become an increasingly important determinant of her educational investment levels as she matures.

The increasing importance of a child's propensity to think analytically has important implications, both for public policy and for economic theory. For economic theory, it implies that an individual's propensity to think analytically is an economically relevant determinant of her decision-making, at least within the domain of educational investment. Thus economic theory should consider whether there might be other decision domains in which the normative decision theory represented by Figure 4.1 might not provide as much insight as its generalisation in Figure 4.2. For public policy, it implies that the effect of a child's increasing self-determinacy in later childhood can outweigh the effect of early cognitive stabilisation. Thus, contrary to most of the literature, we conclude that later interventions could potentially be more effective than early interventions, provided they successfully develop the noncognitive skill of an individual's propensity to think analytically.

2 Existing Knowledge

1 The Empirical Literature

Although the empirical importance of noncognitive skills is well established, the appropriate interpretation of the term 'noncognitive skills' is not (Thiel & Thomsen 2013; Rustichini et al. 2016). The term was coined by Heckman & Rubinstein (2001) as a catch-all for the set of all individual traits and abilities that one would not generally describe as 'cognitive', and so existing operationalisations of noncognitive skill range from the big-five traits of personality psychology, to proxies such as antisocial behaviors and participation in sporting activities (for a selective survey see Humphries & Kosse 2017).[1] Such measures are atheoretical, but it is nevertheless pragmatic to make use of them given their predictive validity and intuitive relevance.

More recently, behavioural economic theory has added to the set of candidate noncognitive skills by identifying preference parameters such as discount factors and risk premia that ought to affect choice behavior. Such preference parameters are typically elicited experimentally, although there is evidence that simple Likert-style survey items may provide equally useful measurements of economic preferences (Dohmen et al. 2011). Recent research has found that these micro-founded noncognitive skills may be comparable to personality traits in terms of their predictive power for life outcomes, and also that these two subsets of noncognitive skills may be somewhat complementary as determinants of those outcomes (Becker et al. 2012; Burks et al. 2015; Rustichini et al. 2016; Almås et al. 2016; Humphries & Kosse 2017). Sadly it is rare for studies that regress life outcomes on preference parameters to explicitly present a model that relates those preference parameters to the outcome in question; indeed the present study is the first that I know of which deliberately analyzes a theoretical model to generate and test hypotheses as to which specific noncognitive skills should affect a given outcome of interest.

Most existing evidence on the importance of noncognitive skills is derived from reduced-form estimations that treat skills as static across time. Of the 27 recent studies into noncognitive skills that were identified by Humphries & Kosse (2017), only one went beyond reduced-form estimations, and none derived micro-foundations for their noncognitive measures. Reduced-form studies provide useful indicative results of associations between observables, but ultimately the causal interpretation of such studies is jeopardised by the potential for confounding vari-

[1] The big-five traits are typically described as: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience, following the NEO-PI taxonomy of McCrae & Costa (1985).

ables. This is particularly true in the context of individual differences in life outcomes, and it can only be partially mitigated by controlling for observables. Of the many studies listed by Humphries & Kosse (2017), two utilise truly longitudinal approaches to tackle this problem. Heineck & Anger (2010) estimate individual random effects to account for unobserved heterogeneity, whilst Heckman, Humphries & Veramendi (2018) estimate an instrumented dynamic choice model to establish the causal effects of education, but not necessarily the causal effects of noncognitive skill levels.

The econometric challenge presented by the simultaneous and dynamic production of both cognitive and noncognitive skills has inspired a small number of researchers to develop explicit econometric models of that production process. The seminal contributions in this field are those of Cunha & Heckman (2008) and Cunha, Heckman & Schennach (2010). These studies estimate the production share of noncognitive skills within the cognitive development process to be small, at 2% or 6% in a child's early years and vanishing thereafter. Comparable results are found by Attanasio et al. (2015) and Hernández-Alava & Popli (2017), whilst Helmers & Patnam (2011) find that noncognitive ability negatively affects cognitive development – although this finding may be less reliable because it is based upon a single period of observation. The results of Coneus, Laucht & Reuß (2012) suggest that these inconsistencies may stem from the inconsistent operationalisation of noncognitive skills between these studies: they decompose noncognitive skills into emotional and mental skills to find that only the latter positively affect cognitive development. Although this conclusion is intuitively plausible, there is also conflicting evidence from Reuß (2011), who tends to draw the opposite conclusion from the same operationalization within the same dataset. The mixed findings of this specialist literature contrast with those of the broader field of the behavioral economics of education, wherein the importance of noncognitive skills for educational production is no-longer in question (Koch, Nafziger & Nielsen 2015; Lavecchia, Liu & Oreopoulos 2016).

The specialist educational production literature is reviewed in Chapter 3 (Embrey 2019d), where I conclude that the original work by Cunha, Heckman & Schennach (2010) still represents the state of the art in that field. Cunha et al. use measurements of antisocial behaviours as a proxy for noncognitive skills, and so in this study we will use a similar set of measures as a benchmark specification. The results from this benchmark can be compared meaningfully to those of Model 1 in Chapter 3, which develops the contribution of Cunha et al. by normalizing the production function, by relaxing over-identifying assumptions within the factor model, and by amending some other minor discrepancies within the code. The difference between the results in Chapter 3 and those of the present benchmark specification is that the former are derived from the US Children of the National Longitudinal Study of Youth, whilst the latter are derived from the UK Millenium Cohort Study.

2 The Theoretical Literature

The objective of our theoretical analysis is to establish a mechanism by which noncognitive skills could affect cognitive development through childhood. Most existing models of educational investment treat cognitive development as a black box, and so provide insufficient insight for our present purpose.[2] However, the canonical model due to Becker (1964) does provide a useful intuition that human capital may be generated through the input of costly effort, and we can further develop that intuition in light of the empirical results of Fiorini & Keane (2014) and Del Boca, Monfardini & Nicoletti (2017), who find that children's time investments are a substantial determinant of their cognitive development. These insights suggest that cognitive production might usefully be modelled as the cumulative outcome of participation in a series of minor educational investment opportunities, an approach which has been formalised by Sjögren & Sällström (2004), Filippin & Paccagnella (2012), MacLeod (2016), and in Chapter 2 (Embrey 2019f). This approach is intuitively appealing and, to my knowledge, it is the only candidate mechanism for cognitive production that has been formalised within the theoretical literature on educational development.

There nevertheless remain considerable variations in the detailed assumptions that underpin economic theories of educational development. For example, the prevailing assumption in that literature is that ability is a fixed endowment (see, for example Bénabou & Tirole 2002; Filippin & Paccagnella 2012; Belzil 2007; Köszegi 2006; Akerlof & Kranton 2002; Wang & Yang 2003; Sjögren & Sällström 2004), whilst only a few authors allow ability to be developed through educational participation (see, for example MacLeod 2016; Embrey 2019f). The empirical reality is that the nascent abilities endowed upon us by nature develop throughout childhood and beyond, and they do so as a function of our developmental environment and experiences (Cunha & Heckman 2010; Heckman & Mosso 2014). That being so, it is possible that some models with static ability may use the term 'ability' as short-hand for a concept of an individual's idiosyncratic capacity to develop ability, in which case this modelling discrepancy would be semantic, rather than fundamental in nature. Presumably this generally is the case, because it is widely accepted throughout the literature that human capital can be developed through the exertion of effort.

[2] Notable 'black box' models include those of Becker (1962, 1964) and Cunha & Heckman (2007).

A more fundamental discrepancy within the literature is therefore the question of whether educational participation produces cognitive development ipso facto, or whether it is necessary to achieve success in order to develop human capital. Here again MacLeod (2016) and Embrey (2019f) diverge from the main body of literature by assuming that, in the words of a somewhat sceptical anonymous reviewer: “in the attempt to understand algebra or to acquire familiarity with a computer, what matters is not to actually achieve this skill, but to have attempted to do so.” Again the discrepancy here may be largely semantic. The fact is that in a modern educational setting, the child who genuinely attempts to master a current learning objective is highly likely to do so, whether independently or through seeking support from peers and from the teacher. It is therefore the subjective perception of success that is in question. Perceived success may be determined by formal grading or by informal comparison with one’s peers, and it is this perception that gives rise to either positive or negative psychic payoffs under the assumptions of Embrey (2019f).[3]

In the following section we maintain the assumptions of Embrey (2019f), and we further analyse that model to derive five candidate noncognitive abilities. We adopt this model: i) because it is the approach that most closely formalises the accepted classical view that cognitive development is produced through the exertion of effort; ii) because the results of Propositions 1-3 below would be unaffected under the alternative interpretation of success (as is also the case for the main results of MacLeod (2016) and Embrey (2019f)); iii) because the conclusions of most alternative models collapse under our preferred interpretation of success, since they frequently depend upon a trade-off between the benefit of learning and the probability of achieving that benefit; and iv) because Observation 1 critically depends upon our preferred interpretation of success, and that observation is strongly supported by our empirical results.

3 The Theoretical Model

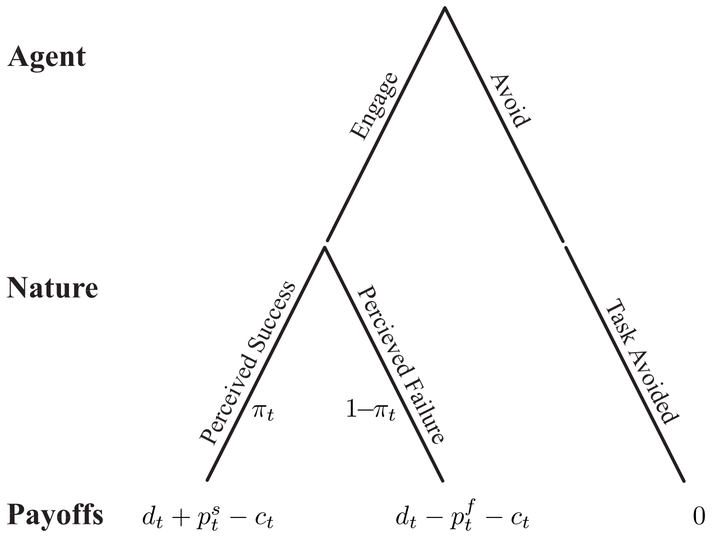
We model the mechanism of educational investment as a series of elemental participation decisions. Thus, cognitive development during the early years of childhood will be produced by participation in games, challenges, and other informal educational opportunities that may be provided by a child’s parents. As the child matures, incremental educational investment opportunities will increasingly be comprised of classwork, homework, and extra-curricular activities within formal

[3] Embrey (2019f) presents a formal theoretical exposition of the modern educational system: in principle opportunity is available to all; in practice the social context and early-years development of disadvantaged children may lead to their not taking up that opportunity until it is too late.

education. In each period t, the agent must therefore decide whether or not to engage with the present educational opportunity, and she will do so according to the extensive form decision problem presented in Figure 4.1. The payoffs relevant to that decision are:

$\boldsymbol{d_t}$ the present value of the human capital developed through task participation,
$\boldsymbol{c_t}$ the direct and opportunity cost of effortful task participation,
$\boldsymbol{p_t^s}$ the psychic payoff to perceived success, and
$\boldsymbol{p_t^f}$ the psychic cost of perceived failure.

Figure 4.1: *A Representative Agent's Participation Decision*



Note: Adapted from Figure 2.1 in Chapter 2 (Embrey 2019f).

Cognitive skill development is therefore modelled as a cumulative process, wherein the stage game presented in Figure 4.1 is played repeatedly. Accordingly, the developmental payoff d_t must take into account the full stream of expected future-period consequences from present-period participation, and so it may be defined recursively by

$$d_t = V_t + \delta \mathbf{E}\Big[\,(U_{t+1}\,|\,engagement_t) - (U_{t+1}\,|\,avoidance_t)\,\Big], \quad \text{for } t < T,$$
$$\text{and} \qquad d_T := V_T; \tag{4.1}$$

where δ is the agent's discount factor, where U_{t+1} is her next-period utility payoff, and where V_t is the present value of the incremental life-cycle benefit of the educational opportunity in period t. Intuitively, V_t enumerates the direct value of the skills developed by the present task, whilst the expectation term enumerates the indirect value of the future developmental opportunities opened up by the present task. This temporal interdependence greatly complicates any analysis of the model, and so we develop here only such analytic machinery as is necessary to derive our candidate noncognitive skills. A full analysis of the equilibrium pathways of the repeated game is presented in Chapter 2 (Embrey 2019f).[4]

We begin with some basic assumptions regarding the agent's payoffs for all periods t. First, all payoffs are well defined maps from the set of all possible events into the set of real numbers. Second, the payoff due to inaction is normalised to zero – this is without loss of generality since utility is defined up to affine transformation. Third, $p_t^s \geq 0$ and $-p_t^f \leq 0$: we assume that success is pleasant and that failure is unpleasant. Fourth, $c_t \geq 0$: we assume the agent has some outside option which is at least as attractive as doing nothing. Fifth, $V_t \geq 0$: we assume that the life-cycle benefit of each educational task is non-negative. Note that this does not necessarily imply that $d_t \geq 0$; counter-examples could occur if present participation were expected to have negative consequences on future parameter values – for example it may be that participation is likely to lead to perceived failure, and that this is expected to affect future decisions by increasing p^f, the agent's aversion to failure, thereafter.

We continue by establishing the stage-game participation condition as implied by these assumptions. In each period t, the agent will engage with the present educational opportunity if and only if her expected utility from doing so exceeds her expected utility from avoiding that opportunity. That is, if and only if:

$$d_t + \pi_t.(p_t^s + p_t^f) \geq c_t + p_t^f, \tag{4.2}$$

where π_t is the agent's anticipated probability of feeling successful conditional upon engagement with the present educational opportunity.[5]

Given specific values for each parameter, Equation (4.2) determines whether the agent will engage with the present educational opportunity. However, ex ante there are multiple potential sources of variation around this clear prediction. First, each of the parameters in Equation (4.2) is likely to depend upon the specific

[4]This formulation is subtly distinct from that developed in Chapter 2, in that here V_t is built into the utility payoffs in each period as it is accumulated, rather than carried back recursively from a realisation in period $T+1$. These formulations are equivalent if educational opportunities are perfect substitutes (as was maintained in Chapter 2), but for the present analysis we do not require that assumption, and so we take this more general approach.

[5]In this formulation, the realisation of perceived success determines which of the psychic payoffs p^s and p^f will become manifest, but it does not affect d_t. Thus we assume that human capital is developed by engagement with each educational opportunity, whilst the perception of success is determined independently, for example by comparisons with one's peers or by the allocation of grades. This assumption was discussed in Section 2, where it was noted that it is not critical for Propositions 1-3. Indeed it is straightforward to adapt the model such that the agent only learns from her perceived successes, and that adaptation would be orthogonal to those propositions. On the other hand, Observation 1 would be weakened or reversed under the adapted model.

nature of the period t task. Thus the realisation c_t, for example, represents a draw from the distribution of the agent's potential participation costs across all possible educational opportunities in period t. Second, it is unlikely that the agent will know her true parameter values for any given task a priori: instead she will estimate these from a signal that is likely to contain an element of stochastic noise. For example the apparent effort requirement of an educational task is likely to depend upon a teacher's presentation of that task. Thus the agent's anticipated participation cost c_t should more properly be considered as $c_t(x, y)$, a draw from the distribution of posterior c_t across all possible task $\times$ signal pairs in period t.

Since only the distributions of each parameter in Equation (4.2) are fixed ex ante, it is sensible to define λ_t as the ex ante probability of participation in period t:

$$\lambda_t := \int_{X_t} \int_{Y_t \,|\, X_t = x} 1\Big[d_t + \pi_t.(p_t^s + p_t^f) \geq c_t + p_t^f\Big] \, df(y \,|\, x) \, df(x), \tag{4.3}$$

where X_t is the set of possible tasks in period t, where Y_t is the set of possible signals in that period, and where $f(y \,|\, x) = f_{Y_t}(y \,|\, X_t = x)$ denotes the conditional probability distribution function of possible signals y for each task x. Here we suppress the function notation $\pi_t(x, y)$, $p_t^s(x, y)$, $p_t^f(x, y)$, $c_t(x, y)$, and $\lambda_t(d_t, \pi_t, p_t^s, p_t^f, c_t)$ for clarity. The integral λ_t exists whenever the set $X_t \times Y_t$ of task $\times$ signal pairs is probability measurable – indeed its result is precisely the probability measure of the subset of pairs for which the participation condition (4.2) is satisfied.

Equation (4.3) provides a clear notion of how each parameter could affect an agent's ex ante probability of participating in the period t educational opportunity. Nevertheless, not all parameters will be effectual in all circumstances. For example, there may be some tasks x for which

$$[1 - \pi_t(x, y)] \, .p_t^f(x, y) > d_t(x, y) + \pi_t(x, y).p_t^s(x, y), \quad \text{for all signals } y : f(x, y) > 0. \tag{4.4}$$

In this case the participation condition (4.2) cannot be satisfied for any positive value of c_t. Intuitively, this is because the probability of success is sufficiently small that the fear of failure exceeds both the developmental payoff and the potential satisfaction from succeeding in the present task, and so participation would not be optimal even if its cost were zero. In such cases we will say that the participation condition (4.2) is not binding at x:

Definition 1 *The participation condition (4.2) is binding for task(s) $\chi \subseteq X_t$ if:*

$$\int_{Y_t \,|\, X_t = x} 1\Big[d_t + \pi_t.(p_t^s + p_t^f) \geq c_t + p_t^f\Big] \, df(y \,|\, x) \ \in (0, 1) \ \text{ for all } x \in \chi.$$

Thus we say that the participation condition is binding for tasks χ if it is possible, but not certain, that the signal y will induce participation.

Where the participation condition is binding, the participation probability λ_t can be affected by the distributions of c_t, p_t^s, p_t^f, and π_t over the set of possible task $\times$ signal pairs. Let us refer to these distributions as C_t, P_t^s, P_t^f, and Π_t respectively. In order to assess the effects of heterogeneity in parameter distributions, we will use the following partial ordering:

Definition 2 *A family of partial orders on parameter distributions:*
For any $\epsilon \in \mathbb{R}$*, write* $\tilde{C}_t >_\epsilon C_t$ *on* $\chi \subseteq X_t :$ *if* χ *occurs with positive probability,*
if $\tilde{c}_t(x,y) > c_t(x,y) + \epsilon$ *for all* $(x,y) \in \chi \times Y_t$ *that satisfy* $f(x,y) > 0$*,*
and if $\tilde{c}_t(x,y) \geq c_t(x,y)$ *for all* $(x,y) \in X_t \times Y_t$ *that satisfy* $f(x,y) > 0$*.*
We define partial orders over P^s, P^f*, and* Π *analogously.*

This partial ordering is intuitive: $\tilde{C}_t >_\epsilon C_t$ requires that all possible realizations of $\tilde{C}_t$ exceed the corresponding realizations of C_t by at least ϵ, and we will say that $\tilde{C}_t >_\epsilon C_t$ on a non-empty subset of tasks χ if the same criterion holds for all tasks in that subset. In the latter case we also require $\tilde{C}_t >_\epsilon C_t$ to be unambiguous – that is we disallow cases in which the inequality could be reversed for an alternative choice of non-empty subset.

We are now ready to derive our candidate noncognitive abilities. The first of these are straightforward interpretations of the parameter distributions C_t, P_t^s, and P_t^f. Each of these distributions describes an aspect of individual heterogeneity at time t, namely: the agent's cost of effortful participation, her drive to succeed, and her resilience to failure. Our fourth candidate noncognitive ability is confidence, which emerges as a bias in the agent's subjective probability of success distribution Π_t. Our final noncognitive ability is the agent's propensity to think analytically, which arises when the normative decision model of Figure 4.1 is relaxed into the generalised decision framework presented in Embrey (2019e).

1 The Cost of Effortful Participation

A traditional view of the direct and opportunity cost of effortful task participation would be that this is a purely environmental factor. However, further consideration reveals that individual heterogeneity could important here too: there is no reason to suppose that all children will require an equivalent amount of effort to focus on school work or to forgo unrelated chatter and classroom antics. As such, it seems intuitively plausible that an exogenous reduction in an individual's cost of effortful task participation could increase their educational participation in expectation. Given the machinery above, we can identify sufficient conditions under which this intuition will hold:

Proposition 1 *A reduction in participation costs C can increase cognitive development:*

1. For discrete tasks X : *If there exists any period t, task $x \in X_t$, and positive $\epsilon \in \mathbb{R}$ such that the participation condition is binding on $\{x\}$ and such that $\tilde{C}_t <_\epsilon C_t$ on $\{x\}$; if signals for task x are continuously distributed over a connected support, and if the parameters c_t, d_t, p_t^s, p_t^f, and π_t are continuous functions of those signals, then $\lambda_t(\tilde{c}_t, \cdot) > \lambda_t(c_t, \cdot)$.*

2. For a continuum of tasks X : *If there exists any period t, tasks $\chi \subseteq X_t$, and positive $\epsilon \in \mathbb{R}$ such that the participation condition is binding on χ and such that $\tilde{C}_t <_\epsilon C_t$ on χ; if tasks and signals are jointly continuously distributed over a connected support, and if the parameters c_t, d_t, p_t^s, p_t^f, and π_t are continuous functions of those tasks and signals, then $\lambda_t(\tilde{c}_t, \cdot) > \lambda_t(c_t, \cdot)$.*

3. *Corollary: a strict increase in participation probability λ_t will lead to a strict increase in total expected human capital development $\mathbf{E}\sum_{\tau=t}^{T} V_\tau.\lambda_\tau.\delta^{\tau-T}$, either if:*

   A. *Tasks are perfect substitutes and the participation constraint remains binding for some $\chi_\tau \subseteq X_\tau$ for all $\tau > t$, or if:*

   B. *Present-period participation weakly increases the probability of future participation. (proof in Appendix A)*

Proposition 1 confirms our intuition that a monotonic decrease in an agent's period t cost distribution will, in quite general circumstances, have a positive effect on their expected participation likelihood in that period. Moreover, the corollary shows that, under reasonable additional assumptions, this increase in participation likelihood will strictly increase the total expected value of the human capital accumulated during the repeated supergame. Neither of these outcomes is certain for any individual, however they generate an empirical hypothesis that is testable at the population level. Children with lower costs of effortful participation over a span of periods are likely to exhibit greater total cognitive development on average during those periods, all else being equal. The empirical challenge, as ever, is to identify a causal effect of C within the constraints of a world where all else is rarely equal.

We discuss our empirical strategy toward that challenge in Section 4. Within that strategy, we measure the latent noncognitive skill of low participation cost using the observations summarised in table 4.1. We believe that the commonality between these measurements at each time period provides a plausible operationalisation of individual heterogeneity in the cost of effortful task participation.

Table 4.1: *Measurements of the cost of effortful task participation ($-C$)*

			Observations by data sweep				
Measurement	Reporter	Domain	3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
Shows interest in test stimuli	Visitor	$\{0, 1, 2, 3, 4\}$	14,368				
Responds well to requests	Visitor	$\{0, 1, 2, 3, 4\}$	14,368				
Combined: –restless; –fidgets (parent); –fidgets (visitor)	Vis. & Par.	$\approx\mathbb{R}\ -IRT^*$	15,356				
Moves to a new activity after finishing task	Parent	$\{0, 1, 2\}$	14,783	14,774	13,489		
Generally obedient	Parent	$\{0, 1, 2\}$	14,783				
Combined: obedient; responds well to requests	Parent	$\{0, 1, 2, 3\}$		14,773	13,488		
Combined: –restless; –fidgets	Parent	$\{0, 1, 2, 3, 4\}$		14,772	13,488	12,822	11,028
Combined: interest; involvement; motivation; concentration	Teacher	$\{0, 1, 2, 3\}$		3,201			
Combined: obedient; –restless; –fidgets	Teacher	$\{0, 1, 2, 3\}$; $\{0, 1\}$			8,712	7,275	
Combined: school interesting; –fed up at scl; –tired at school	Self	$\approx\mathbb{R}\ -IRT^*$			12,817		
Combined: school interesting; like scl; –tired at scl	Self	$\approx\mathbb{R}\ -IRT^*$				12,850	11,351
Combined: scl interesting; like scl; –hard to focus at scl	Self	$\approx\mathbb{R}\ -IRT^*$					11,351
Combined, –often: exhausted; restless; hard to concentrate	Self	$\approx\mathbb{R}\ -IRT^*$					11,125

NOTE: Noncognitive skill measurements in sweep 1 (age 9 months) are pleasantness, wariness, and fussiness.
*Combined using Item Response Theory, see Section 4.1.

2 Drive and Resilience

It is intuitively plausible that an individual's drive to succeed could increase her educational participation. We are able to refine this intuition by identifying our agent's drive to succeed as the parameter p^s, which quantifies the psychic payoff that our agent believes that she will enjoy if she is pleased with her performance. p^s may be derived from extrinsic feedback, such as the receipt of a grade that our agent perceives to be 'good', or it may be derived from an intrinsic sense of satisfaction. We can also separate our agent's drive to succeed from the related concepts of aspiration and thirst for learning, each of which would be manifest through d_t. The distinction between d_t and p^s is not merely semantic. According to our model, high aspiration would motivate educational participation even if that participation were certain to lead to an embarrassingly poor performance, because such participation would nevertheless generate learning. In contrast, a high drive to succeed would only motivate participation if there were some chance that participation would lead to the perception of success.

Conversely, if participation leads to failure then our agent anticipates paying the psychic cost $-p^f$. This payoff is the dual to p^s within our model, and it operationalises a concept of resilience. Intuitively, a more resilient agent has a larger (less negative) value of $-p^f$, and so she anticipates less severe disutility from perceived failure. Thus the resilient agent is more willing to attempt tasks where the probability of failure is non-negligible. Table 4.2 lists the measurements by which we operationalise P^s and P^f, and Proposition 2 formalizes their intuitive importance by supplying precise conditions under which an increase in either P^s or $-P^f$ would increase the likelihood of educational participation.

Proposition 2 *An increase in either drive P^s or resilience $-P^f$ can increase cognitive development:*

1. For discrete tasks X : *If there exists any period t, task $x \in X_t$, and positive $\epsilon \in \mathbb{R}$ such that the participation condition is binding on $\{x\}$ and such that $\tilde{P}^s_t >_\epsilon P^s_t$ on $\{x\}$; if signals for task x are continuously distributed over a connected support, if the parameters c_t, d_t, p^s_t, p^f_t, and π_t are continuous functions of those signals, and if there exists some $\delta > 0$ such that $\pi_t(x, y) > \delta$ for all $y \in Y_t$, then $\lambda_t(\tilde{p}^s_t(x, y), \cdot) > \lambda_t(p^s_t(x, y), \cdot)$.*

2. For a continuum of tasks X : *If there exists any period t, tasks $\chi \subseteq X_t$, and positive $\epsilon \in \mathbb{R}$ such that the participation condition is binding on χ and such that $-\tilde{P}^f_t >_\epsilon -P^f_t$ on χ; if tasks and signals are jointly continuously distributed over a connected support, if the parameters c_t, d_t, p^s_t, p^f_t, and π_t are continuous functions of those tasks and signals, and if there exists some*

Table 4.2: *Measurements of drive (P^s) and resilience ($-P^f$)*

			Observations by data sweep				
Measurement	Reporter	Domain	3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
			drive				
Positive reactions during cognitive testing	Visitor	$\{0,1,2,3,4\}$	14,368				
Beams with pride when praised	Parent	$\{0,1,2,3\}$	13,750				
Likes to work things out for self	Parent	$\{0,1,2\}$	14,783	14,774	13,490		
Sees tasks through to the end	Parent	$\{0,1,2\}$	14,783	14,769	13,485	12,677	11,210
Sees tasks through to the end	Teacher	$\{0,1,2\}$			8,718	7,291	
Likes to do well	Self	$\{0,1,2,3\}$				12,627	
Values hard work	Self	$\{0,1,2\}$					11,347
			resilience				
–Negative reactions during cognitive testing	Visitor	$\{0,1,2,3,4\}$	14,368				
Quickly gets over being upset	Parent	$\{0,1,2\}$	14,783	14,774	13,489		
–Easily Frustrated	Parent	$\{0,1,2\}$	14,783	14,773	13,490		
Cambridge Gambling Task risk score	Elicited	$\approx\mathbb{R}$				12,521	10,707
–Feels bad when does wrong	Self	$\{0,1,2,3\}$				12,512	
Willing to take risks	Self	$\{0,1,...,10\}$					11,222

NOTE: Noncognitive skill measurements in sweep 1 (age 9 months) are pleasantness, wariness, and fussiness.

$\delta > 0$ such that $\pi_t(x, y) < 1 - \delta$ for all $x \in \chi$, $y \in Y_t$, then $\lambda_t(\tilde{p}_t^f(x, y), \cdot) > \lambda_t(p_t^f(x, y), \cdot)$.

3. *Moreover, analogous results hold for P^s for a continuum of tasks, and for P^f for discrete tasks, and the corollary to Proposition 1 applies in all of the above cases.*

Proposition 2 is identical to Proposition 1, save that it demonstrates the importance of Π_t as a mediator of P^s and P^f. Drive to succeed will only affect participation if there is a non-negligible chance of success, whilst resilience is only important if there is a non-negligible chance of failure. Reflexively, Π_t will only be an important determinant of participation if at least one of P^s or P^f is sufficiently large.

3 Confidence

We have seen that an agent's probability of success distribution Π_t characterises her cognitive ability at time t. However, since the agent's payoffs are decision utilities, in precise terms it is her subjective cognitive ability that is represented by Π_t – and there is no guarantee that her subjective beliefs will be unbiased. In our model, bias could arise in the agent's subjective probability of success either through systematic inaccuracy in her prior beliefs, or through systematic inaccuracy in her signal interpretation. Both of these are plausible: for example there is evidence that many individuals over-estimate their own ability (Bénabou & Tiróle 2016), and that teachers and parents overweight praise in their feedback (Mueller & Dweck 1998). It is therefore likely that Π_t will contain an element of true cognitive ability, but also an element of subjective bias.

Positive bias in subjective cognitive ability Π_t is a manifestation of (over-) confidence.[6] As such, we might expect an individual who has greater confidence in her own ability to participate more readily in educational opportunities, since she will overweight the probability of obtaining the positive payoff of success p^s and underweight the probability of obtaining the negative cost of failure p^f. Proposition 3 confirms that confidence will have a positive effect on participation across a broad class of situations; a result which echoes the main theoretical prediction of Filippin & Paccagnella (2012), but which contradicts the predictions of Bénabou & Tirole (2002) and Köszegi (2006). Table 4.3 provides our empirical operationalisation of confidence as a noncognitive skill.

[6](Over-)confidence has been defined in this way by Bénabou & Tirole (2002), by Köszegi (2006), and by Filippin & Paccagnella (2012), amongst others.

Table 4.3: *Measurements of confidence ($\mathbf{E}\,\pi_t$)*

Measurement	Reporter	Domain	Observations by data sweep				
			3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
–Seems nervous or clingy	Parent	$\{0, 1, 2\}$	14,783	14,773	13,488		
–Often seems worried	Parent	$\{0, 1, 2\}$	14,783	14,773	13,488		
Perseveres with difficult tasks	Parent	$\{0, 1, 2\}$	14,783	14,773	13,488		
Volunteers to help others	Parent	$\{0, 1, 2\}$	14,783	14,772	13,487		
Combined: -worried; –nervous; volunteers to help	Parent	$\{0, 1, 2, 3, 4\}$				12,600	11,069
Combined: confident to try new things; perseveres	Teacher	$\{0, 1, 2\}$		3,186			
Combined: –worried; –nervous; –down-hearted; volunteers to help	Teacher	$\{0, 1, ..., 6\}$			8,682		
Combined: –worried; –nervous; volunteers to help	Teacher	$\{0, 1, 2, 3, 4\}$				7,253	
How often teacher thinks you're clever	Self	$\{0, 1, 2\}$			12,672		
Combined, how good at: English; maths; science	Self	$\approx\mathbb{R}\;-IRT^*$				12,801	11,355
How happy with school work	Self	$\{0, 1, ..., 5\}$				12,903	11,170
Combined, Rosenberg self-worth items†	Self	$\approx\mathbb{R}\;-IRT^*$				12,757	11,140
Combined: –worthless, –hopeless, –useless	Self	$\approx\mathbb{R}\;-IRT^*$					11,143

NOTE: Noncognitive skill measurements in sweep 1 (age 9 months) are pleasantness, wariness, and fussiness.
*Combined using Item Response Theory, see Section 4.1. †Satisfied with self; have good qualities; am capable; have value; feel good about self.

Proposition 3 *An increase in Π_t, due either to an increase in current cognitive ability or an increase in confidence, can increase future cognitive development:*

1. For discrete tasks X : *If there exists any period t, task $x \in X_t$, and positive $\epsilon \in \mathbb{R}$ such that the participation condition is binding on $\{x\}$ and such that $\tilde{\Pi}_t >_\epsilon \Pi_t$ on $\{x\}$; if signals for task x are continuously distributed over a connected support, if the parameters c_t, d_t, p_t^s, p_t^f, and π_t are continuous functions of those signals, and if there exists some $\delta > 0$ such that either $p^s(x, y) > \delta$ or $p^f(x, y) > \delta$ for all $y \in Y_t$, then $\lambda_t(\tilde{\pi}_t(x, y), \cdot) > \lambda_t(\pi_t(x, y), \cdot)$.*

2. *Moreover, analogous results hold for a continuum of tasks, and the corollary to Proposition 1 is applicable.*

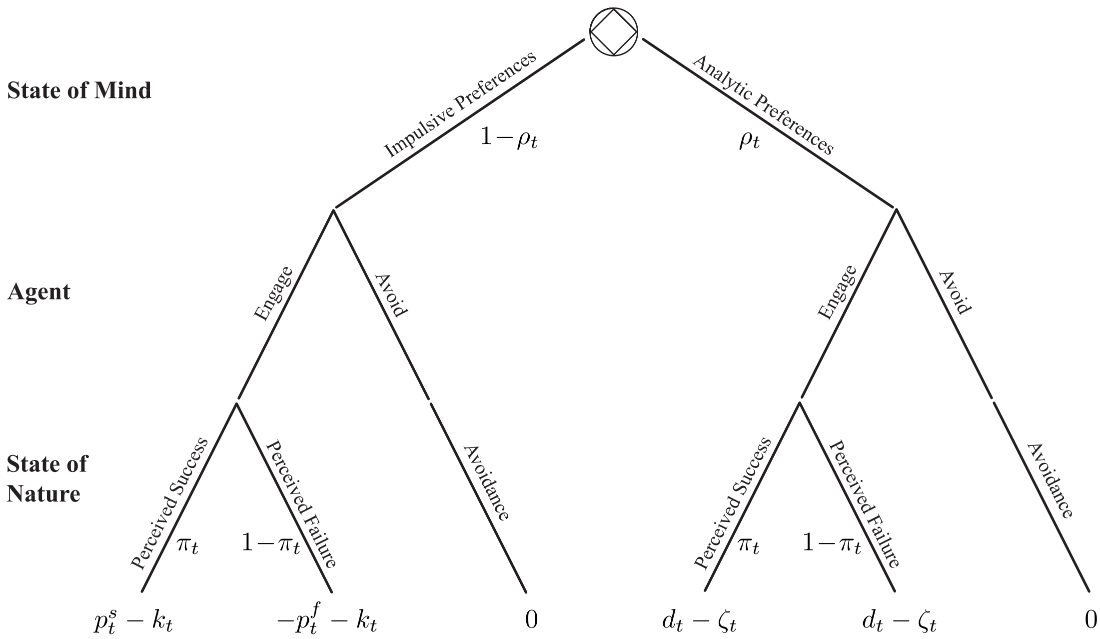
4 State of Mind

Our final theoretically driven noncognitive skill arises when the normative decision problem shown in Figure 4.1 is relaxed into the generalised decision framework of Embrey (2019e). The model presented in Figure 4.1 is predicated upon a strong set of assumptions which imply that all agents are required to act as if they trade off present payoffs against future payoffs in all decision periods. This set of assumptions is particularly restrictive for children, who frequently act without any consideration for the long-term consequences of their actions. On a more fundamental level, this assumption set requires that both childhood desires and adult returns to education can be meaningfully evaluated as quantities along the same unidimensional utility scale, which is difficult to justify.

The simplest generalisation of Figure 4.1 that relaxes the normative assumption set is shown in Figure 4.2. Here the agent will analyse future payoffs with probability ρ_t, or else they will follow their present impulses with probability $(1-\rho_t)$. The corresponding assumption set for this generalization is that all agents are required to act as if they prioritise either impulsive or analytic payoffs in each decision period. This assumption allows decisions to be made without any consideration of their long-term consequences, and, whilst it still requires that all payoffs must admit numerical valuations, it allows those valuations to lie along independent dimensions of the agent's utility space. An extended comparison between the foundational assumptions of the normative decision theory and of the generalised decision theory is provided in Embrey (2019a).

Under the generalised model of Figure 4.2, each of two possible states of mind is represented by a standard Neoclassical decision problem. Thus the psychic payoffs to perceived success and failure only affect the agent under her impulsive state of mind, whilst the long-term developmental payoff to educational participation is only salient under an analytic state of mind. Similarly, we may decompose the cost

Figure 4.2: *A Representative Agent's Participation Decision*



Note: Adapted from Embrey (2019e) with the author's permission.

of effortful task participation c_t into two components k_t and ζ_t, which evaluate the impulsive and analytic aspects of that participation cost respectively. The total cost c_t includes factors such as 'not being bothered', which lies entirely within the impulsive component k_t, and missing out on gossip, which lies within k_t to the extent that it that seems tantalizing in the moment, but which lies within ζ_t to the extent that it has any lasting analytic value. Since the lasting analytic value of classroom gossip is likely to be small, and since children of compulsory schooling age are required to be present in the classroom whether or not they engage with the present educational opportunity, we note that the analytic participation cost ζ_t will typically be very low, whilst k_t will typically be much greater.

These considerations imply that an individual's state of mind ρ_t may be an important noncognitive determinant of her educational development. If the agent adopts an impulsive mindset in period t, then she will participate in the present educational opportunity if and only if her expected enjoyment from feeling successful exceeds both her fear of failure and her relatively high impulsive appraisal of her participation cost. However, if the agent adopts an analytic mindset in period t, then she will participate if and only if the educational benefit of doing so exceeds her relatively low analytic appraisal of her participation cost. We should therefore expect that an increase in the probability ρ_t with which an agent adopts analytic preferences will increase her cognitive development except in rather contrived circumstances.

Observation 1 *An increase in the state-of-mind parameter ρ_t will increase an agent's total cognitive development unless the quantity of cognitive development that would be realised under her best analytic effort to optimize that development is lower than than the accidental cognitive development that would be realised through her pursuit of immediate gratification.*

We will refer to ρ_t as the agent's idiosyncratic propensity to think analytically. It has been argued that this propensity can be measured by the Cognitive Reflection Test of Frederick (2005); indeed it is in that context that the term *propensity to think analytically* was first proposed (Pennycook et al. 2016; Pennycook & Ross 2016). If this is true, then there is substantial evidence to suggest that an individual's propensity to think analytically may have a pervasive influence across a wide range of life outcomes (for reviews see Pennycook, Fugelsang & Koehler 2015; Brañas-Garza, Kujal & Lenkei 2019). That possibility is investigated by Embrey (2019e), where the decision-theoretic model presented in Figure 4.2 is applied to a wide range of decision contexts to find theoretical evidence that an individual's propensity to think analytically could indeed have a pervasive influence over life outcomes.

In the present chapter, we focus on the outcome of cognitive development during childhood. Our data don't include any Cognitive Reflection Test items, however Table 4.4 demonstrates that they do include multiple empirical measures of each individual's propensity to consider the wider consequences of her actions. The nature of these measurements underscores the distinction between an agent's propensity to think analytically and her discount factor. A discount factor determines how inter-temporal utility comparisons are made, and it is typically measured by forcing a test subject to compare two or more alternatives. Such measurements avoid the fundamental question of whether a comparison is truly possible, because they ensure that the alternative payoffs are naturally located upon a single, usually monetary, scale. In doing so, experimental elicitations of a subject's discount factor induce the same pecuniary state of mind for the evaluation of both alternatives. In our situation, we propose that an impulsive state of mind would focus on the present-period payoffs of educational participation, whilst an analytic state of mind would focus on its developmental payoffs. It is not clear that either state of mind would naturally trade-off the two disparate sets of motivations against each other using an inter-temporal discount factor.

Table 4.4: *Measurements for state of mind (ρ)*

			Observations by data sweep				
Measurement	Reporter	Domain	3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
–Impulsive / acts without thinking	Parent	$\{0,1,2\}$	14,783	14,772	13,488		
Chooses her own activities	Parent	$\{0,1,2\}$	14,783	14,774	13,490		
–Easily distracted	Parent	$\{0,1,2\}$	14,783	14,773	13,488		
Stops to think before acting	Parent	$\{0,1,2\}$	14,783	14,767	13,485		
Considerate of others' feelings	Parent	$\{0,1,2\}$	14,783	14,772	13,488		
Combined: considers; shares; –easily distracted; thinks before acting	Parent	$\approx\mathbb{R}\ -IRT^*$				12,766	11,230
How often patient	Parent	$\{0,1,...,10\}$					11,269
How good at controlling emotions	Parent	$\{0,1,2,3,4\}$					11,286
Focus during cognitive testing	Visitor	$\{0,1,2,3,4\}$	14,368				
Combined: considers consequences; awareness of: others; right & wrong	Teacher	$\{0,1,2,3\}$			3,156		
Combined: considers; shares; –easily distracted; thinks before acting	Teacher	$\{0,1,...,7\}$			8,658		
Combined: as above + works independently	Teacher	$\approx\mathbb{R}\ -IRT^*$				7,305	
–How often lose temper	Self	$\{0,1,2\}$			12,813		
–How often get angry	Self	$\{0,1,2\}$				12,662	
–How often school a waste of time	Self	$\{0,1,2,3\}$				12,811	11,345
Combined: –use of cigarettes; drugs	Self	$\{0,1,2,2.5,3\}$					11,121

NOTE: Noncognitive skill measurements in sweep 1 (age 9 months) are pleasantness, wariness, and fussiness.
*Combined using Item Response Theory, see Section 4.1.

4 The Empirical Model

The foregoing section has produced five clear theoretical predictions; namely that cognitive ability development should be increasing in each of: drive, resilience, confidence, the cost of effortful participation, and the propensity to think analytically. We therefore adapt the empirical model pioneered by Cunha, Heckman & Schennach (2010) to test each of these hypotheses. Since detailed discussions of this empirical model are provided by those authors and in Chapter 3 (Embrey 2019d), we will present only a brief résumé here.

Our goal is to analyse the simultaneous production of cognitive and noncognitive skills through childhood. For tractability, we will assume that the relevant inputs at time t are captured by the state vector

$$(\theta_t^C, \theta_t^{NC}, I_t, \vartheta_t^C, \vartheta_t^{NC}),$$

where θ_t^C and θ_t^{NC} denote the child's cognitive and noncognitive skill levels, where ϑ_t^C and ϑ_t^{NC} denote parent cognitive and noncognitive skill levels, and where I_t denotes parental investment. Our operationalisation of each component of this vector through childhood is shown in the tables of Section 3 and Appendix C.

Our object of interest is the technology by which cognitive and noncognitive skills are produced. We will parametrise this technology using the CES functional form, normalised to the sample mean level of each input. Thus we assume that, for individuals i at developmental stage s:

$$\theta_{i,t+1}^C = \left(\gamma_{s,1}^C\left(\frac{\theta_{i,t}^C}{\Theta_t^C}\right)^{\varphi_s^C} + \gamma_{s,2}^C\left(\frac{\theta_{i,t}^{NC}}{\Theta_t^{NC}}\right)^{\varphi_s^C} + \gamma_{s,3}^C\left(\frac{I_{i,t}}{\Theta_t^I}\right)^{\varphi_s^C} + \gamma_{s,4}^C\left(\frac{\vartheta_{i,t}^C}{\Theta_t^{PC}}\right)^{\varphi_s^C} + \gamma_{s,5}^C\left(\frac{\vartheta_{i,t}^{NC}}{\Theta_t^{PNC}}\right)^{\varphi_s^C}\right)^{\frac{1}{\varphi_s^C}} e^{\eta_{i,t,C}}$$

$$\theta_{i,t+1}^{NC} = \left(\gamma_{s,1}^{NC}\left(\frac{\theta_{i,t}^C}{\Theta_t^C}\right)^{\varphi_s^{NC}} + \gamma_{s,2}^{NC}\left(\frac{\theta_{i,t}^{NC}}{\Theta_t^{NC}}\right)^{\varphi_s^{NC}} + \gamma_{s,3}^{NC}\left(\frac{I_{i,t}}{\Theta_t^I}\right)^{\varphi_s^{NC}} + \gamma_{s,4}^{NC}\left(\frac{\vartheta_{i,t}^C}{\Theta_t^{PC}}\right)^{\varphi_s^{NC}} + \gamma_{s,5}^{NC}\left(\frac{\vartheta_{i,t}^{NC}}{\Theta_t^{PNC}}\right)^{\varphi_s^{NC}}\right)^{\frac{1}{\varphi_s^C}} e^{\eta_{i,t,NC}}$$

$$\text{where:} \quad \sum_{j=1}^{5}\gamma_{s,j}^k = 1; \quad \varphi_s^k \in [-\infty, 1]\,; \;\; \forall s;\; k \in \{C, NC\}. \tag{4.5}$$

In Equation (4.5) the $\gamma_{s,j}^k$ are share parameters, that is they represent the share of total resource that one should optimally allocate to input j for the formation of skill k during developmental stage s at the sample mean input level $(\Theta_t^C, \Theta_t^{NC}, \Theta_t^I, \Theta_t^{PC}, \Theta_t^{PNC})$. Thus the share $\gamma_{s,2}^C$ represents an intuitive concept of the 'value' to the average child of period t noncognitive skills θ_t^{NC} in the production of period t+1 cognitive skills θ_{t+1}^C during stage s of childhood development. Equivalently, $\gamma_{s,2}^C$ represents the output elasticity of θ_t^{NC} in the production of θ_{t+1}^C

at the sample mean input level $\boldsymbol{\Theta_t}$.[7]

The $\gamma_{s,2}^C$ are therefore our main parameters of interest, because they quantify the importance of the present candidate noncognitive ability at each stage of childhood cognitive development. However the φ_s^C are also of interest to us, because they indicate the extent to which noncognitive skills can substitute for current cognitive skills in the production of future cognitive skills. The elasticity of substitution between inputs for cognitive development during stage s is given by $1/(1-\varphi_s^C)$, and an elasticity above unity would indicate that the inputs are reasonably substitutable, whereas an elasticity of below unity would indicate that the inputs are mutually complementary.[8]

The remaining aspects of the econometric model are important for its validity, but they are not central to our present purpose. First, each state variable is considered to be latent, since none is directly observable. We therefore require two or more measurement equations that proxy for each state variable in each period in order to identify each as the factor-analytic commonality between its measurements.[9] However, our factor models depart from those of Cunha, Heckman & Schennach (2010) to ensure that we do not impose multiple conflicting sets of identifying restrictions. In particular we normalise the location and scale of each latent factor during period one, but we do not impose any additional conflicting normalisations thereafter. The consequences of imposing additional normalisations were highlighted by Agostinelli & Wiswall (2016b), and they are summarised in Chapter 3.

Second, we require parametric assumptions regarding the error terms $\eta_{i,t,k}$. In the case that investment is endogenous because parents respond to private signals of next-period skills, this term may be correlated with observed investment. We therefore instrument parental investment with (OECD equivalised) household income (controlling for the other input variables), and allow the error term $\pi_{i,t}$ which represents unobserved heterogeneity within that investment equation to appear within $\eta_{i,t,k} := \gamma_{s,6}^k \pi_{i,t} + \nu_{i,t,k}$, where the $\nu_{i,t,k}$ are assumed to be normally distributed with mean zero, and independent across skills and periods.

Third, we require production technologies for parental skills and for family income in analogue to Equation (4.5). Parental cognitive skill is assumed to be

[7]See Embrey (2019c) for a detailed interpretation of share parameters within the normalised CES production function.

[8]In precise terms, as the elasticity diverges to infinity the (normalised) CES form approaches a (normalised) linear production technology with coefficients given by the $\gamma_{s,j}^k$; as the elasticity converges to zero the CES form approaches a Leontief production technology with input requirements given by the $\Theta_{t,j}$; and as the elasticity converges to unity the CES form approaches a Cobb-Douglas production technology with shares given by the $\gamma_{s,j}^k$.

[9]We also require one additional measurement in each period to achieve non-parametric identification; for this and other details of the identification analysis see Cunha, Heckman & Schennach (2010).

constant due to data constraints, whilst parental noncognitive skill and family income are assumed to follow AR(1) processes which ensure that their respective factor-models are just-identified: as such parental noncognitive skills are mapped directly forward into the next period to provide both location and scale for that latent factor, whereas the AR(1) coefficient of family income is freely estimable since we assume that our single measure of family income is accurate in each period, which implies that the scale of that factor is already known.

Our estimation procedure adapts the code used in Chapter 3 (Embrey 2019d), which in turn builds upon the procedure developed by Cunha, Heckman & Schennach (2010). Under this procedure, the technology and ancillary equations are estimated simultaneously via maximum likelihood iteration, which allows for efficient data utilisation. A guide to the use and adaptation of this code is appended to this thesis in Section A3.

1 Data

Our empirical model requires a large-scale longitudinal dataset with multiple measures of child and parental skills and investments, including multiple measures of each candidate noncognitive skill at each period in time. As can be seen from the tables in Section 3 and Appendix C, the UK Millennium Cohort Study meets these requirements.

MCS subjects were born around the year 2000, and the first MCS survey at around age 9 months reached 18,818 children from across the UK, with an oversampling of deprived and ethnic minority areas. We drop 272 of these children because they represent multiple observations from the same household, and so the unobserved heterogeneity within their production functions may not be statistically independent. With this exception, we base our estimates on the available sample at each observation point. The MCS has achieved a lower attrition rate than many longitudinal surveys, with 11,714 families responding in the latest sweep, which is sweep 6 at age 14. Data access is controlled via the UK Data Service, and further information can be found at https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000031.

The MCS compares favourably with the CNLSY data used by Cunha, Heckman & Schennach (2010) and in Chapter 3: it provides suitable measures for all candidate noncognitive abilities; its measures of cognitive ability in sweep 1 are considerably more relevent; and it has a much larger sample size. In reality, the MCS contains too many measures for some latent factors, and so we combine groups of highly correlated measurements together in advance of the main estimation.[10] In complex cases, the measurement groupings were guided by exploratory factor analysis similar to that used by Heckman, Pinto & Savelyev (2013). Where possible we combined related measures using Item Response Theory, which estimates a latent factor score for each individual by allowing each response for each measurement to have separate difficulty and discriminatory properties. Because of this, Item Response Theory makes considerably better use of large-scale data than factor analysis, which allows only one loading parameter for each measurement item[11]. Where data were too restricted or too skewed for Item Response Theory, they were combined by hand into appropriate bands. The Stata code that compiles our working dataset from the many MCS data files is included within our supplementary materials, and further discussion of the data compilation process is provided in our supplementary materials (A3).

5 Results and Discussion

Our main results for the technologies of cognitive and noncognitive skill production are presented in Tables 4.5 and 4.6 respectively. These tables display our estimates of the share parameters and elasticities of substitution from Equation (4.5), while our estimates of the auxiliary parameters and equations are provided in Appendix B.

The first column in Tables 4.5 and 4.6 reproduces known results from the American CNLSY sample for comparison. These are were derived under Model 1 of Chapter 3 (Embrey 2019d), which builds upon the methodology pioneered by Cunha, Heckman & Schennach (2010). The third column of each table replicates that benchmark specification within the larger and more recent dataset provided by the Millennium Cohort Study in the UK.[12] Focusing first on our estimates for the technology of cognitive skill production, we can see that there are substantial differences between the results of columns one and three. Most strikingly, the estimated first-stage input share of present-period cognitive skill is substantially lower in the CNLSY sample than in the MCS sample, with the bulk of that short-fall appearing in the coefficient on parental investments. The discrepancy between these estimates manifests institutional differences, but it is also driven by differences in data availability between the two cohorts.

Empirically, the estimated coefficients are derived from their underlying data.

[10]Each additional measurement within the main estimation would cost seven additional parameters, and the parameter space affects our computational requirements approximately quadratically.

[11]See Embretson & Reise (2000) for an overview.

[12]The dataset is described in Section 4.1; Appendix C details the measures of prosocial and antisocial behaviours that are used in our operationalisation of the benchmark specification.

Table 4.5: *Estimates of the Technology of Cognitive Skill Production: Input shares and elasticities*

	First stage (0–5 years)							
	behaviors			–cost	confidence	resilience	drive	analyticity*
	CNLSY†	\wave 1‡	MCS	$-C$	$\mathbf{E}(\pi)$	$-P^f$	P^s	ρ
Cognitive Skill $\gamma^C_{1,1}$	.342	.651	.835	.785	.817	.822	.755	.831
	(.041)	(.099)	(.007)	(.008)	(.008)	(.007)	(.008)	(.008)
Noncognitive Skill $\gamma^C_{1,2}$	.001	.029	.031	.090	.037	.033	.098	.026
	(.030)	(.027)	(.007)	(.007)	(.007)	(.005)	(.006)	(.008)
Parental Investments $\gamma^C_{1,3}$	.486	.115	.003	.002	.004	.003	.000	.002
	(.111)	(.135)	(.006)	(.005)	(.008)	(.007)	(.008)	(.009)
P. Cog. Skill $\gamma^C_{1,4}$	.086	.105	.128	.121	.135	.137	.145	.135
	(.027)	(.030)	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
P. Noncog. Skill $\gamma^C_{1,5}$	.084	.100	.003	.002	.007	.004	.002	.007
	(.058)	(.035)	(.005)	(.004)	(.005)	(.005)	(.005)	(.005)
e-subs $1/(1-\varphi^C_1)$	.893	.594	.595	.530	.676	.656	.671	.673
	(.136)	(.083)	(.014)	(.012)	(.015)	(.014)	(.012)	(.016)
	Second stage (5–14 years)							
Cognitive Skill $\gamma^C_{2,1}$	.923	.932	.886	.872	.890	.906	.848	.793
	(.015)	(.015)	(.008)	(.008)	(.009)	(.009)	(.010)	(.009)
Noncognitive Skill $\gamma^C_{2,2}$	.002	.004	.035	.050	.024	.009	.066	.138
	(.002)	(.003)	(.010)	(.010)	(.010)	(.010)	(.007)	(.010)
Parental Investments $\gamma^C_{2,3}$	.000	.000	.001	.001	.001	.001	.002	.002
	(.028)	(.024)	(.018)	(.019)	(.019)	(.017)	(.014)	(.015)
P. Cog. Skill $\gamma^C_{2,4}$	.074	.061	.078	.076	.085	.084	.084	.067
	(.019)	(.015)	(.014)	(.014)	(.014)	(.014)	(.011)	(.010)
P. Noncog. Skill $\gamma^C_{2,5}$	.002	.002	.000	.000	.000	.000	.000	.000
	(.026)	(.021)	(.008)	(.008)	(.008)	(.009)	(.007)	(.007)
e-subs $1/(1-\varphi^C_2)$	.500	.482	1.013	1.199	1.078	1.016	1.536	1.427
	(.043)	(.045)	(.088)	(.112)	(.096)	(.102)	(.138)	(.112)

NOTE: Standard errors in parentheses. †Chapter 3 (Embrey 2019d) Model 1; ‡CNLSY omitting wave 1. *The propensity to think analytically.

Whilst most measurements at most waves are reasonably commensurate between the two data sources, the CNLSY is not well-endowed with cognitive measures during its first wave. A motor-social development score is observed during wave one, but only for 9% of the sample, and so Cunha, Heckman & Schennach (2010) use gestation length and birth weight to proxy for early cognitive development. We would therefore expect the estimated self-productivity of early cognitive skills to be deflated in the CNLSY sample because those proxy measurements are likely to be less informative than the more direct measurements provided by the MCS (these are detailed in Table 4.C2). To evaluate the economic significance of this data deficiency, column two of Table 4.5 replicates column one except that it omits the first wave of observations. These results indicate that the majority of the empirical discrepancy between the MCS and CNLSY cohorts could be driven by deficiencies in the CNLSY data.[13]

The remaining discrepancy between columns two and three can be attributed to institutional and cultural factors. For example, enrolment rates in pre-school education are substantially lower in the US than in the UK (OECD 2017), and so we might expect direct parental inputs to have a commensurately larger role in the early cognitive development of the US cohort. Our results show that this is indeed the case: point estimates for parental investments and parental noncognitive skills are substantially larger in the American sample, although the former is estimated rather imprecisely. Our estimates for the second stage of childhood suggest that the two samples converge towards a common production technology for cognitive skills once formal schooling begins, except that the elasticity of substitution between alternative inputs becomes significantly greater for children from the UK Cohort during that stage of development.

We now turn to a comparison between the benchmark MCS results in column three of Table 4.5, and the alternative estimates derived for each candidate noncognitive skill. From the cross-productivity coefficients $\gamma^C_{s,2}$ it is immediately apparent that generic proxy measurements for noncognitive skills tend to under-estimate their importance. However, we can also see that the specific noncognitive skills of confidence and resilience are, if anything, less important in the production of childhood cognitive development than are antisocial and prosocial behaviours.[14] This is perhaps surprising, because the most robust implication from existing economic theory is that confidence should affect cognitive development (see, for example Bénabou & Tirole 2002; Köszegi 2006; Filippin & Paccagnella 2012) and because there is now a substantial body of literature which follows Duckworth et al.

[13] Note that distortions in estimated latent factors from the first wave are also likely to affect future waves in an unpredictable manner, due to the dynamic development of each individual's latent skill levels through time.

[14] Of course, these noncognitive skills could be substantial determinants of other life outcomes.

(2007) in suggesting that grit should be an important determinant of educational outcomes.[15] However, despite the close linguistic relationship between grit and resilience, only two of the twelve grit survey items relate to resilience as defined by P^f. These observations illustrate the benefits of working with precisely defined measures that have been derived through rigorous theoretical analysis.

The remaining columns of Table 4.5 show that the noncognitive skills defined by $-C$, P^s, and ρ are substantial determinants of cognitive development during childhood. During a child's early years, her cost of effortful task participation and her drive to succeed each account for at least 9% of her optimal period-on-period inputs to cognitive development. Since cognitive skills are also shown to be strongly self-productive, the effects of these noncognitive inputs will also accumulate over time, so that by the age of seven approximately 23% of cumulative cognitive development could be attributed to the child's drive to succeed as measured during waves 1-3 of the MCS.[16] Given that inputs accumulate in this way, our estimate that the average child's propensity to think analytically accounts for 14% of her period-on-period cognitive development in later childhood is a striking result.

When taken together, our results provide further evidence that the most important determinant of ongoing cognitive development is current cognitive ability. This suggests that an individual's developmental pathway may often be determined during the early years of her life. However, to the extent that this pathway remains malleable, the factors that have greatest influence over it are her parents' cognitive skill level and her own noncognitive skills – in particular her drive to succeed, her cost of effortful task participation, and her propensity to think analytically. These results are commensurate with the model of educational investment that is depicted in Figure 4.2. During a child's early years, it is the cognitive ability level of a her parents that determines, either directly or indirectly, the educational opportunities that are available to her. Thereafter, as the child approaches compulsory schooling age, the educational input from her parents diminishes in importance. Meanwhile, the child's responses to her early educational opportunities are determined by her impulsive payoffs – chiefly by the trade-off between the pleasure of successful task participation and the cost of exerting the effort to participate. Thereafter, as the child matures, she is increasingly able to impose analytic preferences over her decision-making, and so her impulsive payoffs are invoked less frequently. Thus, once the child reaches compulsory schooling age, her own propensity to think analytically becomes the most important outside

[15]Although a recent meta-analysis questions both the construct validity and the predictive power of grit (Credé, Tynan & Harms 2017).

[16]This value is given by $0.98(1 + 0.755 + 0.755^2)$.

determinant of her ongoing cognitive development.

The results presented in Table 4.5 therefore support the conventional view that developmental pathways for cognitive skills stabilise early in life, but they also challenge the finality of that conclusion. From the prior research reproduced in the first two columns of Table 4.5, we would conclude that only parental cognitive skills can affect a young person's developmental pathway once they reach compulsory schooling age. In contrast, our estimates offer hope that some specific aspects of noncognitive skill can also have a substantial influence over that pathway in later childhood, and that these noncognitive skills can act as economic substitutes for current cognitive ability. Our findings therefore imply that, if a policy intervention could successfully improve a child's propensity to think analytically, then it could have a lasting positive influence over the likelihood that she will engage with each incremental educational opportunity thereafter. This possibility is further substantiated by the results presented in the final column of Table 4.6, which suggest that an individual's propensity to think analytically becomes strongly self-perpetuating from later childhood.

There is accumulating evidence that it may indeed be possible to improve an individual's propensity to think analytically through policy intervention. Several recent papers have developed interventions that target metacognition[17] (Tanner 2012; Zepeda et al. 2015; Casselman & Atwood 2017), and Dignath & Büttner (2008) synthesise a large number of earlier studies that have a similar focus. In addition, there is now a scientific journal dedicated to the study of *Metacognition and Learning*. Furthermore, work is underway to develop an online intervention for economics undergraduates that explicitly targets ρ, and a randomised controlled pilot study of that intervention has produced an eight percentage-point increase in exam performance.[18]

Although our main object of interest is the technology of cognitive skill formation, we also present results on the technology by which each of our candidate noncognitive skills is produced. The first three columns of Table 4.6 suggest that these technologies are largely comparable between the CNLSY and MCS cohorts, and that the most important determinant of early noncognitive skill formation is parental noncognitive skill levels. Our measures of parental noncognitive skill follow Cunha, Heckman & Schennach (2010) in focussing on maternal emotional stability (see Table 4.C3), and so it is unsurprising that early resilience, confidence, and social behaviors are most strongly affected by this production input. In contrast, parental cognitive skills are typically unimportant for child noncognitive skill development, except for specific cases where they probably act as a proxy for a

[17] Metacognition is defined as an awareness and understanding of one's own thought processes.
[18] Pilot study results and details are available from the author upon request.

secure family background: early resilience, behaviors, and the cost of exerting educational effort could each be affected in this way. Interestingly, tangible parental investments could also proxy for a stable family background, and yet these are not significant determinants of any noncognitive skill in Table 4.6, although there is some indication that they may serve to build confidence in later childhood, and that they may affect the later (non-)development of antisocial behaviours within the CNLSY cohort.

The most substantial variation between the production technologies of our candidate noncognitive skills is found in the cross-productive influence of cognitive skills. For example, there is strong evidence in Table 4.6 that one needs to experience early success in order to develop one's drive to succeed. This mechanism will act to reinforce the early stabilisation of cognitive skill, because it implies that children who have comparatively high cognitive skill levels are most likely to continue to engage with educational opportunities in order to seek further gratification. Similarly, column five of Table 4.6 appears to suggest that cognitive skill is important for the development of confidence throughout childhood, however that finding may be an artefact of our measures of confidence: many of these will necessarily contain an element of true ability in addition to confidence (which we defined in Section 3 as perceived ability net of true ability). This is not the case for the cost of effortful task participation. We find little evidence that cognitive skills reduce this psychic participation cost, and so there is limited scope for confounding within our measures of $-C$. These results challenge the main assumption behind the signalling hypothesis of Spence (1973), since they suggest that less cognitively skilled individuals may not experience a substantially greater cost of educational investment than their more-able peers. Our findings for the technology by which $-C$ is produced therefore support the alternative hypothesis that cognitive skills can be developed through educational participation.

Our technology estimates for the production of resilience also have important policy implications. Table 4.6 shows that, as a child matures, it will be increasingly essential for her to experience success in order for her to become resilient to failure. Moreover, we find that ongoing success remains necessary even for children who had high levels of resilience two years previously: by compulsory schooling age, the self-productivity share of previous-period resilience is only 15%, and the elasticity of substitution in the production of resilience is substantially below unity. These estimates provide an alternative explanation for the oft-observed social gradient in risk aversion: that gradient need not imply that less cognitively skilled individuals are less able to calculate expected benefits, instead our results suggest that successfulness itself may allow individuals to be more resilient to the possibility of failure. If so, those of us who lead successful lives should judge less harshly

Table 4.6: *Estimates of the Technology of Noncognitive Skill Production: Input shares and elasticities*

	First stage (0–5 years)							
	behaviors			–cost	confidence	resilience	drive	analyticity*
	CNLSY†	\wave 1‡	MCS	$-C$	$\mathbf{E}(\pi)$	$-P^f$	P^s	ρ
Cognitive Skill $\gamma^C_{1,1}$	.001	.037	.054	.032	.127	.063	.284	.072
	(.039)	(.096)	(.006)	(.004)	(.007)	(.009)	(.011)	(.004)
Noncognitive Skill $\gamma^C_{1,2}$	.693	.884	.712	.856	.748	.697	.643	.845
	(.056)	(.064)	(.003)	(.005)	(.008)	(.008)	(.011)	(.004)
Parental Investments $\gamma^C_{1,3}$	.000	.000	.003	.000	.003	.003	.018	.001
	(.129)	(.124)	(.006)	(.005)	(.012)	(.013)	(.019)	(.006)
P. Cog. Skill $\gamma^C_{1,4}$	.077	.038	.065	.037	.002	.038	.000	.021
	(.029)	(.032)	(.003)	(.002)	(.005)	(.008)	(.008)	(.002)
P. Noncog. Skill $\gamma^C_{1,5}$	.229	.039	.166	.074	.119	.199	.055	.061
	(.068)	(.050)	(.005)	(.003)	(.006)	(.008)	(.009)	(.003)
e-subs $1/(1-\varphi^C_1)$	.642	.493	.641	.775	1.018	.905	.972	1.007
	(.083)	(.136)	(.014)	(.023)	(.049)	(.031)	(.039)	(.036)
	Second stage (5–14 years)							
Cognitive Skill $\gamma^C_{2,1}$	.000	.000	.000	.011	.176	.845	.124	.000
	(.007)	(.008)	(.005)	(.005)	(.014)	(.060)	(.013)	(.004)
Noncognitive Skill $\gamma^C_{2,2}$	.845	.862	.996	.980	.772	.150	.864	.960
	(.014)	(.013)	(.013)	(.008)	(.015)	(.055)	(.010)	(.006)
Parental Investments $\gamma^C_{2,3}$	.055	.069	.001	.002	.046	.000	.012	.013
	(.063)	(.048)	(.010)	(.005)	(.031)	(.004)	(.021)	(.005)
P. Cog. Skill $\gamma^C_{2,4}$	.000	.000	.000	.000	.004	.005	.000	.001
	(.017)	(.012)	(.006)	(.004)	(.022)	(.017)	(.017)	(.003)
P. Noncog. Skill $\gamma^C_{2,5}$	.100	.069	.002	.007	.001	.000	.000	.026
	(.056)	(.041)	(.006)	(.004)	(.013)	(.010)	(.017)	(.004)
e-subs $1/(1-\varphi^C_2)$	.664	.599	3.703	.664	1.072	.446	.867	11.516
	(.054)	(.057)	(17.554)	(.156)	(.069)	(.096)	(.068)	(18.604)

NOTE: Standard errors in parentheses. †Chapter 3 (Embrey 2019d) Model 1; ‡CNLSY omitting wave 1. *The propensity to think analytically.

the economically inefficient risk-aversion of those who are less fortunate – instead we could improve that efficiency by putting policies in place to make educational investment appear less risky for disadvantaged individuals.

6 Conclusion

Noncognitive skills are the focus of an extensive body of literature. That literature has proposed a wide range of potential noncognitive skills, but these are typically derived from intuitive arguments or as the commonality between psychological survey items. In response, this chapter analyses a decision-theoretic model of educational investment to identify five specific mechanisms through which an aspect of noncognitive skill could affect childhood cognitive development. Each of these mechanisms generates a rigorous theoretical definition for one specific noncognitive skill, and it generates a testable hypothesis that childhood cognitive development should be affected by each of those skills. Our candidate noncognitive skills are: one's academic confidence, one's resilience to failure, one's drive to succeed, one's cost of effortful task participation, and one's propensity to think analytically. This contribution develops economic understanding from vague notions of noncognitive skill based upon superficially cogent constructs to specific individual characteristics defined by parameters within a decision-theoretic model.

Our second contribution is to test the importance of each candidate noncognitive skill. Existing empirical evidence for the effects of noncognitive skills is almost exclusively derived from reduced-form estimations, but this approach does not account for the fact that cognitive and noncognitive skills develop simultaneously throughout childhood and beyond. Because of this, the partial effect of noncognitive skills in simple controlled regressions may not provide an accurate estimate of their true developmental importance. We therefore build upon the structural econometric model of Cunha, Heckman & Schennach (2010) to estimate the simultaneous technologies by which both cognitive and noncognitive skills are produced. Our results corroborate our theoretical hypotheses by finding that each candidate noncognitive ability is a statistically significant determinant of childhood cognitive development. In addition, our results allow a comparison between US and UK childcare provision, and our simultaneous technology estimates allow us to suggest not only when, but also *how* public policy interventions might target noncognitive skills to improve the educational outcomes of disadvantaged children.

The central message from existing research into childhood cognitive development is that early intervention is vital, because a child's cognitive developmental pathway stabilises during her early years. Our findings support this message, but with one important qualification. We find that the average child's idiosyncratic propensity to think analytically becomes an increasingly important determinant of her cognitive development as she matures. Between the ages of 5-14 years, this noncognitive skill accounts for 14% of her optimum period-on-period input to cognitive skill production, and it acts as an economic substitute for her current-period cognitive skill level. Since we also estimate a high degree of persistence in this specific noncognitive skill, our results suggest that later interventions may achieve a lasting positive effect on educational development if they target an individual's propensity to think analytically.

This conclusion may also have wider implications, both for public policy and for economic theory. An individual's propensity to think analytically arises as a decision parameter under the generalisation of Expected Utility Theory that was presented in Figure 4.2. Although we apply the generalized decision theory to educational decisions, such as whether to attempt classwork or study for a test, the same model could equally describe health decisions, such as whether to exercise rather than watch T.V.; financial decisions, such as whether to save up for an affordable car rather than purchasing an expensive one on credit; and political decisions, such as whether to take rhetoric at face-value rather than enquiring more deeply. Since our present results provide strong empirical support for this generalised decision theory within an educational context, it could be fruitful for future research to test whether an individual's propensity to think analytically is similarly important in other decision-making contexts. If so, interventions which target that specific noncognitive ability could provide an effective policy lever through which to address a cluster of negative life outcomes.

A Mathematical Appendix

1 Proof of Proposition 1

For the discrete case: Fix $t \leq T \in \mathbb{N}$, $x \in X_t$, and $\epsilon \in \mathbb{R}$ such that $\tilde{C}_t <_\epsilon C_t$ on $\{x\}$ and such that the participation condition (4.2) is binding on $\{x\}$. Then rearrange the participation condition to define $\varrho_t(x,y) := d_t(x,y)+\pi_t(x,y).p_t^s(x,y)+[\pi_t(x,y)-1].p_t^f(x,y)-c_t(x,y)$. This ϱ is a well defined continuous map: $X_t \times Y_t \to \mathbb{R}$ since it is a sum and product of well defined and real-valued payoff functions that are continuous by the conditions of this proposition.

We now construct $\Gamma \subseteq Y_t$ as the set of possible signals at x, that is $\Gamma := \{y : f(x,y) > 0\}$. Further, we construct the subset $Z \subseteq \Gamma$ as the set of signals y for which $-\epsilon < \varrho(x,y) < 0$ and $f(x,y) > 0$. For any signal $y \in Z$, participation in task x would not be optimal under the original participation cost C_t, but participation would be optimal under the reduced participation cost $\tilde{C}_t$. Our goal is therefore to demonstrate that $x \times Z := \{(x \times z) : z \in Z\}$ has positive probability measure in $X_t \times Y_t$, because the agent's ex ante participation probability λ_t would increase by at least that measure if C_t were to be replaced by $\tilde{C}_t$.

We know by the conditions of the proposition that Γ is connected, since it is the support of the probability distribution of the signals Y_t for task x. Thus its image $\varrho(x \times \Gamma)$ is connected, since ϱ is continuous. but we also know that $\varrho(x \times \Gamma)$ contains both positive and negative elements since the participation condition is binding at x (Definition 1). Thus $\varrho(x \times \Gamma)$ is an interval on the real line that contains both positive and negative elements. We may therefore fix $-e$ as an arbitrary negative element of $\varrho(x \times \Gamma)$, and define $\epsilon' := \min(\epsilon, e)$. Then the open interval $(-\epsilon', 0)$ is contained in $\varrho(x \times \Gamma)$, and so its pre-image $\{(x,y) : x = x,\ \varrho(x,y) \in (-\epsilon', 0),\ f(x,y) > 0\}$ is a non-empty subset of Z, and it is also open, since it is the pre-image of an open set under a continuous map. But the probability measure of a non-empty open set with positive probability density is positive, and so we have that any finite decrease in participation cost induces a positive increase in participation likelihood under the conditions of the proposition.

The continuous case is demonstrated by a similar argument. First fix $t \leq T \in \mathbb{N}$, $\chi \in X_t$, and $\epsilon \in \mathbb{R}$ such that $\tilde{C}_t <_\epsilon C_t$ on χ and such that the participation condition (4.2) is binding on χ. Then construct $\Gamma \subseteq Y_t$ as the set of possible signals at χ, that is $\Gamma := \{y : x \in \chi,\ f(x,y) > 0\}$, and its subset Z as $\{y : x \in \chi,\ f(x,y) > 0,\ -\epsilon < \varrho(x,y) < 0\}$. Then $\chi \times \Gamma$ is connected by the conditions of the proposition, and its connected image under ϱ contains both positive and negative elements by Definition 1. As such, it contains an open subset between $-\epsilon$ and zero, the pre-image of which is a non-empty open subset $\{(x,y) : x \in$

$\chi,\ \varrho(x, y) \in (-\epsilon', 0),\ f(x, y) > 0\}$ which therefore has positive probability measure and which satisfies the participation condition under $\tilde{C}_t$ but not under C_t.

For the corollaries: Note that three period t participation outcomes are possible, dependent upon the realised task $\times$ signal pair (x, y). Either (x, y) would induce participation under both C_t and $\tilde{C}_t$, or under neither, or under only $\tilde{C}_t$. Only in the third case would the change to $\tilde{C}_t$ have any effect. In this case, additional human capital will be developed, with a present value of V_t, but we must consider whether this increase could be reversed by the future consequences of present participation. This rather perverse outcome would be possible in general, but not under either of the conditions stated in the proposition.

If tasks are perfect substitutes, then the agent's order of participation may be freely interchanged without altering the net effects of that participation. Thus additional participation in period t can never have a net negative effect in expectation. To see this consider the counterfactual agent 2 who does not participate in period t. Denote by τ the period in which counterfactual agent's total participation 'catches up' with that of agent 1. Then in period τ total participation is the same, only the order of participation differs in the counterfactual, and so future expected outcomes are identical. Thus the set of circumstances under which counterfactual expected participation 'overtakes' agent 1's participation is empty. However, the converse is not true. So long as the participation constraint remains binding, there remains some positive probability that the counterfactual agent would not participate in any future period.

If present-period participation weakly increases the probability of future-period participation, then the fact that $\tilde{C}_t$ strictly increases λ_t necessarily also means that $\tilde{C}_t$ strictly increases $\mathbf{E} \sum_{\tau=t}^{T} V_\tau.\lambda_\tau.\delta^{\tau-T} \mid \tilde{C}_t$

2 Proof of Propositions 2 and 3

The proof of Proposition 2 is almost entirely as above, so we will not repeat it here. The only difference is that we define $\epsilon' := \delta.\min(\epsilon, e)$, such that the pre-image of the interval $(-\epsilon', 0)$ contains only task $\times$ signal pairs such that an epsilon increase in drive or resilience would be sufficient, when multiplied the the probability of success (rsp. failure) to induce participation, even though that increase would be reduced by a factor of π_t (rsp. $(1 - \pi_t)$), since that factor is bounded away from zero by δ under the conditions of the proposition.

The proof of Proposition 3 follows exactly that of Proposition 2, since the effect of an increase in π_t is multiplied by the sum of p_t^s and p_t^f, and since that sum is bounded away from zero by δ under the conditions of the proposition.

B Ancillary Results

Our empirical model was set out in Section 4. That model requires the simultaneous estimation of several ancillary parameters and equations, and this appendix presents those results for the benchmark specification. These benchmark results are representative of those from each candidate operationalisation of noncognitive skill, and the full results for each of these can be found in the output files which we provide alongside our code in the supplementary materials. Our benchmark ancillary parameter estimates are provided in Tables 4.B1 and 4.B2.

Table 4.B1: *Ancillary parameter estimates for the benchmark specification*

	Cog skill prod.		Noncog skill prod.	
	Stage 1	Stage 2	Stage 1	Stage 2
Components of the error term $\eta_{i,t,k}$:				
residual investment $\gamma^k_{s,6}$	.001	.000	.012	.001
(unobserved heterogeneity)	(.012)	(.019)	(.012)	(.015)
variance of idiosyncratic shocks $\nu_{i,t,k}$	.242	.085	.193	.585
	(.003)	(.006)	(.002)	(.012)
AR(1) coefficient on family income	.904			
	(.002)			
Variance of shocks to family income	.024			
	(.000)			
Variance of shocks to residual investment	.277			
	(.005)			
Variance of shocks to parent noncog. skill	.130			
	(.001)			

NOTE: Standard errors in parentheses.

Table 4.B2: *Estimates of the benchmark investment equation coefficients*

	9 months	3 years	5 years	7 years	11 years	14 years
Cognitive Skill $\delta_{t,1}$	.117	.146	.160	-.013	-.031	-.027
	(.014)	(.021)	(.019)	(.019)	(.038)	(.065)
Noncognitive Skill $\delta_{t,2}$	.086	.110	.105	.233	.220	.189
	(.019)	(.024)	(.023)	(.025)	(.033)	(.041)
P. Cog. Skill $\delta_{t,3}$	.083	.205	.356	.323	.324	.343
	(.006)	(.018)	(.016)	(.020)	(.044)	(.067)
P. Noncog. Skill $\delta_{t,4}$	.081	.176	.216	.099	.120	.120
	(.008)	(.020)	(.018)	(.019)	(.035)	(.053)
Family Income $\delta_{t,5}$	.198	.486	.871	.954	1.002	1.071
	(.011)	(.033)	(.033)	(.040)	(.088)	(.134)

NOTE: Standard errors in parentheses.

C Measurement

The measurements used to operationalise each candidate noncognitive skill were detailed in Tables 4.1 – 4.4. This appendix sets out our measurements of: the benchmark proxy for child noncognitive skill, child cognitive skill, parental cognitive and noncognitive skills, and parental investments in Tables 4.C1 – 4.C4 respectively.

Table 4.C1: *Measurements of prosocial and antisocial behaviours*

Measurement	Reporter	Domain	Observations by data sweep				
			3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
–Combined: fights/bullies; argues; spiteful; temper; moody	Parent	$\approx\mathbb{R}$ $-IRT^*$	14,783	14,773	13,488		
–Combined: fights/bullies; lies; steals	Parent	$\{0, 1, 2, 3\}$				12,821	11,083
Combined: helpful; has ≥1 friend	Parent	$\{0, 1, ..., 8\}$ $-IRT^*$	14,783				
Combined: helpful; has ≥1 friend	Parent	$\{0, 1, 2\}$		14,773	13,489	12,822	11,150
–Combined: destructive; harmful to others	Visitor	$\{0, 1, 2\}$	13,795				
Social development score	Teacher	$\{0, 1, ..., 9\}$	11,855	(separate scale used in England)			
Combined: –steals; –lies; –temper; helps; ≥1 friend	Teacher	$\{0, 1, 2, 3, 4\}$			8,725		
Combined: –steals; –lies; –misbehaves; helps;	Teacher	$\{0, 1, 2, 3, 4\}$				7,211	
How often behave well in class	Self	$\{0, 1, 2\}$			12,760		
–How often misbehave in class	Self	$\{0, 1, 2\}$				12,813	
–Combined: raucous; shoplift; graffiti; vandalism	Self	$\{0, 1, 2, 3\}$				12,682	
–Combined: raucous; shoplift; graffiti; vandalism	Self	$\{0, 1, ..., 7\}$					10,931
–Combined: bullies: siblings; other children; online	Self	$\{0, 1, 2, 3, 4, 5\}$					11,192
–Combined: last 12 months: fought; used weapon; stolen	Self	$\{0, 1, 2\}$					11,192

NOTE: Noncognitive skill measurements in sweep 1 (age 9 months) are pleasantness, wariness, and fussiness.
* Combined using Item Response Theory, see Section 1.

Table 4.C2: *Measurements of child cognitive skill level*

Measurement	Reporter	Domain	Observations by data sweep					
			9 months	3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
Combined: waves bye; extends arms; nods yes; hands together; gives toy	Parent	$\approx\mathbb{R}\ -IRT^{*}$	16,675					
Combined: sits up; stands (holding on); moves around; walks a few steps	Parent	$\approx\mathbb{R}\ -PFA^{\dagger}$	16,675					
Combined: smiles; grabs; holds; passes	Parent	$\approx\mathbb{R}\ -IRT^{*}$	16,675					
Birthweight (cleaned with admin. data)	Parent	$\approx\mathbb{R}$	17,210					
Combined: speech concerns; –needs help for tasks; speech understandable; –various motion problems	Parent	$\approx\mathbb{R}\ -PFA^{\dagger}$		14,571				
Bracken school readiness	Visitor	$\approx\mathbb{R}$		13,790				
Vocab test	Visitor	$\approx\mathbb{R}$		14,514	14,961	13,412	12,994	10,781
Picture similarlity test	Visitor	$\approx\mathbb{R}$			14,951			
Patterns test	Visitor	$\approx\mathbb{R}$			14,904	13,552		
Maths test	Visitor	$\approx\mathbb{R}$				13,574		
School subject difficulties	Parent	$\{0,1,2,3,4\}$				13,733		
Teacher assessment in core subject areas	Teacher	$\approx\mathbb{R}\ -IRT^{*}$				11,854	8,611	7,280
Working memory task errors	Visitor	$\approx\mathbb{R}$					12,589	
Decision test	Visitor	$\approx\mathbb{R}$					12,552	10,708

NOTE: †Combined using Principal Factor Analysis; *Combined using Item Response Theory, see Section 4.1.

Table 4.C3: *Measurements of parental skills*

Measurement	Reporter	Domain	Observations by data sweep					
			9 months	3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
Parental Cognitive Skills (assumed constant)								
Highest qualification level (NVQ scale)	Parent	$\{0, 1, 2, 3, 4, 5\}$	16,995					
Years of schooling	Parent	$\{0, ..., 8\}$	17,270					
–Difficulties: form-filling; arithmetic; reading	Parent	$\{0, 1, 2, 3\}$	17,280					
Vocab test	Parent	$\{0, 1, ..., 20\}$						11,057
Parental noncognitive skills								
Combined: learning empowers; education improves parenting; –computer use too much effort	Parent	$\approx\mathbb{R}\ -IRT^*$	16,144					
–Combined: feel: useless; no good; a failure	Parent	$\approx\mathbb{R}\ -IRT^*$	16,138					
Combined: have control; get what want; run own life	Parent	$\approx\mathbb{R}\ -IRT^*$	16,139					
–Combined: miserable; worried; scared; annoyed; upset; jittery; enraged; heart races	Parent	$\approx\mathbb{R}\ -IRT^*$	16,155					
–Combined: Tobacco use; alchohol use; also drug use in sweeps 3 and 6	Parent	$\{0, 1, 2, 3\}^\dagger$	17,311	15,385	15,166	13,778	13,178	11,578
How satisfied with life	Parent	$\{0, 1, 2, ..., 9\}^\ddagger$	16,131	13,462	14,160	13,024	12,252	10,859
–Kessler psych distress: depressed; hopeless; restless; effortful; nervous; worthless	Parent	$\approx\mathbb{R}\ -IRT^*$		13,575	14,358	13,199	12,334	10,881

NOTE: †\{0,1,2,3,4\} in sweeps 5 and 6. ‡\{0,1,...,10\} in sweep 5. *Combined using Item Response Theory, see Section 4.1.

Table 4.C4: *Measurements of parental investments*

Measurement	Reporter	Domain	Observations by data sweep					
			9 months	3 yrs	5 yrs	7 yrs	11 yrs	14 yrs
Frequency plays with friends	Parent	$\{0, 1, 2, 3, 4\}$	16,670					
Combined: value of: Stimulating child; talking to child; cuddling child	Parent	$\{0, 1, 2, 3, 4\}$	16,152					
Combined: access to: heating; phone; web; car; garden	Parent	$\{0, 1, ..., 5\}$	17,311					
Patience with child	Parent	$\{0, 1, 1.5, 2, 3\}$	16,127					
Combined: Observes: toys; encouragement; caressing; conversing; praise; parent introduces visitor	Visitor	$\approx\mathbb{R}\ -IRT^*$		14,176				
Pianta relationship: affectionate; comforting; valued; in-tune; sharing; confiding; –moody	Parent	$\approx\mathbb{R}\ -IRT^*$		13,750				
Together: read; count; alphabet; songs; sport; draw; reads with others	Parent	$\approx\mathbb{R}\ -IRT^*$		15,346				
Together: sports; stories; music; arts/crafts; games; exercise; park; read; active	Parent	$\approx\mathbb{R}\ -IRT^*$			15,172	13,782		
Together: reading; writing; maths	Parent	$\approx\mathbb{R}\ -IRT^*$			14,979	13,760		
Visits: Library; sports classes; cultural; historical; animal; funfair; cinema; sports events	Parent	$\approx\mathbb{R}\ -IRT^*$			15,176	13,788		
How often have weekend fun with the family	Self	$\{0, 1, 2\}$				12,822		
Combined: help with hwork; enforce hwork	Parent	$\approx\mathbb{R}\ -IRT^*$					13,145	
Excercise: with family; at club or class	Parent	$\approx\mathbb{R}\ -IRT^*$					13,179	
I wish family could afford to buy me more	Self	$\{0, 1, 2, 3\}$					12,734	
Together: how often: games; exercise; library; talk about things that matter	Parent	$\approx\mathbb{R}\ -IRT^*$					13,180	
Extra tuition	Parent	$\{0, 1\}$					13,161	11,526
How often talk about things that matter	Parent	$\{0, 1, 2, 3\}$						11,575

NOTE: *Combined using Item Response Theory, see Section 4.1.

Chapter References

Agostinelli, Francesco & Matthew Wiswall (2016b). “Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development”. In: NBER Working paper, No. 22441. Cambridge.

Akerlof, George A & Rachel E Kranton (2002). “Identity and schooling: Some Lessons from the Economics of Education”. In: *Journal of Economic Literature* 40.4, pp. 1167–1201.

Almås, Ingvild et al. (2016). “What explains the gender gap in college track dropout? Experimental and administrative evidence”. In: *American Economic Review* 106.5, pp. 296–302.

Attanasio, Orazio et al. (2015). “Estimating the production function for human capital: Results from a randomized control trial in Colombia”. In: National Bureau of Economic Research working papers, no. 20965.

Becker, Anke et al. (2012). “The Relationship between Economic Preferences and Psychological Personality Measures”. In: *Annual Review of Economics* 4, pp. 453–478.

Becker, Gary S (1962). “Investment in Human Capital: A Theoretical Analysis”. In: *Journal of Political Economy* 70.5, pp. 9–49.

— (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: National Bureau of Economic Research.

Belzil, Christian (2007). “Subjective beliefs and schooling decisions”. In: GATE Working Papers, No. 2007-17 07-17. Écully: Groupe d’Analyse et de Théorie Économique.

Bénabou, Roland & Jean Tirole (2002). “Self-Confidence and Personal Motivation”. In: *The Quarterly Journal of Economics* 117.3, pp. 871–915.

Bénabou, Roland & Jean Tiróle (2016). “Mindful Economics: The Production, Consumption, and Value of Beliefs”. In: *Journal of Economic Perspectives* 30.3, pp. 141–164.

Brañas-Garza, Pablo, Praveen Kujal & Balint Lenkei (2019). “Cognitive reflection test: Whom, how, when”. In: *Journal of Behavioral and Experimental Economics* 82.March, p. 101455.

Burks, Stephen V et al. (2015). “Cognitive skills, personality, and economic preferences in collegiate success”. In: *Journal of Economic Behavior and Organization* 115, pp. 30–44.

Casselman, Brock L. & Charles H. Atwood (2017). “Improving General Chemistry Course Performance through Online Homework-Based Metacognitive Training”. In: *Journal of Chemical Education*, acs.jchemed.7b00298.

Coneus, Katja, Manfred Laucht & Karsten Reuß (2012). “The role of parental investments for cognitive and noncognitive skill formation - Evidence for the first 11 years of life”. In: *Economics and Human Biology* 10.2, pp. 189–209.

Credé, Marcus, Michael C Tynan & Peter D Harms (2017). “Much Ado About Grit: A Meta-Analytic Synthesis of the Grit Literature”. In: *Journal of Personality and Social Psychology* 113.3, pp. 492–511.

Cunha, Flávio & James J Heckman (2007). “The Technology of Skill Formation”. In: *The American Economic Review* 97.2, pp. 31–47.

— (2008). “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Journal of Human Resources* 43.4, pp. 738–782.

— (2010). “Investing in our Young People”. In: IZA discussion paper, No.5050. Bonn: Forschungsinstitut zur Zukunft der Arbeit.

Cunha, Flávio, James J Heckman & Susanne M Schennach (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Econometrica* 78.3, pp. 883–931.

Del Boca, Daniela, Chiara Monfardini & Cheti Nicoletti (2017). “Parental and Child Time Investments and the Cognitive Development of Adolescents”. In: *Journal of Labor Economics* 35.2, pp. 565–608.

Dignath, Charlotte & Gerhard Büttner (2008). “Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level”. In: *Metacognition and Learning* 3.3, pp. 231–264.

Dohmen, Thomas et al. (2011). “Individual risk attitudes: Measurement, determinants, and behavioral consequences”. In: *Journal of the European Economic Association* 9.3, pp. 522–550.

Duckworth, Angela L et al. (2007). “Grit: Perseverance and passion for long-term goals.” In: *Journal of Personality and Social Psychology* 92.6, pp. 1087–1101.

Embretson, S E & S P Reise (2000). *Item Response Theory for Psychologists*. Mahwah: Lawrence Erlbaum Associates.

Embrey, Iain Peter (2019a). “A Review of Economic Dual-Self Modelling”. Lancaster.

— (2019c). “On the Benefits of Normalization in Production Functions”. Lancaster.

— (2019d). “Re-Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Under Review*.

— (2019e). “States of nature and states of mind: a generalized theory of decision-making”. In: *Theory and Decision*.

Embrey, Iain Peter (2019f). “The Education Trap: Could a grades-focussed educational system be perpetuating poverty in advanced economies?” In: *Under Review.*

Filippin, Antonio & Marco Paccagnella (2012). “Family background, self-confidence and economic outcomes”. In: *Economics of Education Review* 31.5, pp. 824–834.

Fiorini, Mario & Michael P Keane (2014). “How the Allocation of Children’s Time Affects Cognitive and Noncognitive Development”. In: *Journal of Labor Economics* 32.4, pp. 787–836.

Frederick, Shane (2005). “Cognitive Reflection and Decision Making”. In: *Journal of Economic Perspectives* 19.4, pp. 25–42.

Heckman, James J, John Eric Humphries & Gregory Veramendi (2018). “Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking”. In: *Journal of Political Economy* 126.S1, S197–S246.

Heckman, James J & Stefano Mosso (2014). “The Economics of Human Development and Social Mobility”. In: *Ann. Rev. Econ.* 6.2006, pp. 689–733.

Heckman, James J, Rodrigo Pinto & Peter Savelyev (2013). “Understanding the Mechanisms Through which an Influential Early Childhood Program Boosted Adult Outcomes”. In: *American Economic Review* 103.6, pp. 2052–2086.

Heckman, James J & Y Rubinstein (2001). “The importace of noncognitive skills: lessons from the GED testing”. In: *the American Economic Review* 92.2, pp. 145–149.

Heineck, Guido & Silke Anger (2010). “The returns to cognitive abilities and personality traits in Germany”. In: *Labour Economics* 17.3, pp. 535–546.

Helmers, Christian & Manasa Patnam (2011). “The formation and evolution of childhood skill acquisition: Evidence from India”. In: *Journal of Development Economics* 95.2, pp. 252–266.

Hernández-Alava, Mónica & Gurleen Popli (2017). “Children’s Development and Parental Input: Evidence From the UK Millennium Cohort Study”. In: *Demography* 54.2, pp. 485–511.

Humphries, John Eric & Fabian Kosse (2017). “On the interpretation of noncognitive skills - What is being measured and why it matters”. In: *Journal of Economic Behavior and Organization* 136, pp. 174–185.

Koch, Alexander, Julia Nafziger & Helena Skyt Nielsen (2015). “Behavioral Economics of Education”. In: *Journal of Economic Behavior & Organization* 115.1, pp. 3–17.

Köszegi, Botond (2006). “Ego Utility , Overconfidence , and Task Choice”. In: *Journal of the European Economic Association* 4.4, pp. 673–707.

Lavecchia, Adam M, Heidi Liu & Philip Oreopoulos (2016). “Behavioral Economics of Education: Progress and Possibilities”. In: *The Handbook of the Economics of Education,* ed. by Eric A Hanushek, Stephen Machin & Ludger Woessmann. Vol. 5. Amsterdam: North-Holland. Chap. 1, pp. 1–74.

MacLeod, W. Bentley (2016). “Human capital: Linking behavior to rational choice via dual process theory”. In: *Labour Economics* 41, pp. 20–31.

McCrae, Robert R & Paul T Jr Costa (1985). “Updating Norman' s “ Adequate Taxonomy ”: Intelligence and Personality Dimensions in Natural Language and Questionnaires”. In: *Journal of Personality and Social Psychology* 49.3, pp. 710–721.

Moffitt, T. E. et al. (2011). “A gradient of childhood self-control predicts health, wealth, and public safety”. In: *Proceedings of the National Academy of Sciences* 108.7, pp. 2693–2698.

Mueller, Claudia M & Carol S Dweck (1998). “Praise for Intelligence Can Undermine Children's Motivation and Performance”. In: *Journal of Personality and Social Psychology* 75.1, pp. 33–52.

OECD (2017). *Starting Strong 2017: Key OECD Indicators on Early Childhood Education and Care*. Paris: OECD Publishing, p. 192.

Pennycook, Gordon, Jonathan A. Fugelsang & Derek J. Koehler (2015). “Everyday Consequences of Analytic Thinking”. In: *Current Directions in Psychological Science* 24.6, pp. 425–432.

Pennycook, Gordon & Robert M. Ross (2016). “Commentary: Cognitive reflection vs. calculation in decision making”. In: *Frontiers in Psychology* 7.JAN, pp. 2014–2017.

Pennycook, Gordon et al. (2016). “Is the cognitive reflection test a measure of both reflection and intuition?” In: *Behavior Research Methods* 48.1, pp. 341–348.

Reuß, Karsten (2011). “Determinants of personality and skill development in the Socio-emotional environment during childhood”. In: MPRA Paper No. 82818. Munich.

Rustichini, Aldo et al. (2016). “Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation”. In: *Journal of Behavioral and Experimental Economics* 64, pp. 122–137.

Segal, Carmit (2013). “Misbehavior, Education, and Labor Market Outcomes”. In: *Journal of the European Economic Association* 11.4, pp. 743–779.

Sjögren, Anna & Susanna Sällström (2004). “Trapped, Delayed and Handicapped”. In: *CEPR Discussion Paper Series*. CEPR Discussion Paper Series, No. 613. London: Centre for Economic Policy Research.

Spence, Michael (1973). “Job Market Signaling”. In: *The Quarterly Journal of Economics* 87.3, p. 355.

Tanner, Kimberly D. (2012). “Promoting student metacognition.” In: *CBE life sciences education* 11.2, pp. 113–120.
Thiel, Hendrik & Stephan L. Thomsen (2013). “Noncognitive skills in economics: Models, measurement, and empirical evidence”. In: *Research in Economics* 67.2, pp. 189–214.
Wang, X. Henry & Bill Z. Yang (2003). “Why Competition may Discourage Students from Learning? A Behavioral Economic Analysis”. In: *Education Economics* 11.2, pp. 117–128.
Zepeda, Cristina D. et al. (2015). “Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study”. In: *Journal of Educational Psychology* 107.4, pp. 954–970.

Chapter 5: **An Holistic Conclusion**

1 The Present Contribution

This thesis has presented theoretical and empirical analyses of the pathways that lead to poverty. Although poverty is characterised by a complex cluster of negative socio-economic outcomes, it is now well-established that the pathways which lead into poverty are mediated through the effects of educational development in childhood. Education transforms pre-existing characteristics into the cognitive and noncognitive skills that determine adult outcomes, and educational engagement decisions have a causal effect upon that transformation (Sparkes & Glennerster 2002; Hobcraft 2002; Cahuc, Zylberberg & Carcillo 2014).[1] Accordingly, the present analyses have aimed to advance our understanding of the divergent developmental pathways that solidify as children progress through their compulsory education in advanced economies. This chapter reviews my progress toward that aim, and it outlines the broader implications of my results.

[1] These are indirect and direct mediation effects respectively – see Heckman & Pinto (2015) for a rigorous discussion.

In Chapter 2 I presented a theoretical explanation for divergence in developmental pathways. That explanation arises because childhood development is modelled as the cumulative consequence of (non-)participation in a series of incremental educational opportunities. Initially, those opportunities will be provided by parents, but formal educational provision will take an increasingly important role as the child develops, and so this model is able to encapsulate both the direct and the indirect mediating effects of education in the production of adult outcomes. The direct effect of education is manifest in the quality and quantity of the educational opportunities that it supplies, net of any endogenous response by parents to that provision. The indirect effect of education is the variation in eventual skill levels that can be attributed to skill differences at enrolment: my results show that these initial differences will be attenuated over time for two sufficiently similar agents, but that they will be greatly exacerbated for two agents who begin education above and below their respective threshold ability levels. This is the Education Trap: children who have sufficient early disadvantage will endogenously separate onto a divergent low-investment pathway at equilibrium.

This model has several important implications for policy and for society. First, the endogenous separation of children with less than a critical threshold level of ability provides the first viable theoretical explanation for very low educational investment amongst disadvantaged children. Second, this mechanism also contributes a rigorous explanation for the well-known result that early interventions in cognitive development are the most effective (see, for example Francesconi & Heckman 2016): because the low-investment equilibrium rapidly diverges away from the critical threshold ability level, later interventions will require an increasingly large impact to propel their recipient onto a high-investment pathway. Third, these conclusions mandate a paradigm shift in intervention policy such that its target should become the educational system rather than the child. To this end I have derived concrete recommendations that might move the current grades-focussed system toward a learning-focussed system that could not sustain a separating equilibrium. Finally, because our model shows that the apparent educational underinvestment of disadvantaged children may be an equilibrium response to an exogenous endowment of early disadvantage, we must conclude that the victims of poverty may not be wholly responsible for their situation.

Chapter 3 is primarily econometric in nature. Its contribution is therefore to build upon the seminal work of Cunha, Heckman & Schennach (2010), and to make the improved estimation procedure available for future research. My most important improvements to the estimation strategy were to replace its misleading anchoring procedure with explicit CES normalisation and to correct over-identifying restrictions within the factor-analytic measurement model, although I also made several more minor amendments to the code.[2] Together, these improvements uncover substantial and statistically significant anomalies within the main results of Cunha, Heckman & Schennach (2010), and so this chapter also makes an independent contribution to our understanding of the technologies of cognitive and noncognitive skill production. In particular, the evidence for cross-productivity between cognitive and noncognitive skills is substantially weaker than previously believed, whilst the already-compelling evidence for early investment in cognitive skills is now substantially stronger than previously believed.

Chapter 4 builds upon each of the preceding chapters. Its theoretical contribution is to derive five candidate noncognitive skills from the model presented in Chapter 2, and its empirical contribution is to test the impact of each of those noncognitive skills upon cognitive development during childhood. I find strong statistical evidence that each candidate noncognitive skill positively affects ongoing cognitive development, and I also present estimates of the technologies by which each of those noncognitive skills is produced. These results therefore advance our current knowledge in the area by suggesting which specific noncognitive skills are likely to provide the most effective intervention opportunities at each stage of childhood development. In addition, these results are highly informative for intervention design because each specific noncognitive skill is precisely defined as a parameter within a decision-theoretic model. These precise definitions contrast with many existing constructs of noncognitive skill, which are typically only defined by linguistic associations.

In sum, the results from Chapters 2, 3 and 4 shed new light upon the pathways that lead young people into poverty. I have provided a novel and yet intuitively appealing mechanistic model of educational investment, and I have found strong empirical support for that model. Moreover, because that model is based upon tangible investment decisions, its payoffs are directly interpretable – both as aspects of those investment opportunities and as aspects of an individual's noncognitive skill set. These interpretations are dual: for example the payoff that a given individual expects to derive from perceived success in a given task depends both upon the characteristics of the task and upon their own individual characteristics. In principle, policy-makers and practitioners could intervene to manipulate either set of characteristics, although to date interventions overwhelmingly target the individual (see reviews in Almlund et al. (2011) and Lavecchia, Liu & Oreopoulos (2016)). Current practice notwithstanding, a priori it would seem likely that interventions which target the educational system could be both more effective and less costly than individual-level intervention, and this suggests that such systemic interventions could represent a valuable and underexploited set of policy leavers

[2] A rigorous discussion of CES normalisation can be found in Embrey (2019c).

with the potential to reduce the incidence of poverty in advanced economies.

2 The Wider Implications of this Contribution

The present contribution suggests a number of implications for future research and for public policy. An immediate implication is that future work might follow the specific recommendations of Chapter 2 to develop and pilot educational systems in which engagement becomes attractive to disadvantaged children. In addition, where traditional individual-level interventions remain in use, an important aspect of those interventions should become the explicit teaching and implicit modelling of analytic decision-making, because the results presented in Chapter 4 have shown that an individual's propensity to think analytically is an important determinant of their ongoing educational development. However, these results also raise the possibility that an individual's propensity to think analytically might be an important causal determinant across the wider domain of multi-dimensional poverty. I therefore conclude this thesis by discussing that possibility.

My model of education as a repeated participation game was explicitly presented as a specialisation of a generally applicable modelling approach. That approach recognises that people in their everyday lives might not always act according to the complex high-level planning decisions of normative microeconomic theory, and that instead some outcomes for some people could arise as the cumulative consequence of a series of elemental participation decisions. In these cases, elemental participation decisions could be seen as the nano-foundations of traditional microeconomic investment decisions, in analogue to the manner in which individual objective maximisation provides the micro-foundations for macroeconomic decision outcomes. Moreover, because this thesis has demonstrated that the nano-founded approach can provide both explanatory power and novel insight into educational decision-making, it seems possible that it may also provide new insight within other microeconomic decision contexts.

Under the hypothesis that a single, aggregated, investment decision is made to determine important life outcomes, it may appear beyond doubt that the decision-maker will take due account of the long-term consequences of that decision. However, if such a high-level decision is never taken, if instead its outcome is arrived at as the cumulative consequence of a series of incremental participation decisions, then that commonly maintained assumption is no-longer beyond question. In Figure 4.2 of Chapter 4 I presented a generalisation of expected utility theory that admits the possibility that some elemental participation decisions may be taken without any consideration of their long-term consequences. That decision theory nests Expected Utility Theory at the point where all individuals display a uniformly perfect propensity to think analytically. However, the results of Chapter 4 provide strong empirical support for meaningful heterogeneity in individuals' propensity to think analytically, in addition to the support that they provide for the nano-founded model more broadly.

It is nevertheless possible that the results presented in this thesis may not be externally valid beyond the educational investment decisions of children. In particular, the generalised decision theory may only be pertinent for those childhood decisions because children are particularly likely to act without considering the consequences of their actions – indeed action without consideration for its consequences is almost synonymous with childish behaviour. Future research should therefore test the nano-founded approach and the generalised decision theory within other decision contexts, for example by applying the methodology of chapter 3 to establish whether the an individual's propensity to think analytically has a substantial and statistically significant effect upon their production of additional socio-economic outcomes.

On an intuitive level it is plausible that adults may also, on occasion, act without considering the consequences of their actions. This idea is not new: in 1759 Adam Smith wrote that many people act "variously and accidentally, depending on whether mood, inclination, or self-interest happens to be uppermost" (p.276), and in around 380 B.C.E. Plato wrote "You may observe that in children: from their earliest years they are full of spirit, but some of them seem to me never to acquire reason, and most of them do so only quite late" (p.130). Moreover, it seems likely that people living with the debilitating urgency of poverty are less likely than most to deliberate analytically upon each elemental participation decision, and so impoverished outcomes in one decision domain could cause impoverished decision-making in others. Even in the absence of such causation, it is likely that an individual's propensity to think analytically in one decision domain will be correlated with their propensity to do so in other decision domains, and so if some individuals do indeed follow the incremental approach to microeconomic decision-making, and if those incremental decisions are made according to the generalised decision theory, then we should expect to see a clustering of negative socio-economic outcomes.

The results presented in this thesis provide both theoretical and empirical support for a nano-founded and generalised decision theory within the domain of childhood educational investment decisions. If future work were to corroborate these findings in additional decision domains, then the implications of these innovations could be far-reaching. At the time of writing, there is some indicative support for the generalised decision theory within the wider literature. In Embrey (2019e) I sketch theoretical applications of the generalised decision theory that provide new insights into chronic unemployment and into economic behaviours such as the crowding-out effect. In Embrey (2019a) I pursue a complementary avenue of research by providing a thorough discussion of the foundational assumptions of the generalised decision theory, and by placing that theory within the existing dual-self literature in economics. In addition, there is authoritative psychological work which proposes that the Cognitive Reflection Test of Frederick (2005) provides a meaningful measure of an individual's propensity to think analytically (Pennycook et al. 2016; Pennycook & Ross 2016), and there is a large correlational literature that documents associations between performance in the Cognitive Reflection Test and outcomes across multiple decision domains (for reviews see Pennycook, Fugelsang & Koehler 2015; Brañas-Garza, Kujal & Lenkei 2019). Initial indications therefore suggest that nano-founded and generalised decision theories might indeed provide important insights into the pathways that lead to poverty, although considerably more work is required to formalise their implications for additional dimensions of poverty, and to test for causal evidence of those implications.

Chapter References

Almlund, Mathilde et al. (2011). "Personality Psychology and Economics". In: *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen Machin & Ludger Woessmann. 1st ed. Vol. Volume 4. Amsterdam: North-Holland. Chap. 1, pp. 1–181.

Brañas-Garza, Pablo, Praveen Kujal & Balint Lenkei (2019). "Cognitive reflection test: Whom, how, when". In: *Journal of Behavioral and Experimental Economics* 82.March, p. 101455.

Cahuc, Pierre, André Zylberberg & Stéphane Carcillo (2014). *Labour Economics*. 2nd. Cambridge: MIT press.

Cunha, Flávio, James J Heckman & Susanne M Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Econometrica* 78.3, pp. 883–931.

Embrey, Iain Peter (2019a). "A Review of Economic Dual-Self Modelling". Lancaster.

— (2019c). "On the Benefits of Normalization in Production Functions". Lancaster.

— (2019e). "States of nature and states of mind: a generalized theory of decision-making". In: *Theory and Decision*.

Francesconi, Marco & James J Heckman (2016). "Child Development and Parental Investment: Introduction". In: *Economic Journal* 126.596, F1–F27.

Frederick, Shane (2005). "Cognitive Reflection and Decision Making". In: *Journal of Economic Perspectives* 19.4, pp. 25–42.

Heckman, James J & Rodrigo Pinto (2015). "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs". In: *Econometric Reviews* 34, pp. 6–31.

Hobcraft, John (2002). "Social Exclusion and the Generations". In: *Understanding Social Exclusion*. Ed. by J. Hills, J. Le Grand & D. Piachaud. Oxford: Oxford University Press. Chap. 5, pp. 62–83.

Lavecchia, Adam M, Heidi Liu & Philip Oreopoulos (2016). "Behavioral Economics of Education: Progress and Possibilities". In: *The Handbook of the Economics of Education,* ed. by Eric A Hanushek, Stephen Machin & Ludger Woessmann. Vol. 5. Amsterdam: North-Holland. Chap. 1, pp. 1–74.

Pennycook, Gordon, Jonathan A. Fugelsang & Derek J. Koehler (2015). "Everyday Consequences of Analytic Thinking". In: *Current Directions in Psychological Science* 24.6, pp. 425–432.

Pennycook, Gordon & Robert M. Ross (2016). "Commentary: Cognitive reflection vs. calculation in decision making". In: *Frontiers in Psychology* 7.JAN, pp. 2014–2017.

Pennycook, Gordon et al. (2016). "Is the cognitive reflection test a measure of both reflection and intuition?" In: *Behavior Research Methods* 48.1, pp. 341–348.

Plato (c.380 B.C.E.). "The Republic". In: *The Republic of Plato, 1906*. Ed. by A.D. Lindsay & R. Bambrough. London: J.M. Dent & Sons Ltd.

Smith, Adam (1759). *The Theory of Moral Sentiments*. London: A. Millar.

Sparkes, Jo & Howard Glennerster (2002). "Preventing Social Exclusion: Education's Contribution". In: *Understanding Social Exclusion*. Ed. by J. Hills, J. Le Grand & D. Piachaud. Oxford: Oxford University Press. Chap. 11, pp. 178–201.

Appendix: Supplementary materials to the main chapters

1 Introduction to the Supplement

I have compiled supplementary materials in support of each of the main chapters, and these will be published alongside those chapters in due course. Some of these materials cannot feasibly be reproduced here, but should any interested party wish to be supplied with these then please contact the author via i.embrey@lancs.ac.uk. The supplementary materials that are available upon request comprise:

- For all chapters the code that generates my results (variously in Matlab, Stata, and Fortran);
- For Chapter 3 the final dataset;
- For Chapter 4 a spreadsheet detailing the structure and derivation of the final dataset (in addition to the code that compiles that dataset). Instructions on how to obtain access to the Millenium Cohort Study data for Chapter 4 are included in Section 1 of that chapter.

However there are also some supplementary materials that can be appended to this thesis. These include robustness exercises pertaining to Chapter 2, a guide to the code pertaining to Chapter 3, and a commentary on the data compilation for Chapter 4. These are now set out according to the below table of contents.

Appendix Contents

2 Supplement to Chapter 2

1 Further discussion of Figure 2.3

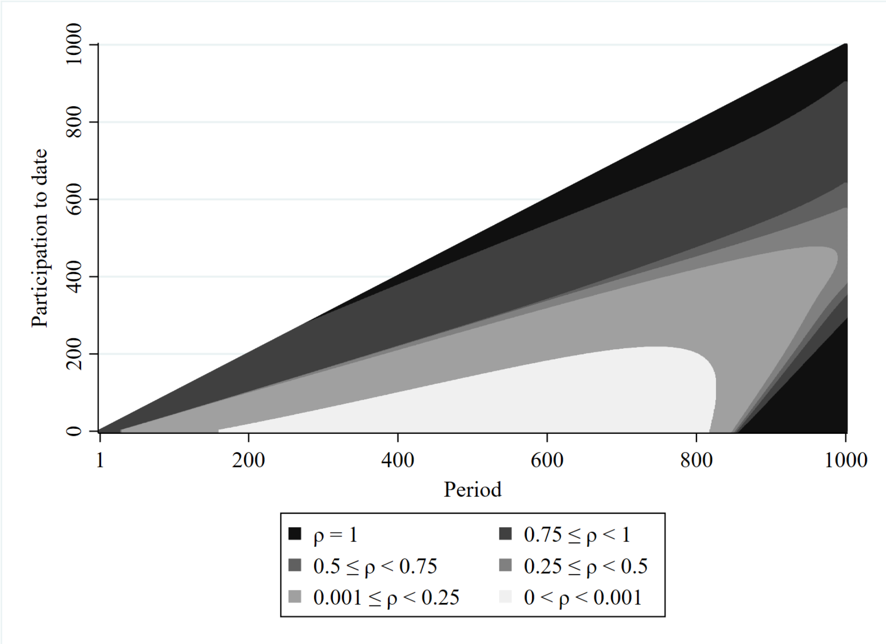
Figure 2.3 of the main text is reproduced below. This figure shows the equilibrium strategy for our preferred specification when agents possess private information as to their realised probability of achieving perceived success in each period. In doing so, this figure captures much of the numerical outcome from our preferred specification: note for example that we can read off the benchmark solution as a dichotomisation around $\rho = 0.5$. For this reason we will carry out our robustness checks by computing the analogue of Figure A.1 for each alternative specification.

Before undertaking the robustness exercises it is worth taking a moment to analyse the implications of Figure A.1 in greater detail. For example, we can see that the participation probability is always positive. This is because $p^s > c$, and so there is always some chance that a sufficiently positive realisation of π_t would lead to a positive expected payoff. Nevertheless, the probability of such a positive draw being realised may be negligible – the lightest shading here indicates $\rho < 0.001$.

We can also observe that Figure A.1 extends our analytic results by demonstrating that the optimal quantity of postponed participation will be decreasing in an agent's prior educational level. This is because we have assumed diminishing life-cycle returns to education ($V'' < 0$), and so an agent with greater prior ability will have less incentive to invest in her education, until and unless she crosses the critical ability threshold for the high-participation equilibrium.

Finally, Figure A.1 shows an interesting upward curve at end of the high-participation region. This is because, as the number of future periods decreases, the case for participation as a means to increase future expected psychic payoffs becomes less overwhelming.

Figure A.1: *The Equilibrium Strategy with Private Information*

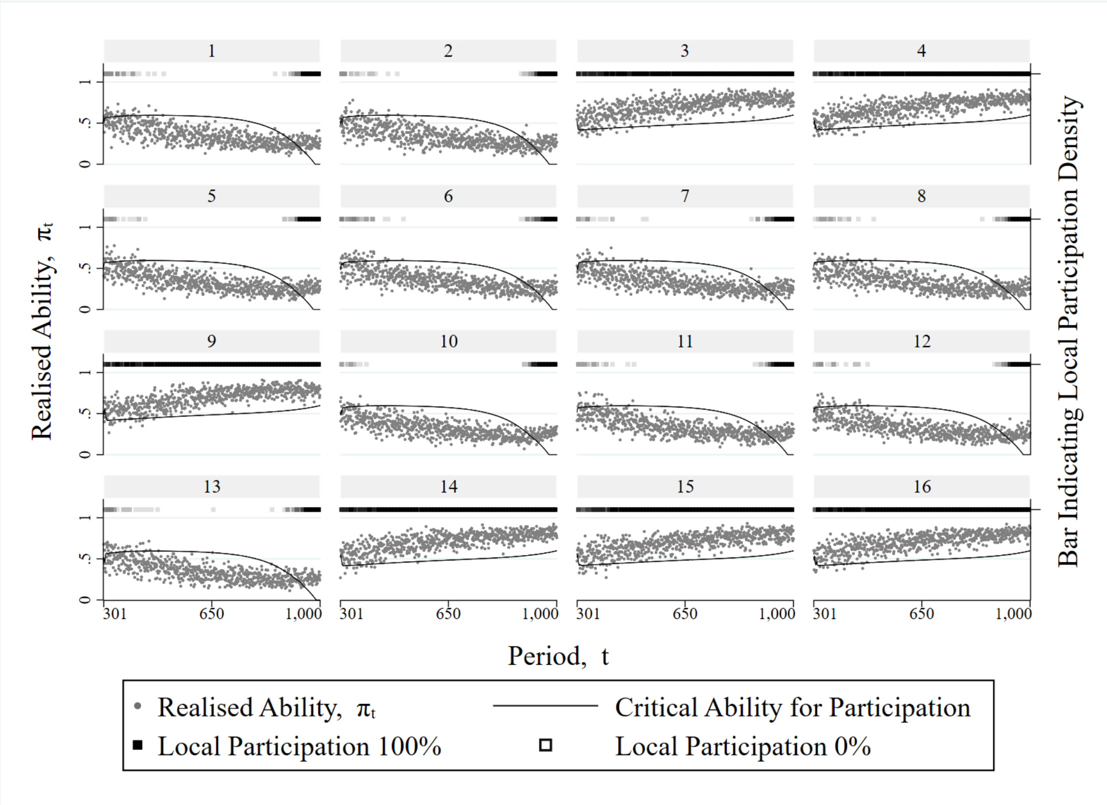


The equilibrium strategy S^* for the specification detailed in Table A.1, when participation may be conditioned upon the realisation $\pi_t(n)$. Events are shaded according to the probability $\rho_t(n)$ that the agent's realised ability $\pi_t(n)$ would be large enough to induce equilibrium participation.

2 Additional simulations cf. Figure 2.4

As stated in the text these (and other) simulations suggest that the individual developmental pathways presented in Figure 4 are representative.

Figure A.2: *Simulated Ability Development for 16 additional Agents*



Simulated relative ability development for 9 representative agents, all of whom are endowed with $n = 157$ in period $t = 301$. Variation occurs because each agent receives a private signal as to her realised probability of success in each period. The bar above each panel is shaded to indicate the local participation density.

3 Introduction to the robustness exercises

The preferred specification detailed in Table 2.1 of the main chapter is reproduced here. This section provides the rationale for that specification, which is intended to reflect reality as closely as possible whilst remaining tractable. This supplement demonstrates that the numerical results presented in the paper are remarkably robust to alternative specifications.

Table A.1: *The parametric assumptions of Table 2.1*

Parameter	Assumption	Notes
Number of periods	$T = 1,000$	Robust to parameter variation.
Initial ability distribution	$\Pi_1 \sim Beta[2.5, 2.5]$	As used by Filippin & Paccagnella 2012; Robust to truncated normal; Robust to parameter variation.
Π_t update magnitude	$\iota = 0.05$	Robust to parameter variation.
Value of education	$V(n) = V(T)\left[1-\left(\frac{99}{100}\right)^n\right]$	Robust to parameter variation; Robust to linear form.
Maximum participation benefit	$V(T) = 10,000$	Robust to parameter variation.
Psychic payoffs	$p^s = p^f = 5$	Robust to parameter variation; Robust to asymmetric values.
Participation cost	$c = 1$	Robust to parameter variation.
Discount rate	$\delta = 0.99$	Reasonable cf. † and ‡; Robust to parameter variation.

† Benhabib, Bisin & Schotter (2010); ‡ Frederick, Loewenstein & O'Donoghue (2002).

4 Rationale for the preferred specification

The first main functional form assumption of Table A.1 is that of the anticipated benefit $V(n)$ of obtaining any given education level n. Commensurate with the intuition that education should add value, but at a diminishing marginal rate, my preferred specification is for $V(n)$ to approach its limiting value $V(T)$ exponentially - here at the rate of one hundredth of the remaining distance each period. Over 1,000 periods, those incremental developments would sum to 99.996% of $V(T)$.

The second main functional form assumption determines the effect of each participation decision on agents' skill development, as parameterised by their probability of success distribution $\Pi_t(n)$. Since $\Pi_t(n)$ is a distribution of probabilities, it is natural to model it as Beta distributed. The Beta-distribution is a bell-shaped curve on the support $[0, 1]$, which also has several other desirable properties: it is very convenient to update the Beta distribution in response to participation or avoidance since its conjugate posterior is also Beta distributed, and, under that updating process, the mean of the Beta distribution[1] exhibits diminishing marginal effects of both participation and avoidance in a particularly intuitive manner.

The salient aspects of the parameter-size assumptions are generally their relative magnitudes. For example, there would be no effect on optimal strategies if each of p^s, p^f, c, and $V(T)$ were multiplied by any constant, and it is only the ratio of the initial *Beta* parameters to their update magnitude that is important under commitment. The specific parameter values therefore represent a somewhat arbitrary choice motivated by common sense: for example $T = 1,000$ provides a tractable many-period model, and $\delta = 0.99$ represents an annualised discount factor of around 0.53, which is reasonable given the widely-varying empirical data in that area (Frederick, Loewenstein & O'Donoghue 2002; Benhabib, Bisin & Schotter 2010). Intuition suggests that the absolute size of these parameters is qualitatively unimportant, and this conclusion is thoroughly supported by the extensive robustness checks presented hereinafter.

[1] The mean of $Beta(\alpha, \beta) = \frac{\alpha}{\alpha+\beta}$ — strictly, the computation of optimal actions under commitment requires only an assumed form for the mean, not the entire distribution.

5 A guide to the replication code

We supply all of the code used in both the main chapter and these supplementary materials. The files are summarised here, and introduced in their appropriate sections below. All code is commented to explain the purpose of each block, and each block starts with a fully commented parameter specification area, wherewith the user may freely adapt our specifications as they desire. The reader may replicate our results by running the code 'as-is', and they may extend our investigations by altering these parameters. Table A.2 details the purpose of each code file, and the interrelations between code files. In summary, we carry out most computations in Matlab, before importing to Stata for visualisation.

Table A.2:
A guide to the replication code files

Code File	Contents
"education_trap.do"	Stata code for all figures presented in the main chapter.
"education_trap.m"	Matlab code for all figures presented in the main chapter; feeds into "education_trap.do" for visualisation.
"education_trap_supp.do"	Stata code for the figures presented in this supplement.
"education_trap_supp.m"	Matlab code that feeds into "education_trap_supp.do".

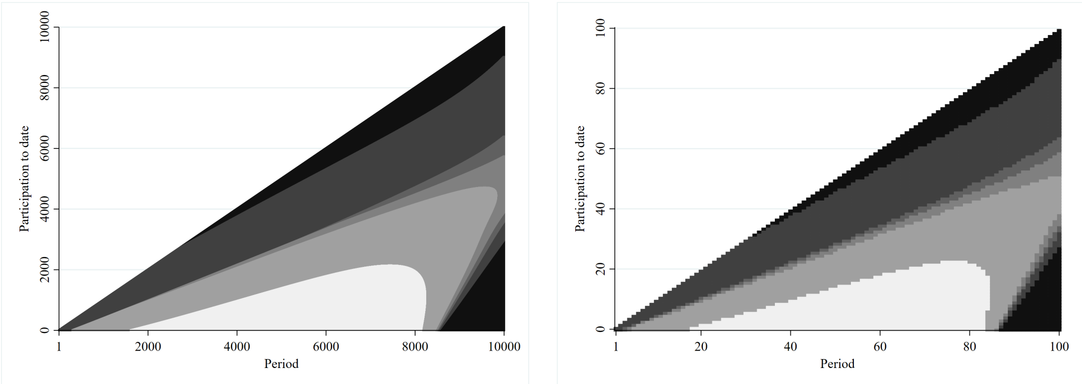
6 Robustness with respect to T, ι, δ, β, Π_1, VT, $V(n)$ convergence rate, $p^s = p^f$, and imbalance in $p^s \neq p^f$ or in initial ability endowment

The code for the main chapter allows for the straightforward manipulation of any parameters for any figure. Here we present a comprehensive set of robustness checks that show our qualitative findings remain even under substantial deviations around the preferred specification. We present output corresponding to Figure 2.3 of the main paper, since this best encapsulates the full results of our model. Figure 2.3 of the main paper is reproduced for convenience as Figure A.1 above, which also provides the legend that applies to all remaining figures.

We now carry out stringent robustness checks for each of the model parameters: The output below illustrates the effects of: an order or magnitude variation around T, a 5-fold increase and decrease in each of ι, $p^s = p^f$, VT and $V(n)$ convergence rate, a 5-fold imbalance in initial ability endowment and in p^s cf. p^f, substantial variation in the discount parameter δ from 0.99 ($\equiv$ 0.53 annually) to 0.997 ($\equiv$ 0.83 annually) and 0.97 ($\equiv$ 0.15 annually), and the introduction of hyperbolic discounting with $\beta = 0.9$ and $\beta = 0.5$. A short commentary is provided after each Figure.

Each of these many changes has the expected quantitative effects and almost all leave the solution qualitatively unchanged. The only changes to affect our qualitative results are those where the initial balance between p^s and either p^f or c is changed dramatically, in which case the following outcomes can arise: First, $p^s = p^f$ could be made so small relative to c that the cost of participation precludes any agent from attaining a high-investment equilibrium pathway – Figure A.12 shows that $p^s = p^f = 2$ closely resembles the benchmark $p^s = p^f = 5$, whereas if $p^s = p^f = 1$ then the high-participation pathway largely vanishes. Second, the ratio $p^s : p^f$ could become so skewed that full participation would be optimal in almost any event, or that non-participation would be optimal in the majority of events. These findings show that our model is highly robust to reasonable levels of parameter variation, and it also illustrates the comparative static results discussed in the main chapter, in particular regarding the importance to parents and educators of maximising p^s and minimising p^f.

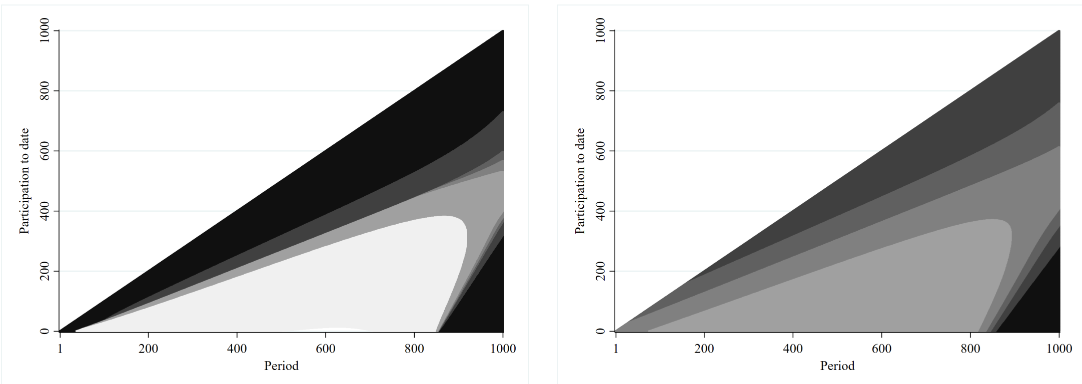
Figure A.3: ***cf. Fig. 2.3:***
Variation from number of periods $T = 1,000$



(A) $T = 10,000$ (B) $T = 100$

We can see from Figure A.3 that a ten-fold increase in the number of periods produces a result that is qualitatively indistinguishable from our benchmark specification. Similarly, a ten-fold reduction in periods is distinguishable only because it reduces the 'resolution' of the output.

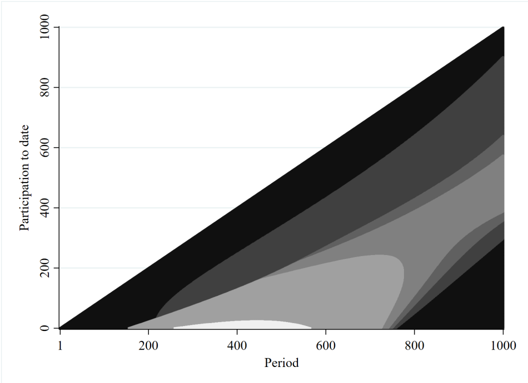
Figure A.4: ***cf. Fig. 2.3:***
Variation from update magnitude $\iota = 0.05$



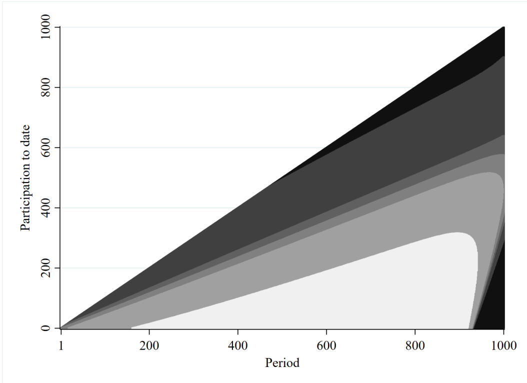
(A) $\iota = 0.25$ (B) $\iota = 0.01$

We can see from Figure A.4 that a larger update parameter would make the high-participation pathway easier to attain, and the low-participation pathway more entrenched, but it would not alter our qualitative findings. Conversely, a smaller update parameter would make the high-participation pathway harder to attain, and the low-participation pathway less entrenched, but again it would not alter our qualitative findings.

Figure A.5: ***cf. Fig. 2.3:***
Variation from discount rate $\delta = 0.99$



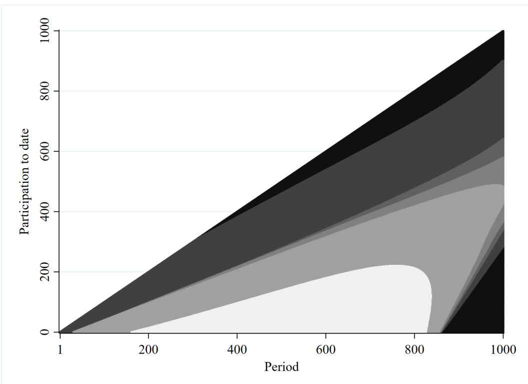
(A) $\delta = 0.997$



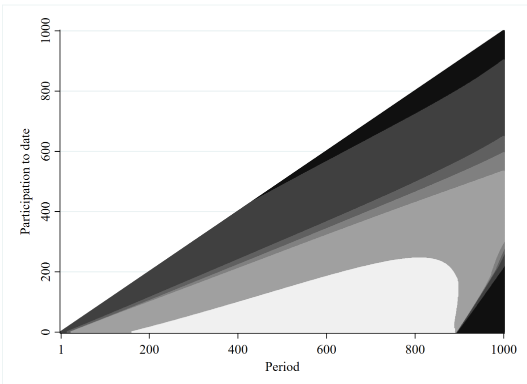
(B) $\delta = 0.97$

It can be seen from Figure A.5 that a smaller discount rate would make the high-participation pathway more attractive, and the low-participation pathway less attractive, but that it would not alter our qualitative findings. It can also be seen from Figure A.5 that the low-participation pathway entails greater participation when developmental benefits are less discounted. Conversely, a larger discount rate would make the high-participation pathway less attractive, and the low-participation pathway more attractive, but again it would not alter our qualitative findings. It can also be seen in Figure A.5 that the duration of the high-t low-n participation region is reduced when discounting is greater.

Figure A.6: ***cf. Fig. 2.3:***
Introduction of present bias β



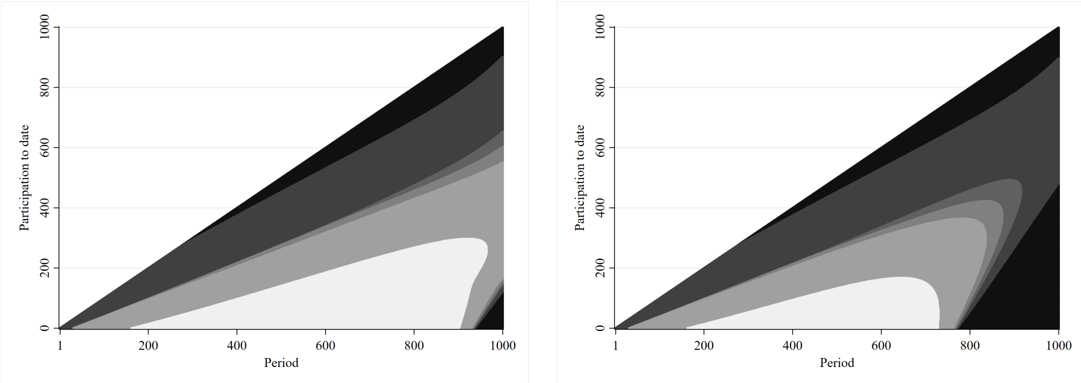
(A) $\beta = 0.9$



(B) $\beta = 0.5$

We can see from Figure A.6 that introducing present-bias into our model does not affect our qualitative findings. Quantitatively, it has very little effect on the high-participation pathway, but reduces the amount of participation that should optimally be made by disadvantaged children as they approach the end of compulsory schooling.

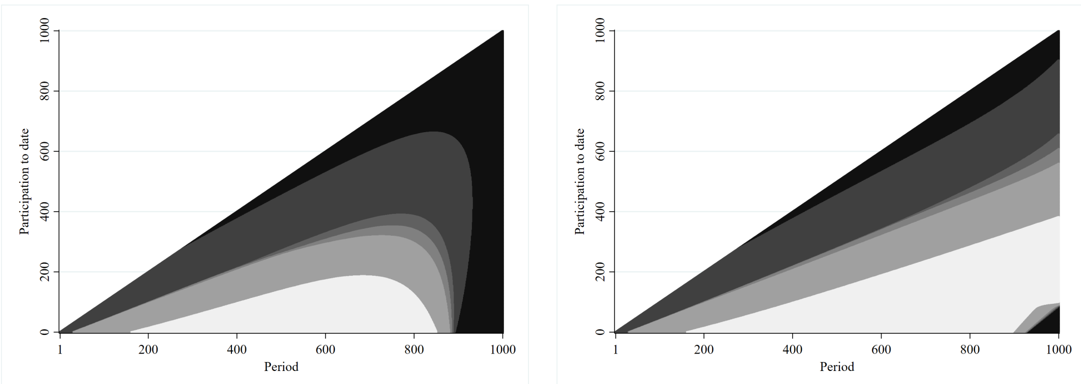
Figure A.7: ***cf. Fig. 2.3:***
Variation from attainment $VT = 10,000$



(A) $VT = 2,000$ (B) $VT = 50,000$

We can see from Figure A.7 that changing the life-cycle payoff for human-capital accumulation does not affect our qualitative findings. Quantitatively, it has very little effect on the high-participation pathway, but again there is a larger quantitative effect for disadvantaged children: a lower benefit from schooling would contract the high-t low-n participation region, whilst a greater benefit from schooling would expand the high-t low-n participation region.

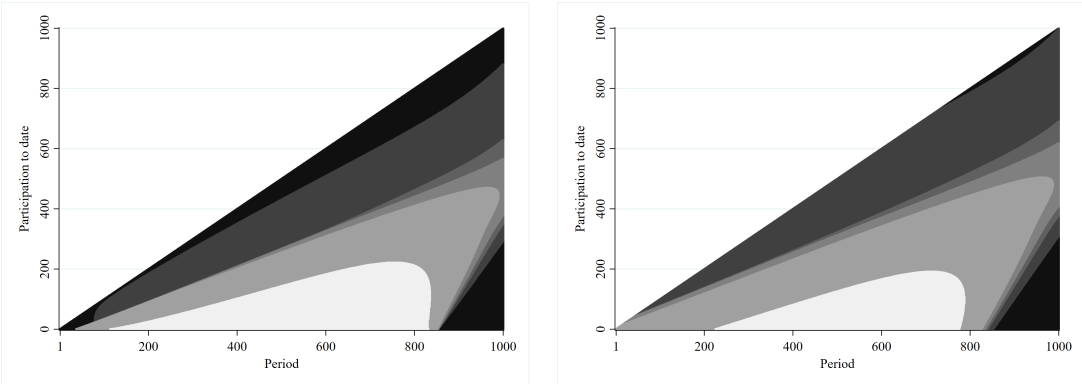
Figure A.8: ***cf. Fig. 2.3:***
Variation from a rate of convergence towards the maximum value of attainment of one hundredth of the remaining distance per period



(A) *one five-hundredth* (B) *one twentieth*

We can see from Figure A.8 that a slower convergence toward maximum schooling benefit would extend the high-t low-n participation region to include much larger values of n, because increased prior participation would diminish the marginal product of education by less. We can also see that this change would not alter our qualitative findings: Not only is the effect relatively minor, but it is also situated away from the relevant equilibrium pathways. As we would expect, the converse effect is evident for a faster convergence towards the maximum schooling benefit, although this again does not affect our qualitative findings.

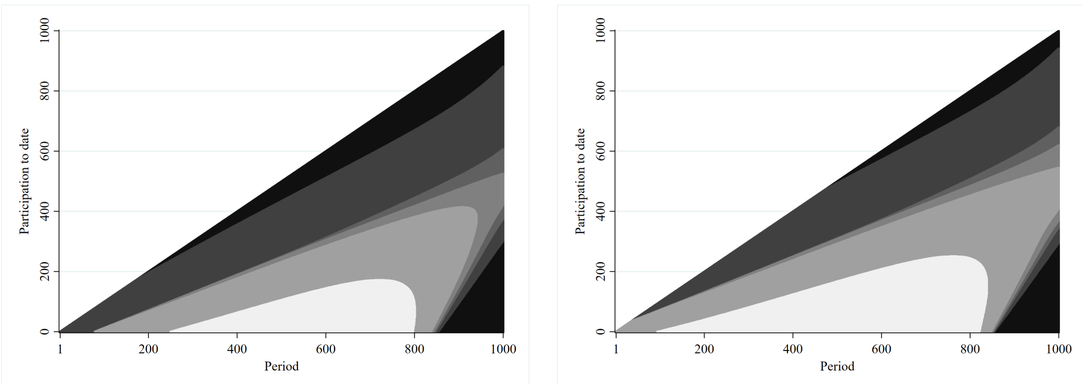
Figure A.9: ***cf. Fig. 2.3:***
Symmetric variation from initial ability stability Π_1 ~*Beta(2.5, 2.5)*



(A) Π_1 ~*Beta(0.5, 0.5)* (B) Π_1 ~*Beta(12.5, 12.5)*

We can see from Figure A.9 that the effect of variation in the stability and precision of our agents' ability distribution has very little qualitative or quantitative effect on our findings.

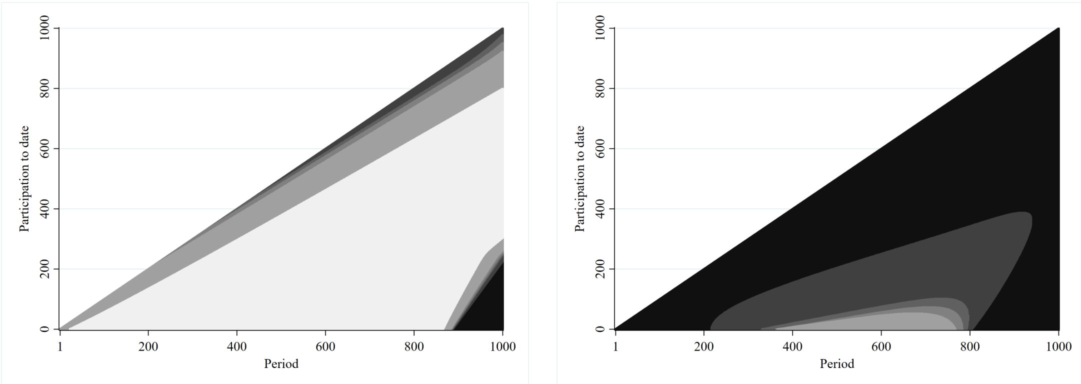
Figure A.10: ***cf. Fig. 2.3:***
Asymmetric variation from initial ability stability Π_1 ~*Beta(2.5, 2.5)*



(A) Π_1 ~*Beta(5, 2)* (B) Π_1 ~*Beta(2, 5)*

We can see from Figure A.10 that an advantageous skew to initial ability would make the high-participation pathway easier to attain, and the low-participation pathway commensurately more difficult to attain, but that the effect on our qualitative findings is minimal. The converse effect is true for a disadvantageous skew in initial ability. The important conclusion here is that an agent's innate ability (dis)advantage could quite easily be reversed by the opposite (dis)advantage in her early years development opportunities.

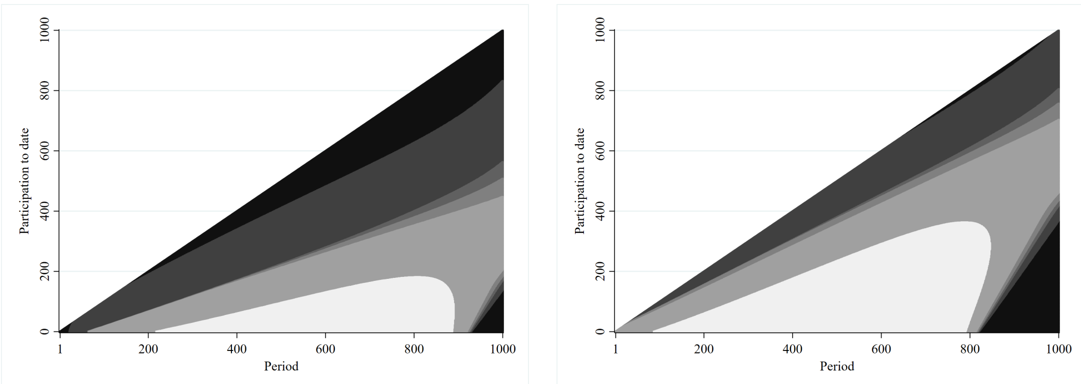
Figure A.11: ***cf. Fig. 2.3:***
Asymmetric variation from psychic payoffs $p^s = p^f = 5$



(A) $p^s = 2$; $p^f = 10$ (B) $p^s = 10$; $p^f = 2$

It can be seen from Figure A.11 that if the psychic cost of failure were substantially larger than the psychic payoff to success, then the high-participation pathway would be almost impossibly hard to attain. This clearly does not reflect reality, but nevertheless our qualitative conclusions would change little. Conversely, if the psychic payoff to success were substantially larger than the psychic cost of failure, then the high-participation pathway would be much easier to attain - in this case it could become almost certain except for children who experience a protracted lack of educational opportunities. This figure therefore illustrates the discussion of the main chapter, which emphasised the importance of adopting pedagogical practices to increase p^s and decrease p^f.

Figure A.12: ***cf. Fig. 2.3:***
Symmetric variation from psychic payoffs $p^s = p^f = 5$



(A) $p^s = p^f = 25$ (B) $p^s = p^f = 2$

We can see from Figure A.12 that the effect of symmetric change in psychic payoffs is less marked than an asymmetric change. In either case, our qualitative findings are unchanged, although a substantial reduction in those payoffs makes the high-participation pathway rather harder to attain. An even larger reduction to $p^s = p^f = 1$ would change our qualitative findings, since in this case almost all decisions would be dominated by c the cost of participation, and so the high-participation equilibrium would no-longer exist.

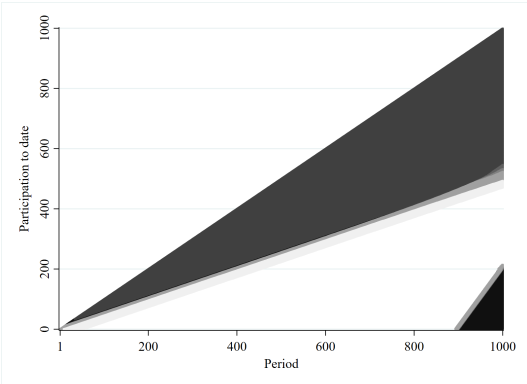
7 Robustness with respect to truncated normal ability distribution, agent naïvety, linear returns to education $V''(n) = 0$, and increasing returns to education $V''(n) > 0$

The final set of robustness exercises confirm that the qualitative results of the chapter are not an artefact of the specific (albeit sensible) choices of beta-distributed ability, or of exponentially diminishing participation benefit. The first panel of Figure A.13 shows that our qualitative findings are robust to an alternative truncated-normal ability distribution, and Panel C of Figure A.13 shows that our assumption of diminishing returns to education mainly affects outcomes towards the end of compulsory schooling, and that effect is concentrated away from most equilibrium pathways. Neither alternative assumption affects the existence of the education trap.

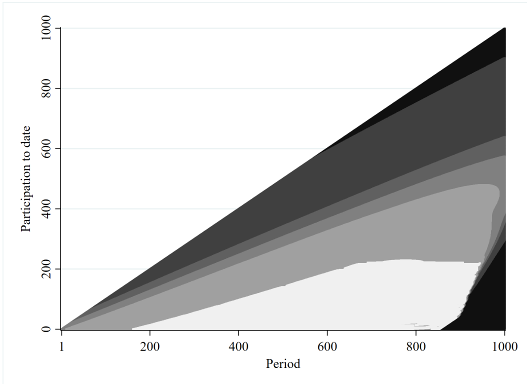
For completeness, we also confirm that our findings are robust to increasing returns to education and to naïvety. In Panel D of Figure A.13 we can see that increasing returns to education would only amplify the effect of linear returns, and that again these effects are concentrated into the final periods of the model and away from the important equilibrium pathways.

Naïve agents are affected by psychic payoffs this period, but are not aware that they will also be so affected next period. The naif therefore provides an archetype of the most extreme version of the human tendency to discount any visceral influences over our future behaviour, as described by Loewenstein (1996). One would expect naïve behaviour to be somewhere between sophisticated and myopic behaviour (myopes do not consider the existence of any future decision periods – we showed robustness to myopia in the main chapter), and this does indeed seem to be the case. certainly our qualitative results are robust to naïvety.

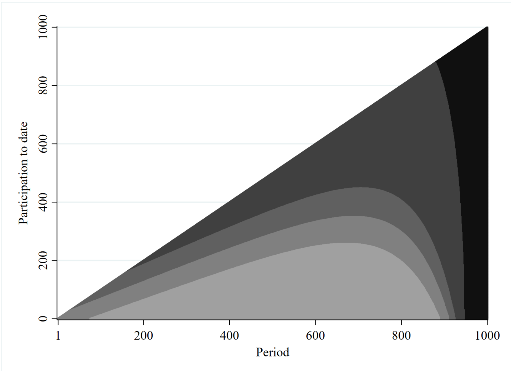
Figure A.13: ***cf.* *Fig.* *2.3*:**
Alternative functional forms and sophistication constraints



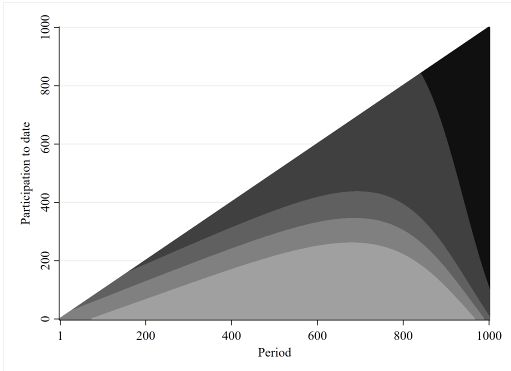
(A) $\Pi \sim$*Truncated normal*



(B) *Naïve agents*



(C) $V''(n) = 0$



(D) $V''(n) > 0$

3 Supplement to Chapter 3: A guide to the code

1 The basics of Fortran

Fortran is a somewhat old-fashioned programming language with more pedantic semantics than most. On the other hand it makes extremely fast programmes.

Programming in Fortran is a two-step process; one first compiles the code into an executable programme, then one runs that programme. I used the free IDE 'codeblocks' to edit the source files, and if you wish to do likewise then you will be able to directly open the .cbp project files. If you use an alternative editor you can completely ignore those files. The code itself is written in Fortran .f90 files. I then used WinSCP to transfer the .f90 files to the high-end computing cluster at Lancaster University, taking care to ensure that they were transferred in text mode. Finally I used a Putty command-line interface to compile and run the code, the latter via a bash .com script. The bash scripts and a text file listing the relevant commands to running and debugging in Intel Fortran are included with this guide.

A couple of 'interesting' details to watch out for in Fortran:

- The line length cannot exceed 72 characters including initial space characters (the size of a computer card). This includes comments, although most compilers will allow comments to overrun – but watch out if using !$OMP as this looks like a comment to the compiler and so might not throw up a warning if it's over line-length (this cost me about a week of my life). lines can be carried over using '...' but it's probably best to avoid this where possible.

- It's good practice to start any module, subroutine, etc with the command: 'implicit none'; you will then need to define in painstaking detail all variables which you wish to use, and, in the case of functions or subroutines that are passed inputs, their 'intent', ie. do they merely go in to the function (in), are they generated by the function (out), or both (inout). Wherever a function is passed to another function it will need to be present in the latter via a dummy 'interface' block, where the variable definitions and intents need to match those of the actual function.

2 An overview of the code

The (Model 2) code contains the following files (in order of increasing seniority):

minimization.f90 This file contains the minimisation algorithms which we will apply to iteratively reduce the log-likelihood. You probably don't need to change it. The current configuration requires the process to converge twice (once through dfpmin, and again through dfpmin2). 1st convergence does not imply continual convergence of re-runs because dfpmin progressively builds up a numerical approximation to the Hessian matrix of the function 'func' that it is passed, however as the Hessian will in general change at different points, this process is imperfect. Hence restarting it every (initially) 250 iterations improves convergence rate, but can change the local minimum that is converged towards when these exist – possibly for the 'better' and possibly for the 'worse'. The parameter itchunk allows you to change the frequency of restarting the Hessian approximation process, separately for dfpmin and dfpmin2. Similarly you can set the maximum number of iterations for each process via their respective itmax parameters. dfpmin operates by finding the current numerical gradient (calling 'gradient'), then searching along that direction for a suitably large decrease in LL (calling 'lnsrch').

integration.f90 This file is a library of numerical integration procedures, some of which are called upon. You will not need to change it.

globvar.f90 This file defines all those variables that will be accessible from all (of the below) other modules. They will be stored and hold their values globally. You may well wish or need to change some of these entries if using different data or trying to do different things. The first 6 parameters in-particular may well need to be changed, and possibly the nsubgp parameter. This file is especially well commented.

normalizations.f90 This file sets (many of) the parameters to be estimated to initial values (it is not especially related to CES normalisation). The code optimises a likelihood function over a many-dimensional parameter space (1223 in the case of Model 2), and the initial values for each of these parameters is set by the file 'point_fix_...', where '...' is specified in line 29 of main.f90. However, there are more parameters that could conceivably be optimised over, and so normalisation takes care of setting these to sensible numbers (such as 0 where a parameter is abstracted from). You will probably only need to change this file if adding additional parameters to the model, since many of its most relevant parameters are overridden by 'point_fix' anyway.

matrix.f90 This file is similar to 'integration.f90' in that it provides a library of routines that are called in other files... things to do with matrices like inverting them and finding eigenvalues and so forth.

policy.f90 This file provides what CHS refer to internally as the policy-function: it is the mapping which ensures that the indirect effects of the 'policy equation' (aka. investment equation) are incorporated into the factor-loadings for each of the state variables at each time period. Since it loops over the measures (27-43) which are used to derive the latent state variable of parental investment, the values 27 and 43 will need to be changed if the positions (ie columns within the data) of the set of factors that measure parental investments changes for any reason (eg. if others are added, or if a different dataset is used).

utilities.f90 This file provides a set of functions and subroutines that sit within the likelihood function of 'likeIPE.f90' (amongst other places). These routines include those which compute the unscented transform, and which map out the transition and policy functions. The transition functions (CES and linear) map out the evolution of the state variables through time, and the CEStransition function is the main object of interest in the whole exercise: it encodes equation 2 of the main chapter, and its parameters are (mapped onto) those that are published in Tables 2 and 3 of the main paper. The linear transition function maps out the laws of motion of the other state variables (equations 5 of the main paper.) the policy function is CHS-code terminology for the investment function (equation 3). Thus, if you are content with the present formulations of these functions then you will not need to change this file, but if you wish to adapt them then this is the place to do it! Note that, as discussed in the main paper, the state variables are measured and encoded in log terms, and so the CEStransition function is transformed into that which operates on log variables to produce log variables. The other functions operate directly on log variables (as can be seen in the paper).

WriteResults.f90 This file does what it says on the tin. It's fairly self explanatory, and will need to be adapted if and only if there are additional variables which you create and which to know the estimates for. The routine 'writeresults2' writes the 'transition_shares' files, which transform the main coefficients of the model into the parameters of interest.

mappings.f90 The parameters over which the likelihood function is maximised are contained in the variable 'theta' defined in main.f90 and passed to various subroutines. Mappings.f90 contains two subroutines: 'dimtheta' is used just once at the start to write the file 'check.out' which organises the information required to dissect which variables will be extracted from which locations within 'theta', and how big theta should therefore be. It is very-much parallel to the 'getpar' routine that actually extracts the current state variables from the vector 'theta'. These are not all just taken as written – for example some parameters have restricted domains (such as [-∞,1] for complementarity parameters), and so mappings transforms the variables such that they lie within the permissive range. Theta is then just a vector of numbers which is transformed by mappings to define the variables and parameters (listed in global.f90) that the likelihood function in likeIPE.f90 uses to calculate the log likelihood of the model given the data. Theta is therefore the argument of the 'loglikelihood_IPE' function that is passed to it via 'dfpmin' in line 39 of main.f90. All of the iterative fitting and convergence of the model is done within that line. Parallel changes to both subroutines will be needed if you wish to change/add to/remove any of the parameters to be estimated.

probability.f90 This is another ancillary file which provides subroutines, such as a numerical approximation to the normal distribution. As with matrix.f90 and integration.f90, these would all just be standard commands within most programming languages. Changes are unlikely to be needed.

initialize.f90 This file transforms the data into an array within fortran, and extracts individuals' datapoints into their appropriate arrays within the model. This will need to be changed if the data used changes in any way, but CHS have made it quite easy to do so:

- 'nvar' is the number of columns in the dataset,
- 'data_aug3.raw' is the filename of the dataset, which needs to be saved in the same directory as the compiled code programme file (which will be named 'a.out' by default).
- 'nequation' gives the number of measurement equations for each period, as it says on the tin,
- 'stage' specifies the developmental stage of each period - here the first three transitions are stage 1, and stage 2 thereafter: stage can be chosen to match your theoretical prior on whether there are multiple stages of development, and how the periods of data are mapped into them.
- 'ly' gives the location of the measurement variables for the latent factors (state variables): for example here the first measurement is gestation length and it is in column 21 of the dataset. The remaining measurements are then contiguous through to column 88.
- 'ldy' gives the location of the missing indicators for the measurements. Here, the missing indicator for gestlength is in column 98, and the other indicators

are contiguous thereafter, so line 32 of 'initialize.f90' shortcuts typing these locations out by simply adding 77 to the values of 'ly'. One could replace this code with a full list of values similar to that which defined 'ly' if one wished to do so. In this code, each variable for each individual is twinned with a missing indicator: so row 1 of column 98 would contain a '0' if gestlength were missing for individual 1, and it would contain a '1' if gestlength were available for individual 1. It is important to generate these missing indicators and to tell the code where they are located via 'ldy'.

- 'lsubgp' is analagous and tells the code the location (column number) of the variables which indicate membership of a certain subgroup, incase you wish to normalise your CES transition function to the mean value of a particular subgroup rather than of the whole population. This choice is set via 'CES-normopt', which will be left at '0' to normalise at the sample mean values of each state variable. If CESnormopt=0, then there is no need to specify any data within 'lsubgp'

- 'equindex' maps out for each period which of the (in this case 64) measurement variables are (in principle) available during that period. For example, '1' only appears in period $t = 1$ because gestlength is only measured at birth. The numbers here relate to the entries within 'ly', so that 1 relates to the location ly(1)=21 within the data. Note that the length of each period's 'equindex' vector should match that specified in 'nequation'.

- 'lx' sets out for each time period where the covariates X are located within the dataset. CHS (and I) control for: a constant term, and dummies to indicate: whether the assessments were collected age age +1yrs relative to others (since the measurements are biennial), gender, whether the child's mother was under 20 at birth, whether the child was born in cohort years 1987-2001, as well as the child's age in the year 2004. These are located in the columns of the dataset indicated by 'lx'. There are 6 of them, as indicated by the parameter 'nx' in globvar.f90, and all 6 are used as controls in all measurement equations in all periods as indicated by the fact that 'mx' is set ='nx' in all cases in line 59 of initialize.f90.

- Hereafter nothing should need changing, unless new state variables are introduced or similar, because the file simply takes the information supplied above and uses it to define the datapoints for each individual within their relevant arrays.

stdev.f90 This file contains a series of subroutines that map back and forth between a vector 'thetagrad' of the (non-measurement) model parameters and their named parameter arrays within fortran. The various subroutines feed into lines 44-63 of main.f90, which calculates standard errors via the standard hessian inverse approach for MLE. In summary:

- 'dimthetagrad' determines the length of the vector 'thetagrad', that is the number of parameters that standard errors will be calculated for. Note that inverting a matrix of, for example, 1233 rows and columns would be computationally infeasible, so CHS exclude the measurement factor loadings and X coefficients that make up the lion's share of those variables.[2]

- 'transform' maps the fortran variables into 'thetagrad'

- 'getparstdev' provides the opposite mapping; this is called within 'logdensity' on line 226 of 'likeIPE.f90', which therefore calculates the log likelihood based upon the current values of 'thetagrad'; this, in turn, allows the numerical gradient to be calculated for each element of thetagrad.

- 'sdpar' then extracts the estimated variances from the inverted Hessian matrix on line 63 of main.f90.

There is then a sister-set of subroutines with '2' at the end. These run analogously to the above, but transforming the model parameters into parameters of interest ahead of calculating standard deviations. for example complementarity parameters are transformed into elasticities of substitution, and γ coefficients from the main technology equation (2) are transformed into shares, as indicated in the main results tables 2 and 3 of the main chapter. The section of code in main.f90 that calls the original set of subroutines is mirrored in the 'stdev2' subroutine that constitutes 'stdevIPE.f90'.

likeIPE.f90 This file contains the likelihood function. It is useful to read Cunha (2011), where the estimation procedure is described in detail. I have added comments to identify some lines of the code with the equations presented in that paper.

The first block calculates the probability (likelihood contribution) that the observed measurements for individual i in period $t = 1$ would have taken place, and then updates the factor means ('a') and factor variance-covariance matrix 'P' for each individual i based upon those observations via the Kalman Filter.

[2]This procedure will be valid if the covariance of the measurement system with the remaining (of interest) parameters is zero, the extent to which this is assumption is restrictive could probably be debated, but we will not do so here.

Lines 73-88 then save the estimated factor score for each individual, and calculate group mean levels – this requires transforming to the true (not log) factor scores, averaging, then taking logs, because the mean of the log factor scores would not give the log of the mean factor scores.

The next block of code (‘do t=2,ntime’) starts in period 2. This is because its job is to predict the present-period factor scores from the last-period scores via the transition equations using the unscented transform. It is followed by the measurement block which was described in the preceding paragraph, which now essentially assesses how well the transition equations predicted the next-period measurements on each factor score.

The final block of code, from line 166 to 184, carries out the shrinkage operation to penalise the log likelihood for variations in real terms between the investment equation coefficients within each stage.

‘probcont’ has kept track of all the contributions to log likelihood, whilst ‘probextra’ has included only the shrinkage contribution and the contribution from measurements of parental income. The former contribution doesn’t exist under the original CHS ‘anchoring’ normalisation procedure, and the latter doesn’t exist under alternative models which don’t correct for endogenous investment (such as CHS Table IV). Thus, when analogous ‘probextra’ contributions are included to remove the contribution of the CHS anchoring procedure, we can compare the remaining core log likelihood across all models. (In fact, it makes sense to exclude contributions form parental income measurements anyway, since these are assumed to be measured-without-error.)

‘logdensity’ runs analogously to ‘loglikelihood_IPE’, except that it calculates the likelihood contribution of each individual separately to feed into the standard error calculation process. There is no need to calculate facvar type data here, as that already exists because ‘loglikelihood_IPE’ has already been called. Also, because these routines don’t run nearly so often as ‘loglikelihood_IPE’, they are coded using an additional ‘if’ statement, that costs a little in speed and adds somewhat in the typeset elegance of the .f90 file. ‘logdensity2’ is identical to ‘logdensity’, save that it operates with the transformed ‘thetagrad’ via ‘getparstdev2’ in line 354.

stdevIPE.f90 This mirrors lines 44-63 of ‘main.f90’, but calling the alternative set of mapping routines ‘...2’ from ‘stdev.f90’ to calculate standard errors for the transformed parameters of interest. These are then printed in ‘transition_shares.out’ files. See the final paragraph in ‘stdev.f90’ notes above.

main.f90 This file runs the show by calling upon all the other files' subroutines as necessary. It starts by counting dimtheta, it then pulls values from 'point_fix.out' and maps them into the programme variables via 'getpar'. The 'estimates' procedure writes the results implied by 'point_fix.out', before the work of the programme is done by iteratively maximising the log likelihood using 'dfpmin'. 'estimates' then writes the results implied once that process is complete, before lines 44-63 calculate standard errors. the final 'estimates' call on line 65 overwrites the previous calls with output that now includes those standard errors. 'stdev2' on line 66 calls the stdev procedures for the transformed parameters of interest, which are printed via the call to 'writeresults2' on line 839 of 'stdev.f90'.

3 The data file

The supplementary materials to CHS Cunha, Heckman & Schennach (2010) provide a detailed discussion of the necessary arrangement of the datafile.[3] This is therefore not repeated here, but rather summarised very briefly. Each row of the data should represent one individual at one particular time-point. Stata refers to this arrangement as 'long form'; see 'help reshape' in Stata. missing indicators are needed for each variable as a separate column - see the discussion of 'ldy' (amongst others) within 'initialize.f90' above. The data are needed in .raw format, which can be produced from Stata using a command such as: 'outfile using data_aug3 , wide' where the 'wide' option stops Stata from wrapping lines of data into new lines of the .raw file, which would mess everything up.

Two additional cautionary notes: first, your life will be much simpler if you ensure that all variables are positively coded in your datafile, such that larger values indicate an increase in the latent factor. The alternative would be to correctly initialise negative factor loadings, but this would be more likely to lead to error, and it would inhibit your ability to inspect final estimated factor loadings for any anomalies.

Second, if your datafile is very large then there could be a (single-thread) memory requirement spike early in the programme where 'initialize.f90' loads the datafile into memmory before reading its contents into their assigned variables and dropping the raw dataset; this could cause a 'sigsev' or related error. As an indication of memory requirements during this initial spike, my experience suggests that they could reach slightly over double the size of your .raw datafile. Memory allocation can be increased in a bash job script by a command such as: '#$ -l h_vmem=1G'.

[3] A very useful spreadsheet detailing their data and its constituent variables is also available there.

4 Parallel processing

The code is endowed with OMP (Open Multi-Processing) commands to parallelise the 'gradient' subroutine within 'minimization.f90'. All OMP code lines begin with the tag '!$OMP', which serial Fortran compilation will disregard because the initial '!' is read as a comment. Thus the OMP code has no effect when the programme is compiled using the single-processor compilation command described within the file 'commands.txt' that is included with this guide. That file also includes a parallel-processor compilation command (which simply includes the flags '-qopenmp -static-intel'), under which the OMP code becomes operative. I parallelise the 'gradient' subroutine because the vast majority of the CPU work for the programme is done through that subroutine (it calculates the effect of a small change in each successive model parameter on the global log-likelihood during each iteration), and so parallel processing with n cores decreases run-time by very nearly a factor of n.

In addition to the OMP code in the 'gradient' subroutine, there is also one line of OMP code at the end of 'globvar.f90', and two lines in the 'dfpmin' subroutine of 'minimization.f90'. The former ensures that variables that update before or during each likelihood calculation are stored separately for each thread (else mayhem would ensue as parallel threads attempted to write and read each other's updates to shared variables), whilst the latter ensures that it is the master thread that executes the initial benchmark likelihood calculation (this is important because if that calculation were dispatched to an alternative cpu, that cpu might not have the initialised variable values because there has not yet been any implied synchronisation since no OMP parallel command has yet been reached). For a good guide to OMP in Fortran, see https://www.openmp.org/wp-content/uploads/F95_OpenMPv1_v2.pdf.

Once the code has been compiled into a parallelised executable (by default 'a.out'), we require a slightly different bash script to run that executable – this is provided in the file 'run_parallel.com' (cf. 'run_single.com') that is supplied with this guide. If you experience difficulty in reading and editing '.com' files, simply change the extension to .txt and open with your favourite text editor.

5 Point_fix.out

This file presented a large part of the challenge in understanding the functioning of the programme, and its development will form a major component of applying this code to any new dataset. The code in 'normalizations.f90' will, in principle, allow us to comment-out the reading of 'point_fix' from 'main.f90', and to include it again once the programme has converged for the first time by using 'point.out' from that first convergence as the start-point for future runs. This approach would allow us to assess the output from that first run for any contra-indications, (such as negative or negligible factor loadings; inexplicable parameter estimates; and so forth), which could inform us as to how to move from a current local minimum or corner solution in the parameter space to an improved model fit elsewhere in the parameter space.

The first problem with this approach is that, at present, the only adaptations that we could make to point.out would be the untargeted addition of random noise to all estimates. This approach is a useful robustness check, and it helped me in the initial replication of CHS, but it is not always the most effective way of reaching a global minimum within the parameter space. To target specific irregularities with the current point estimate we require an intimate knowledge of the 'getpar' routine within 'mappings.f90'. By working through this file line by line, we can map each of the 1200+ components of the point_fix vector onto its eventual model parameter. This enables us to test the implications of specific perturbations to the measurement equations or other parameters.[4] From the perturbed point, the model may return to around the current minimum, possibly with slightly improved or worsened log-likelihood, or it may diverge toward an alternative local minimum, which may have greater or lesser log-likelihood than the current point.

The second problem with this approach is that, in my experience, the crude initial values contained within 'normalizations.f90' do not suffice to get close to the global optimum point. For this we require a rather more informed initial point, in particular with regard to the measurement equations. To derive a more informed initial point for the measurement parameters 'beta' and 'H', I wrote the stata do file 'point-fix-derivation.do': this produces, in the correct order, the regression coefficients of each control variable in each measurement equation as initial values for the 'beta', and the natural logarithm of the sample variance for each measurement as initial estimates for the 'H'. This approach, when augmented with a few different sets of very small random noise, allowed me to replicate the original CHS results before moving on to adapt and develop their econometric model.

[4]Typically modifications to only the technology equations won't achieve a better fit – the measurement equations and auxiliary parameters such as the initial covariance matrix very much drive these results.

4 Supplement to Chapter 4: Data compilation

1 Overview

Access to the Millenium Cohort Study Data is controlled via the UK Data Service, and the survey overview can be found at: https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000031. Neither secure access nor special liscence access are needed for the data used in this study, and so it is relatively accessible for replication and extension studies.

The Stata code which compiles our dataset from the various study files is attached to this supplement. This code extracts, collates, cleans and synthesises data from 41 separate raw data files, which (following UK Data Service approval) can be downloaded from the website above under the study numbers (SN): 4683, 5350, 5614, 5795, 6411, 6847, 6848, 7464, and 8156.

The data cleaning and synthesis process is simple in principle, though somewhat more involved in practice. Many variables required only the removal or relabelling of missing responses and responses such as 'not sure'. For other variables, excessively skewed or sparse responses required some straightforward recoding. In cases where there were large numbers of closely aligned measurements, these were combined into lower-dimensional scores using exploratory factor analysis and Item Response Theory. The approach taken here is exemplified in Section 3 below. Finally, some key variables were required to be nonmissing - these comprise the control variables which appear in each measurment equation, and parental income which instruments for investment. A brief overview of the approach taken for each of these variables is now supplied in Section 2. An overview of all final variables, their usage, and their source variables in the raw data is supplied in the supplementary file varlist.xlsx.

2 Cleaning and imputing key variables

Full details and additional comments can be found in the Stata code file. An overview for each key variable is provided here. The most major requirement for imputation comes during the first wave, because around 600 first-wave non-respondants did respond in the second wave. In most cases, key variables for the first wave can be imputed directly from their values in the second wave.

teenmum *Dummy to indicate teenage motherhood; control variable.* For almost all individuals this is reported directly (in at least one of the first two waves), however a few ($< 0.1\%$) are missing this data. These were imputed using parent age, interview date, and child birth date information from the household grid datafiles.

gender *Dummy to indicate gender; control variable.* This is defined as per the wave 1 observation when available, and the wave 2 observation otherwise. There are no remaining missing observations.

english_at_home *Dummy to indicate whether English is spoken at home; control variable.* Waves 2 and 3 are used to plug gaps in wave 1 where needed.

age *the child's age; control variable.* Available in all sweeps, but at various accuracies: days, months, or year tenths dependent upon the sweep. Not available in 22 child×sweep instances out of 104,142 in final dataset – in these cases child age is set to its mean value for that sweep.

faminc *log of the family's OECD equivalized monthly income in real terms (2015 money); instrument for the investment equation.* OECD equivalized income is a key derived variable in the data, and it is observed in almost all useable child×sweep instances (it's hard to be certain of exactly what proportion, since useable cases are only defined after the cleaning and imputation process, but I'm confident that it is in excess of 99%.) In these cases, all that is needed is to adjust income into real terms, by using the CIPH price index data from the UK Office of National Statistics – this is their prefered price index and it includes housing costs. Details and data on this index can be found at: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/consumerpriceinflationincludesall3indicescpihcpiandrpiqmi. The OECD equivalized income takes account of family size and composition, and so it is designed to reflect more accurately the financial wealth : requirement ratio of a household (see http://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf) This makes it more suitable than raw income for instrumenting the investment equation.

For the few relevant instances where this variable is unavailable, values were imputed using a standard Mincerian wage equation for each period, augmented with a dummy to indicate whether both natural parents were present. To do this, the mother's age and years of schooling were used, or, where necessary, also imputed from data in other sweeps. For example, the increase in age between any two sweeps can be estimated very accurately using OLS, and years of schooling can be carried backward, unless it becomes greater than age in which case it is corrected to current age.

3 Combining closely related variables

A Simple confirmatory case

The MCS includes many apparently closely related measures of maternal psychological distress in sweep 1: amdepr00 – get depressed; amworr00 – get worried; amrage00 – get enraged; amscar00 – get scared; amupse00 – get upset; amkeyd00 – get annoyed; amnerv00 – get nervous; amhera00 – heart races. To confirm that it was appropriate to combine these, the following code was used (line 710 in the attached Stata .do file):

```
factor amdepr00 amworr00 amrage00 amscar00 amupse00 amkeyd00...
amnerv00 amhera00 , ml factors(2)
factor amdepr00 amworr00 amrage00 amscar00 amupse00 amkeyd00...
amnerv00 amhera00 , ml factors(1)
local obs= e(N)
polychoric amdepr00 amworr00 amrage00 amscar00 amupse00 amkeyd00...
amnerv00 amhera00 , pw
mat polychoric = r(R)
paran , all graph iter(500) saving(scree_pnc1-s1) mat(polychoric)...
n(‘obs’) centile(95)
```

The output from the first two commands allows us to compare, amongst other things, the proportion of the variance of each variable that remains unexplained under the principal factor model. In this instance, Table A.3 below shows that a second factor would meaningfully reduce the unexplained variance of just one variable, which suggests that there would be little value in deriving two latent factors from these measurements.

Table A.3: *Unexplained variance of each measurement*

	Uniqueness	
Variable	1-factor	2-factor
depr	.62	.62
worr	.76	.60
rage	.90	.88
scar	.74	.75
upse	.63	.61
keyd	.75	.71
nerv	.70	.61
hera	.76	.76

NOTE: a comparison of the unexplained variance of measures of maternal psychological distress under 1-factor and 2-factor latent factor models.

The paran programme in Stata is explicitly designed to test the dimensionality of a set of measurements Dinno (2009). (Here we use paran with polychoric correlations since our data are ordinal, but this has little effect on the results). Part of the 'paran' output is reproduced in Table A.4, and these results can be analysed on two levels. First, we can compare the total explained variance R^2 for each of the first n components - here we see that R^2 is 61% using just the principal component, and that this only increases by 9% when a component is included. Dual to this observation is the eigenvalue analysis of the measurement covariance matrix. If all of the variables were statistically independent then all eigenvalues would be identically 1. Thus unity provides a rule of thumb against which to judge the incremental contribution of each additional component. Intuitively, values greater than 1 explain more than 'one variable's worth' of the total variance, and so under the classical assumption that measurement error is uncorrelated between measures, any factor with an eigenvalue greater than unity necessarily has incremental explanatory power beyond the greatest possible contribution of measurement error. Of course, due to random chance we would not expect variables that were 'truly' independent to be statistically independent, and so the paran programme also runs the parallel analysis of Horn (1965) to obtain a corrected retention criterion. (Although, as can be seen from Table A.4 the effect of this is small when the number of measurements is reasonable). Thus a standard guide to the dimensionality of a set of variables would be to retain a number of factors equal to the number of adjusted eigenvalues that exceed unity. In this example, Horn's criterion unambiguously indicates that these variables measure a single underlying factor.

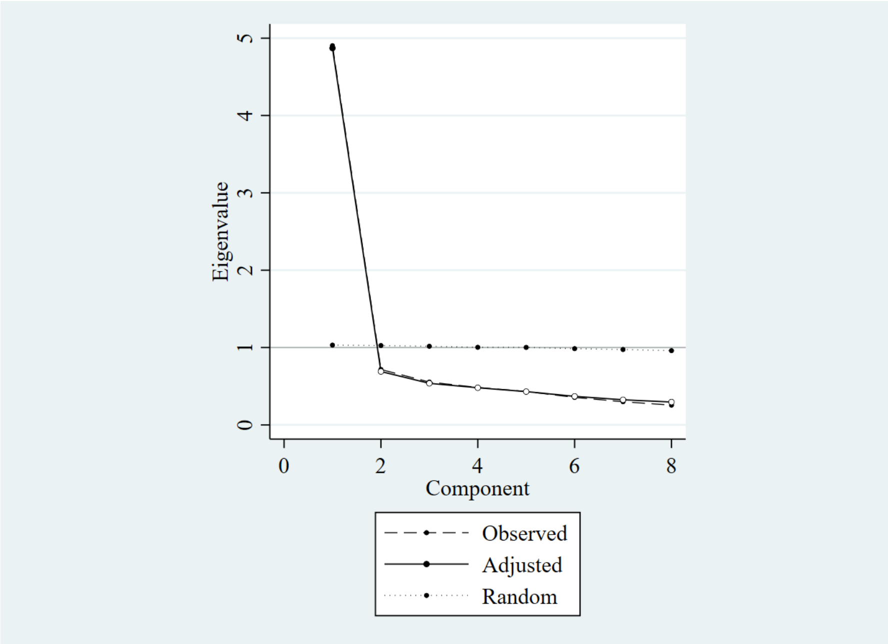
Table A.4: *Eigenvalues of the covariance matrix*

Component	Eigenvalue	Horn-adj	R^2
1^{st}	4.90	4.87	.61
2^{nd}	.72	.69	.08
3^{rd}	.55	.54	.07
4^{th}	.48	.48	.06
5^{th}	.43	.43	.05

NOTE: the largest five eigenvalues of the covariance matrix of maternal psychological distress measurements during sweep 1: unadjusted; adjusted to take account of random chance (Horn 1965); and the proportion of the total variance explained by each additional component.

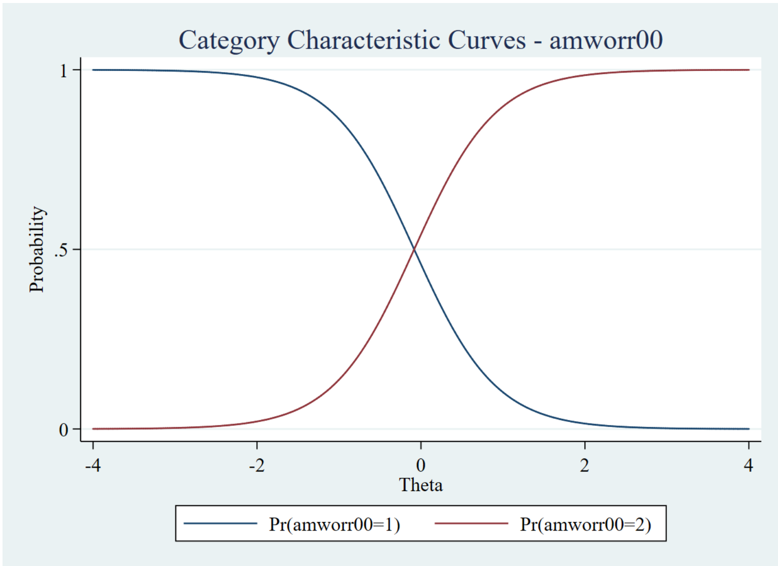
A visual approach to assess the dimensionality of a set of variables would therefore be to plot the eigenvalues of their covariance matrix onto a graph. Figure A.14 shows such a graph, and it also plots the counterfactual retention criteria of (Horn 1965), which can be seen to vary slightly around unity. Cattel (1978) famously likened this graph to a steep mountain, with a 'scree' that extends from below the main slope of the mountainside. From the start of this scree onwards, one can infer that each additional component captures little more than the pure measurement error of one particular measurement, which suggests that we should retain a factor for each eigenvalue that lies strictly above the scree slope. Again, in this instance it is unambiguous that only one component should be retained.

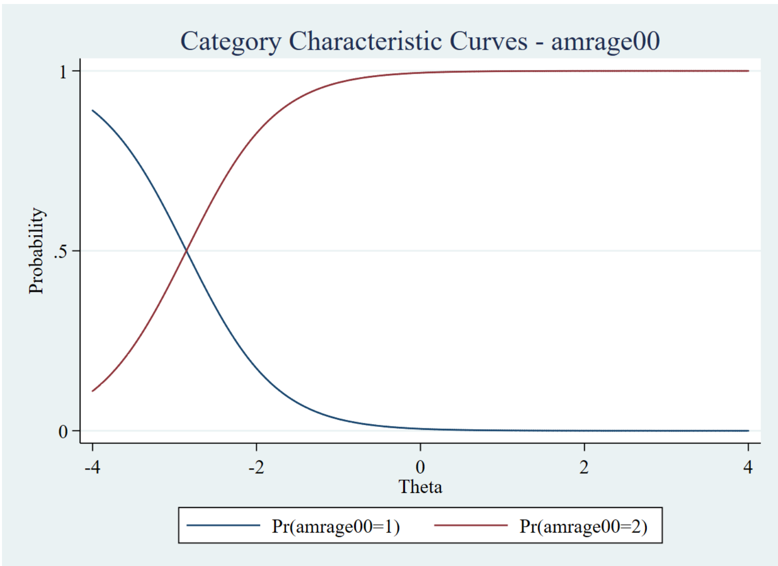
Figure A.14: *A scree plot for sweep 1 maternal psychological distress measurements*



Now that we have established that these variables load upon a single latent factor, we will estimate each individual's score on that factor using Item Response Theory (IRT). A simpler approach would be to use estimated factor loadings to compute the underlying latent factor as a weighted linear combination of the measurements, but IRT utilises the available information more efficiently. Under IRT, each response to each measurement is described by specific difficulty and discrimination parameters. The former is effectively the location of that measurement as projected onto the latent factor scale: for example a survey response 'I get depressed every day' would be a considerably stronger statement than the response 'I get depressed a few times per year', and so it would have a considerably greater difficulty parameter. The discrimination of each response is a measure of the variance of the latent factor score amongst individuals who give that response. A response such as 'I am not sure whether I get depressed' would probably have very low discrimination and hence a very flat curve along the latent axis. A restriction on the usefulness of IRT is that the latent factor distribution across each response is assumed to be normal, and so IRT does not perform well with highly skewed data. It is informative to examine the fitted IRT curves post-estimation, as a final confirmation of the validity of the latent construct. These plot the probability density distribution of the latent factor scores (theta) of the individuals who selected each response. In this particular case the underlying variables are binary, so there are only two response curves for each measurement (this is highly unusual within the MCS data). Figure A.15 shows two representative examples of these – both have fairly strong discrimination, but the first is considerably more 'difficult' than the second.

Figure A.15: *Representative IRT plots for s1 maternal psychological distress items*





A more complex example

We now carry out the same analysis for the full set of possible child noncognitive measurements from sweep 1. These are: amhapna0 – pleasant when nappy changing; amunfaa0 – pleasant when in unfamiliar places; ambrusa0 – pleasant when washing; amfeeda0 – pleasant when feeding interrupted; aminjua0 – pleasant when injured; ambatha0 – wary of strange person bathing; amwarya0 – wary of strange adult after 15 minutes; ambshya0 – wary of strange child; amfreta0 – wary of strange places; amsleea0 – fussy when sleeping; amfubsa0 – fussy when waking; amfuasa0 – fussy when sleeping in new place; and amcrysa0 – fussy when she does not get what she wants.

Table A.5 summarizes output from factor analyses that retain various numbers of factors. We can see that a single factor is inadequate since it would leave unexplained over 90% of the variance of 5 separate measurements. When compared with the 2-factor model, a third factor substantially reduces the unexplained variance of amunfaa0, amfubsa0, and amfuasa0, and makes moderate improvements to several other variables. The addition of a fourth factor makes more marginal improvements, with only amhapna0 being substantially better explained. This analysis is more equivocal than that of the previous example, but it suggests that three, or perhaps more factors might be needed to adequately capture the information in these data.

Table A.5: *Unexplained variance of each measurement*

	Uniqueness			
Variable	1-factor	2-factor	3-factor	4-factor
hapn	.98	.85	.83	.71
unfa	.84	.81	.69	.70
brus	.97	.78	.73	.74
feed	.97	.80	.77	.70
inju	.94	.88	.87	.80
bath	.88	.87	.87	.87
wary	.70	.62	.62	.62
bshy	.71	.63	.64	.64
fret	.60	.54	.53	.53
slee	.79	.79	.76	.76
fubs	.91	.86	.68	.67
fuas	.92	.87	.76	.76
crys	.89	.79	.70	.70

NOTE: a comparison of the unexplained variance of measures of child noncognitive skill under 1-factor up to 4-factor latent factor models.

Table A.6 summarizes the output from the paran programme. We can see that three variables clearly surpass Horn's retention criterion, and that a fourth factor is reasonably far below that threshold level. However, the first three factors together do not explain as much of the variance of these data as did a single principal component in the previous example.

Table A.6: *Eigenvalues of the covariance matrix*

Component	Eigenvalue	Horn-adj	R^2
1st	3.07	3.02	.24
2nd	1.77	1.73	.14
3rd	1.33	1.30	.10
4th	.94	.92	.07
5th	.84	.82	.06

NOTE: the largest five eigenvalues of the covariance matrix of child noncognitive skill measurements during sweep 1: unadjusted; adjusted to take account of random chance (Horn 1965); and the proportion of the total variance explained by each additional component.

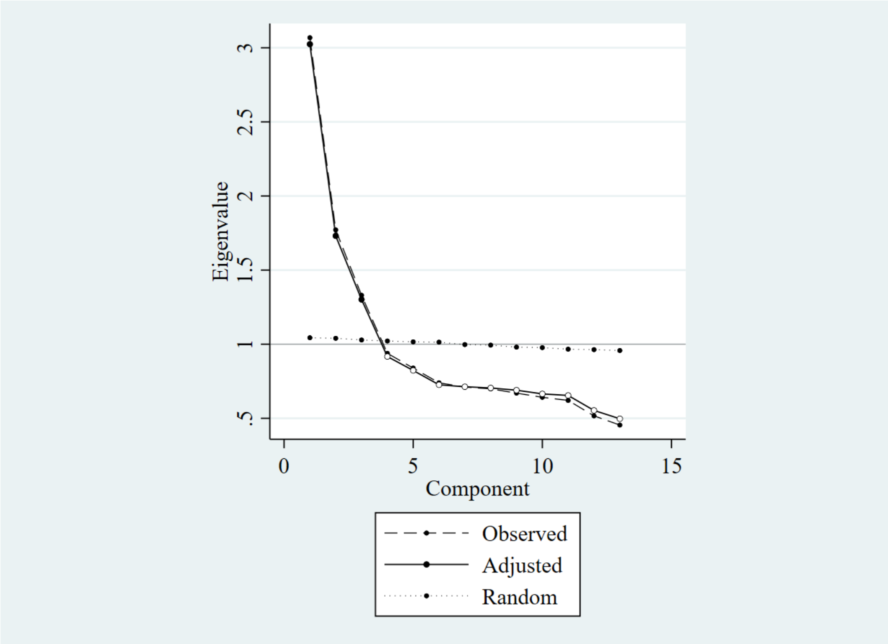
Figure A.16 shows the scree and counterfactual plots for these variables. From this image it is clear that a three-factor structure is appropriate, although the transition from mountainside to scree is much less distinct than in the previous case.

Once we have identified that three factors should be retained, we need to determine how the measurements should be allocated between these factors. To do this we need to carry out a factor rotation, which amounts to selecting an informative basis for the three-dimensional vectorspace that is spanned by the first three principal factors. Many possible criteria have been developed for factor rotation, some of which impose orthogonality between the factors and some of which allow those factors to become oblique (correlated). We do not require orthogonal factors for our main analysis, but it is good practice to check that one's allocation decisions are robust to the type of rotation used. The following commands perform a leading rotation of each type (orthogonal varimax and oblique oblimin), and their output is summarized in Table A.7.

```
factor amhapna0 amunfaa0 ambrusa0 amfeeda0 aminjua0 ambatha0...
amwarya0 ambshya0 amfreta0 amsleea0 amfubsa0 amfuasa0 amcrysa0...
, ml factors(3)
rotate
matrix list e(r_L)
rotate, oblique oblimin
```

In analyzing Table A.7 our objective is to allocate each measurement to the most appropriate factor grouping. To do this, we wish to select the largest factor loading from each row of (either half of) the table, and ideally we would prefer for this decision to be clear-cut. The rotation criteria used here both aim to make this

Figure A.16: *A scree plot for sweep 1 child noncognitive measurements*



allocation process clear-cut: the underlying mathematical tool for each rotation is to maximise the sum of the squares of the factor loadings. It is essential that rotation(s) such as these be performed prior to allocating the variables between factors, because the original factor analysis produced factors in a hierarchical manner, and so the first factor to be derived is likely to have the largest loading for the majority of the measurements regardless of the underlying factor structure. We can see from Table A.7 that the first five variables predominantly load onto factor 3, that the next four variables predominantly load onto factor 1, and that the last three variables predominantly load onto factor 2. (Of course the factor numbers are arbitrary). Only 'amsleea0' is difficult to assign. In this case I decided to allocate it to factor 2 as indicated by the oblimin rotation, since the varimax rotation is unnecessarily constrained.

Finally, once the measurements have been allocated to three separate factors, we estimate those latent factors using item response theory, and check the item characteristic curves as described in the previous subsection. Since the allocation of amsleea0 was ambiguous, it will be particularly important to check its characteristic curves. Figure A.17 shows the item characteristic curves for amsleea0 and amwarya0, and we can see that each response to each of these survey item provides useable information as to the individual's latent factor score. The curves for amsleea0 are flatter than would be ideal, which shows that the measurement is not perfectly aligned with the underlying factor, but each response is nevertheless well-defined and conveys a distinct signal of an individual's latent factor score. If the curves were substantially flatter than these we would conclude that

Table A.7: *Rotated factor loadings for a 3-factor model*

	Orthogonal – Varimax			Oblique – Oblimin		
Variable	factor 1	factor 2	factor 3	factor 1	factor 2	factor 3
hapn	.03	-.12	.40	.09	-.07	.40
unfa	-.33	-.00	.45	-.29	.09	.45
brus	.01	-.13	.51	.08	-.06	.51
feed	-.01	-.13	.47	.07	-.07	.47
inju	-.13	-.12	.31	-.08	-.07	.30
bath	.35	-.06	.03	.34	.03	-.01
wary	.61	.05	-.02	.62	-.01	.01
bshy	.60	.09	.01	.60	.03	.05
fret	.68	.09	-.07	.67	.02	-.03
slee	.35	.33	-.07	.30	.31	-.01
fubs	.09	.55	-.05	.00	.57	.03
fuas	.10	.47	-.07	.02	.48	-.00
crys	.12	.51	-.18	.02	.50	-.11

NOTE: Factor loadings for a 3-factor model of sweep 1 child noncognitive skills, as rotated to optimize the varimax and oblimin criteria respectively.

the measurement did not represent the underlying factor, and so we would omit it from that factor and either test its suitability for an alternative factor, include it independently, or drop it from the dataset. If two responses from a particular measurement had a similar distribution across factor scores, then we would infer that those two responses carried similar information about that latent factor, and we would consider combining those response categories.

Chapter References

Benhabib, Jess, Alberto Bisin & Andrew Schotter (2010). “Present-bias, quasi-hyperbolic discounting, and fixed costs”. In: *Games and Economic Behavior* 69.2, pp. 205–223.

Cattel, R B (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum Press.

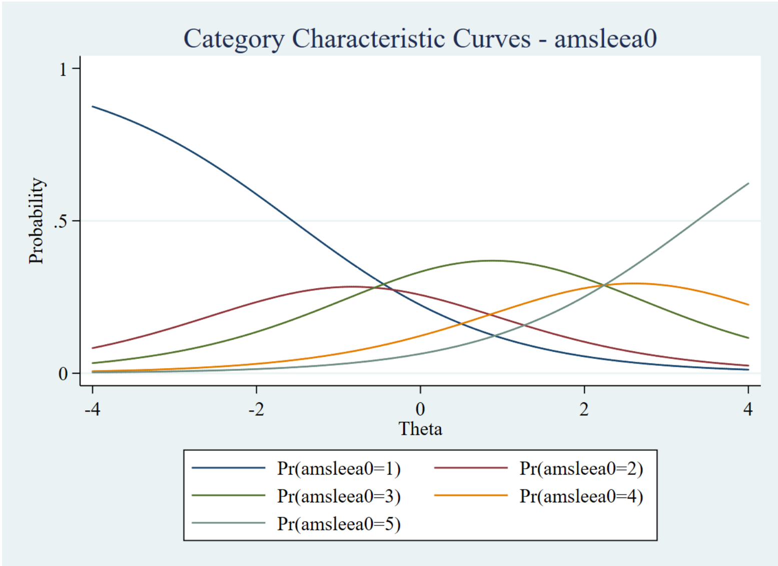
Cunha, Flávio (2011). “Recent developments in the identification and estimation of production functions of skills”. In: *Fiscal Studies* 32.2, pp. 297–316.

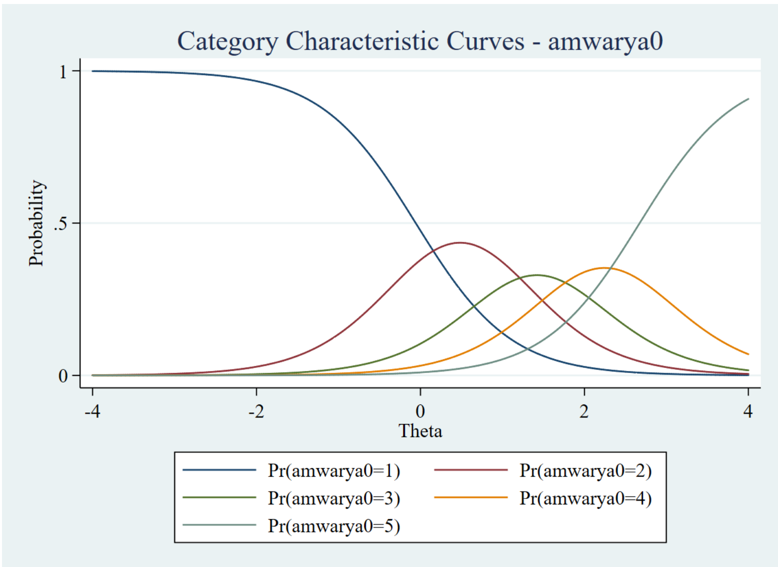
Cunha, Flávio, James J Heckman & Susanne M Schennach (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Econometrica* 78.3, pp. 883–931.

Dinno, Alexis (2009). “Implementing horn’s parallel analysis for principal component analysis and factor analysis”. In: *Stata Journal* 9.2, pp. 291–298.

Filippin, Antonio & Marco Paccagnella (2012). “Family background, self-confidence and economic outcomes”. In: *Economics of Education Review* 31.5, pp. 824–834.

Figure A.17: *Representative IRT plots for s1 child noncognitive skill items*





Frederick, Shane William, George Loewenstein & Ted O'Donoghue (2002). "Time Discounting and Time Preference: A Critical Review". In: *Journal of Economic Literature* XL.June, pp. 351–401.

Horn, John L (1965). "A Rationale and Test for the Number of Factors in Factor Analysis". In: *Psychometrika* 30.2, pp. 179–185.

Loewenstein, George (1996). "Out of control: Visceral influences on behavior". In: *Organizational Behavior and Human Decision Processes* 65.3, pp. 272–292.

Consolidated References

Agostinelli, Francesco & Matthew Wiswall (2016a). “Estimating the Technology of Children’s Skill Formation”. In: NBER Working Paper, No. 22442. Cambridge.

— (2016b). “Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development”. In: NBER Working paper, No. 22441. Cambridge.

Akerlof, George A & Rachel E Kranton (2002). “Identity and schooling: Some Lessons from the Economics of Education”. In: *Journal of Economic Literature* 40.4, pp. 1167–1201.

Almås, Ingvild et al. (2016). “What explains the gender gap in college track dropout? Experimental and administrative evidence”. In: *American Economic Review* 106.5, pp. 296–302.

Almlund, Mathilde et al. (2011). “Personality Psychology and Economics”. In: *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Stephen Machin & Ludger Woessmann. 1st ed. Vol. Volume 4. Amsterdam: North-Holland. Chap. 1, pp. 1–181.

Arthur, W. Brian (1989). “Competing Technologies, Increasing Returns, and Lock-In by Historical Events”. In: *The Economic Journal* 99.394, pp. 116–131.

Attanasio, Orazio, Costas Meghir & Emily Nix (2015). “Human Capital Development and Parental Investment in India”. In: Economic Growth Center Discussion Paper, No. 1052. New Haven.

Attanasio, Orazio et al. (2015). “Estimating the production function for human capital: Results from a randomized control trial in Colombia”. In: National Bureau of Economic Research working papers, no. 20965.

Attanasio, Orazio et al. (2017). “Human capital growth and poverty: Evidence from Ethiopia and Peru”. In: *Review of Economic Dynamics* 25, pp. 234–259.

Bandura, Albert & Dale H. Schunk (1981). “Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation”. In: *Journal of Personality and Social Psychology* 41.3, pp. 586–598.

Becker, Anke et al. (2012). “The Relationship between Economic Preferences and Psychological Personality Measures”. In: *Annual Review of Economics* 4, pp. 453–478.

Becker, Gary S (1962). “Investment in Human Capital: A Theoretical Analysis”. In: *Journal of Political Economy* 70.5, pp. 9–49.

— (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* New York: National Bureau of Economic Research.

Belzil, Christian (2007). “Subjective beliefs and schooling decisions”. In: GATE Working Papers, No. 2007-17 07-17. Écully: Groupe d’Analyse et de Théorie Économique.

Bénabou, Roland & Jean Tirole (2002). “Self-Confidence and Personal Motivation”. In: *The Quarterly Journal of Economics* 117.3, pp. 871–915.

Bénabou, Roland & Jean Tiróle (2016). “Mindful Economics: The Production, Consumption, and Value of Beliefs”. In: *Journal of Economic Perspectives* 30.3, pp. 141–164.

Benhabib, Jess, Alberto Bisin & Andrew Schotter (2010). “Present-bias, quasi-hyperbolic discounting, and fixed costs”. In: *Games and Economic Behavior* 69.2, pp. 205–223.

Black, Paul & Dylan Wiliam (1998). “Inside the Black Box: Raising Standards Through Classroom Assessment”. In: *Phi Delta Kappan* 80.2, pp. 139–148.

Blackwell, Lisa S, Kali H Trzesniewski & Carol S Dweck (2007). “Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention in Child”. In: *Child Development* 78.1, pp. 246–263.

Bradley, Steve (Forthcoming). “The effects of Quasi-Markets in Education”. In: *Economics of Education,* ed. by Steve Bradley & Colin Green. 2nd. Amsterdam: Elsevier.

Brañas-Garza, Pablo, Praveen Kujal & Balint Lenkei (2019). “Cognitive reflection test: Whom, how, when”. In: *Journal of Behavioral and Experimental Economics* 82.March, p. 101455.

Bridgeland, John M, John J Jr Dilulio & Karen Burke Morison (2006). *The Silent Epidemic: Perspectives of High School Dropouts.* Tech. rep. Washington: Civic Enterprises & Peter D. Hart Research Associates.

Burks, Stephen V et al. (2015). “Cognitive skills, personality, and economic preferences in collegiate success”. In: *Journal of Economic Behavior and Organization* 115, pp. 30–44.

Cahuc, Pierre, André Zylberberg & Stéphane Carcillo (2014). *Labour Economics.* 2nd. Cambridge: MIT press.

Carneiro, Pedro & James J Heckman (2002). “The Evidence on Credit Constraints in Post- Secondary Schooling”. In: *The Economic Journal* 112.April 2001, pp. 705–734.

Casselman, Brock L. & Charles H. Atwood (2017). “Improving General Chemistry Course Performance through Online Homework-Based Metacognitive Training”. In: *Journal of Chemical Education*, acs.jchemed.7b00298.

Cattel, R B (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum Press.

Christenson, Sandra L, Amy L Reschly & Cathy Wylie (2012). *Handbook of research on student engagement.* New York: Springer.

Coneus, Katja, Manfred Laucht & Karsten Reuß (2012). “The role of parental investments for cognitive and noncognitive skill formation - Evidence for the first 11 years of life”. In: *Economics and Human Biology* 10.2, pp. 189–209.

Conti, Gabriella, James J Heckman & Sergio Urzua (2010). “The Education-Health Gradient”. In: *American Economic Review* 100.2, pp. 234–238.

Credé, Marcus, Michael C Tynan & Peter D Harms (2017). “Much Ado About Grit: A Meta-Analytic Synthesis of the Grit Literature”. In: *Journal of Personality and Social Psychology* 113.3, pp. 492–511.

Cunha, Flávio (2011). “Recent developments in the identification and estimation of production functions of skills”. In: *Fiscal Studies* 32.2, pp. 297–316.

Cunha, Flávio & James J Heckman (2007). “The Technology of Skill Formation”. In: *The American Economic Review* 97.2, pp. 31–47.

— (2008). “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Journal of Human Resources* 43.4, pp. 738–782.

— (2010). “Investing in our Young People”. In: IZA discussion paper, No.5050. Bonn: Forschungsinstitut zur Zukunft der Arbeit.

Cunha, Flávio, James J Heckman & Susanne M Schennach (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Econometrica* 78.3, pp. 883–931.

De La Grandville, Olivier (1989). “In Quest of the Slutsky Diamond”. In: *The American Economic Review* 79.3, pp. 468–481.

Del Boca, Daniela, Chiara Monfardini & Cheti Nicoletti (2017). “Parental and Child Time Investments and the Cognitive Development of Adolescents”. In: *Journal of Labor Economics* 35.2, pp. 565–608.

Dignath, Charlotte & Gerhard Büttner (2008). “Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level”. In: *Metacognition and Learning* 3.3, pp. 231–264.

Dinno, Alexis (2009). “Implementing horn’s parallel analysis for principal component analysis and factor analysis”. In: *Stata Journal* 9.2, pp. 291–298.

Dohmen, Thomas et al. (2011). "Individual risk attitudes: Measurement, determinants, and behavioral consequences". In: *Journal of the European Economic Association* 9.3, pp. 522–550.

Duckworth, Angela L et al. (2007). "Grit: Perseverance and passion for long-term goals." In: *Journal of Personality and Social Psychology* 92.6, pp. 1087–1101.

Edlin, Aaron S & Chris Shannon (1998). "Strict Single Crossing and the Strict Spence-Mirrlees Condition: A Comment on Monotone Comparative Statics". In: *Econometrica* 66.6, pp. 1417–1425.

Embretson, S E & S P Reise (2000). *Item Response Theory for Psychologists.* Mahwah: Lawrence Erlbaum Associates.

Embrey, Iain Peter (2019a). "A Review of Economic Dual-Self Modelling". Lancaster.

— (2019b). "Education as a Participation Game: An Explanation for Educational Under- Investment". Lancaster.

— (2019c). "On the Benefits of Normalization in Production Functions". Lancaster.

— (2019d). "Re-Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Under Review.*

— (2019e). "States of nature and states of mind: a generalized theory of decision-making". In: *Theory and Decision.*

— (2019f). "The Education Trap: Could a grades-focussed educational system be perpetuating poverty in advanced economies?" In: *Under Review.*

Emmer, Edmund T, Carolyn M Everston & Linda M Anderson (1980). "Effective Classroom Management at the Beginning of the School Year". In: *The Elementary School Journal* 80.5, pp. 219–231.

Figlio, David N. & Cecilia Elena Rouse (2006). "Do accountability and voucher threats improve low-performing schools?" In: *Journal of Public Economics* 90.1-2, pp. 239–255.

Filippin, Antonio & Marco Paccagnella (2012). "Family background, self-confidence and economic outcomes". In: *Economics of Education Review* 31.5, pp. 824–834.

Fiorini, Mario & Michael P Keane (2014). "How the Allocation of Children's Time Affects Cognitive and Noncognitive Development". In: *Journal of Labor Economics* 32.4, pp. 787–836.

Francesconi, Marco & James J Heckman (2016). "Child Development and Parental Investment: Introduction". In: *Economic Journal* 126.596, F1–F27.

Frederick, Shane (2005). "Cognitive Reflection and Decision Making". In: *Journal of Economic Perspectives* 19.4, pp. 25–42.

Frederick, Shane William, George Loewenstein & Ted O'Donoghue (2002). "Time Discounting and Time Preference: A Critical Review". In: *Journal of Economic Literature* XL.June, pp. 351–401.

Griliches, Zvi (1977). "Estimating the Returns to Schooling: Some Econometric Problems". In: *Econometrica* 45.1, pp. 1–22.

Harmon, Colm, Hessel Oosterbeek & Ian Walker (2000). *The Returns to Education: A Review of Evidence, Issues and Deficiencies in the Literature*. December. London: Centre for the Economics of Education, pp. 1–51.

Heckman, James J (2006). "Skill Formation and the Economics of Investing in Disadvantaged Children". In: *Science* 312.5782, pp. 2005–2007.

Heckman, James J, John Eric Humphries & Gregory Veramendi (2018). "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking". In: *Journal of Political Economy* 126.S1, S197–S246.

Heckman, James J & Stefano Mosso (2014). "The Economics of Human Development and Social Mobility". In: *Ann. Rev. Econ.* 6.2006, pp. 689–733.

Heckman, James J & Rodrigo Pinto (2015). "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs". In: *Econometric Reviews* 34, pp. 6–31.

Heckman, James J, Rodrigo Pinto & Peter Savelyev (2013). "Understanding the Mechanisms Through which an Influential Early Childhood Program Boosted Adult Outcomes". In: *American Economic Review* 103.6, pp. 2052–2086.

Heckman, James J & Y Rubinstein (2001). "The importace of noncognitive skills: lessons from the GED testing". In: *the American Economic Review* 92.2, pp. 145–149.

Heineck, Guido & Silke Anger (2010). "The returns to cognitive abilities and personality traits in Germany". In: *Labour Economics* 17.3, pp. 535–546.

Helmers, Christian & Manasa Patnam (2011). "The formation and evolution of childhood skill acquisition: Evidence from India". In: *Journal of Development Economics* 95.2, pp. 252–266.

Heredia, Diana Jimena Martinez (2017). "El rol de la inversión en la función de producción de habilidades cognitivas: evidencia de Colombia, Chile y Perú". In: Serie Documentos Cede, No. 2017-12. Bogotá.

Hernández-Alava, Mónica & Gurleen Popli (2017). "Children's Development and Parental Input: Evidence From the UK Millennium Cohort Study". In: *Demography* 54.2, pp. 485–511.

Hicks, John R (1932). *The Theory of Wages*. London: MacMillan and Co. Ltd.

Hills, J., J. Le Grand & D. Piachaud (2002). *Understanding Social Exclusion*. Oxford: Oxford University Press.

Hobcraft, John (2002). “Social Exclusion and the Generations”. In: *Understanding Social Exclusion*. Ed. by J. Hills, J. Le Grand & D. Piachaud. Oxford: Oxford University Press. Chap. 5, pp. 62–83.

Horn, John L (1965). “A Rationale and Test for the Number of Factors in Factor Analysis”. In: *Psychometrika* 30.2, pp. 179–185.

Humphries, John Eric & Fabian Kosse (2017). “On the interpretation of non-cognitive skills - What is being measured and why it matters”. In: *Journal of Economic Behavior and Organization* 136, pp. 174–185.

Jensen, Robert (2010). “The (Percieved) Returns to Education and the Demand for Schooling”. In: *The Quarterly Journal of Economics* 125.2, pp. 515–548.

Koch, Alexander, Julia Nafziger & Helena Skyt Nielsen (2015). “Behavioral Economics of Education”. In: *Journal of Economic Behavior & Organization* 115.1, pp. 3–17.

Köszegi, Botond (2006). “Ego Utility , Overconfidence , and Task Choice”. In: *Journal of the European Economic Association* 4.4, pp. 673–707.

Laibson, David (1997). “Golden Eggs and Hyperbolic Discounting”. In: *The Quarterly Journal of Economics* 112.2, pp. 443–477.

Lavecchia, Adam M, Heidi Liu & Philip Oreopoulos (2016). “Behavioral Economics of Education: Progress and Possibilities”. In: *The Handbook of the Economics of Education,* ed. by Eric A Hanushek, Stephen Machin & Ludger Woessmann. Vol. 5. Amsterdam: North-Holland. Chap. 1, pp. 1–74.

Loewenstein, George (1996). “Out of control: Visceral influences on behavior”. In: *Organizational Behavior and Human Decision Processes* 65.3, pp. 272–292.

MacLeod, W. Bentley (2016). “Human capital: Linking behavior to rational choice via dual process theory”. In: *Labour Economics* 41, pp. 20–31.

McCrae, Robert R & Paul T Jr Costa (1985). “Updating Norman’ s “ Adequate Taxonomy ”: Intelligence and Personality Dimensions in Natural Language and Questionnaires”. In: *Journal of Personality and Social Psychology* 49.3, pp. 710–721.

McMahon, Walter W (2004). “The Social and External Benefits of Education”. In: *International Handbook on the Economics of Education*. Ed. by Geraint Johnes & Jill Johnes. Cheltenham: Edward Elgar Publishing Ltd. Chap. 6, pp. 211–259.

Moffitt, T. E. et al. (2011). “A gradient of childhood self-control predicts health, wealth, and public safety”. In: *Proceedings of the National Academy of Sciences* 108.7, pp. 2693–2698.

Mueller, Claudia M & Carol S Dweck (1998). “Praise for Intelligence Can Undermine Children’s Motivation and Performance”. In: *Journal of Personality and Social Psychology* 75.1, pp. 33–52.

Nicoletti, Cheti & Birgitta Rabe (2014). “School inputs and skills: Complementarity and self-productivity”. In: IZA Discussion Paper Series, No. 8693. Bonn.

O’Donoghue, Ted & Matthew Rabin (1999). “Doing It Now or Later”. In: *The American Economic Review* 89.1, pp. 103–124.

Ockham, William of (1323). “Summa Totius Logicae”. In: *Summa Logicae, 1974*. Ed. by Ph Boehner, G Gal & S Brown. Olean: St. Bonaventure University.

OECD (2017). *Starting Strong 2017: Key OECD Indicators on Early Childhood Education and Care*. Paris: OECD Publishing, p. 192.

Oreopoulos, Philip (2007). “Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling”. In: *Journal of Public Economics* 91.11-12, pp. 2213–2229.

Oreopoulos, Philip & Kjell G Salvanes (2011). “Priceless: The Nonpecuniary Benefits of Schooling”. In: *Journal of Economic Perspectives* 25.1, pp. 159–184.

Pavan, Ronni (2016). “On the Production of Skills and the Birth-Order Effect”. In: *Journal of Human Resources* 51.3, pp. 699–726.

Pennycook, Gordon, Jonathan A. Fugelsang & Derek J. Koehler (2015). “Everyday Consequences of Analytic Thinking”. In: *Current Directions in Psychological Science* 24.6, pp. 425–432.

Pennycook, Gordon & Robert M. Ross (2016). “Commentary: Cognitive reflection vs. calculation in decision making”. In: *Frontiers in Psychology* 7.JAN, pp. 2014–2017.

Pennycook, Gordon et al. (2016). “Is the cognitive reflection test a measure of both reflection and intuition?” In: *Behavior Research Methods* 48.1, pp. 341–348.

Pissarides, Christopher A (1990). *Equilibrium Unemployment Theory*. 1st. Oxford: Blackwell Publishers Ltd.

— (2000). *Equilibrium Unemployment Theory*. 2nd. Cambridge: MIT press.

Plato (c.380 B.C.E.). “The Republic”. In: *The Republic of Plato, 1906*. Ed. by A.D. Lindsay & R. Bambrough. London: J.M. Dent & Sons Ltd.

Reuß, Karsten (2011). “Determinants of personality and skill development in the Socio-emotional environment during childhood”. In: MPRA Paper No. 82818. Munich.

Rouse, Cecilia Elena (2004). “Low-income students and college attendance: An exploration of income expectations”. In: *Social Science Quarterly* 85.5 SPEC. ISS. Pp. 1299–1317.

Rustichini, Aldo et al. (2016). “Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation”. In: *Journal of Behavioral and Experimental Economics* 64, pp. 122–137.

Samuelson, Paul A. (1938). “A Note on the Pure Theory of Consumers’ Behaviour”. In: *Economica* 5.17, pp. 61–71.

Samuelson, Paul Anthony & William Dawbney Nordhaus (2009). *Economics*. 19th ed. New York: McGraw-Hill Irwin, p. 702.

Savage, Leonard J (1954). *The Foundations of Statistics*. New York: John Wiley & Sons Inc.

Segal, Carmit (2013). "Misbehavior, Education, and Labor Market Outcomes". In: *Journal of the European Economic Association* 11.4, pp. 743–779.

Shoda, Yuichi, Walter Mischel & Philip K. Peake (1990). "Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions". In: *Developmental Psychology* 26.6, pp. 978–986.

Sjögren, Anna & Susanna Sällström (2004). "Trapped, Delayed and Handicapped". In: *CEPR Discussion Paper Series*. CEPR Discussion Paper Series, No. 613. London: Centre for Economic Policy Research.

Smith, Adam (1759). *The Theory of Moral Sentiments*. London: A. Millar.

Sparkes, Jo & Howard Glennerster (2002). "Preventing Social Exclusion: Education's Contribution". In: *Understanding Social Exclusion*. Ed. by J. Hills, J. Le Grand & D. Piachaud. Oxford: Oxford University Press. Chap. 11, pp. 178–201.

Spence, Michael (1973). "Job Market Signaling". In: *The Quarterly Journal of Economics* 87.3, p. 355.

Tanner, Kimberly D. (2012). "Promoting student metacognition." In: *CBE life sciences education* 11.2, pp. 113–120.

Thaler, Richard H. & Hersh M. Shefrin (1981). "An Economic Theory of Self-Control". In: *Journal of Political Economy* 89.2, pp. 392–406.

Thiel, Hendrik & Stephan L. Thomsen (2013). "Noncognitive skills in economics: Models, measurement, and empirical evidence". In: *Research in Economics* 67.2, pp. 189–214.

Wang, X. Henry & Bill Z. Yang (2003). "Why Competition may Discourage Students from Learning? A Behavioral Economic Analysis". In: *Education Economics* 11.2, pp. 117–128.

West, Martin R. & Paul E. Peterson (2006). "The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments". In: *Economic Journal* 116.510, pp. 46–62.

Zepeda, Cristina D. et al. (2015). "Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study". In: *Journal of Educational Psychology* 107.4, pp. 954–970.