

# EDUCATIONAL CHOICES AND SOCIAL INTERACTIONS: A FORMAL MODEL AND A COMPUTATIONAL TEST

Gianluca Manzo

## ABSTRACT

*In their authoritative literature review, Breen and Jonsson (2005) claim that 'one of the most significant trends in the study of inequalities in educational attainment in the past decade has been the resurgence of rational-choice models focusing on educational decision making'. The starting point of the present contribution is that these models have largely ignored the explanatory relevance of social interactions. To remedy this shortcoming, this paper introduces a micro-founded formal model of the macro-level structure of educational inequality, which frames educational choices as the result of both subjective ability/benefit evaluations and peer-group pressures. As acknowledged by Durlauf (2002, 2006) and Akerlof (1997), however, while the social psychology and ethnographic literature provides abundant empirical evidence of the explanatory relevance of social interactions, statistical evidence on their causal effect is still flawed by identification and selection bias problems. To assess the relative explanatory contribution of the micro-level and network-based mechanisms hypothesised, the paper opts for agent-based computational*

---

**Class and Stratification Analysis**

**Comparative Social Research, Volume 30, 47–100**

**Copyright © 2013 by Emerald Group Publishing Limited**

**All rights of reproduction in any form reserved**

**ISSN: 0195-6310/doi:10.1108/S0195-6310(2013)0000030007**

*simulations. In particular, the technique is used to deduce the macro-level consequences of each mechanism (sequentially introduced) and to test these consequences against French aggregate individual-level survey data. The paper's main result is that ability and subjective perceptions of education benefits, no matter how intensely differentiated across agent groups, are not sufficient on their own to generate the actual stratification of educational choices across educational backgrounds existing in France at the beginning of the twenty-first century. By computational counterfactual manipulations, the paper proves that network-based interdependencies among educational choices are instead necessary, and that they contribute, over and above the differentiation of ability and of benefit perceptions, to the genesis of educational stratification by amplifying the segregation of the educational choices that agents make on the basis of purely private ability/benefit calculations.*

**Keywords:** Intergenerational educational mobility; educational choices; social interactions; social networks; agent-based models; analytical sociology

## INTRODUCTION

By themselves, quantities just aid assessments, but when parameters marshal them into an array of numbers, then they can trigger alternative thinking. (White, 2000, p. 505)

About 15 years ago, Goldthorpe (1996) and Breen and Goldthorpe (1997) considered the large temporal stability of class differentials in educational attainment as one of the *explananda* that 'pose(s) an evident theoretical challenge'. While recent comparative empirical analyses (see Breen, Luijkx, Müller, & Pollak, 2009; Breen, Luijkx, Muller, & Pollak, 2010) have demonstrated that the change-resistance of inequality of educational opportunity is less pronounced than previous studies suggested (see, mainly, Shavit & Blossfeld, 1993), it is indisputable that, despite the generalised improvement of living standards and the variety of educational policies aimed at counteracting the partly social constructed ability gaps across social groups, the socioeconomic status of the family in which individuals live still exerts a substantial influence on their educational outcomes. In France, the country on which this paper focuses, while the equalising trend in educational opportunity is extremely clear (see Thélot & Vallet, 2000), although quantitatively modest (see Vallet & Selz, 2007, p. 69), individuals'

social backgrounds continue pervasively to impact on the highest educational level that they reach (see Selz & Vallet, 2006), on the kind of upper secondary tracks that they choose (see Duru-Bellat, Kieffer, & Reimer, 2011; Ichou & Vallet, 2011) as well as on their chances of entering the most prestigious tertiary level educational institutions (see Albouy & Wanecq, 2003). In light of this empirical evidence, therefore, it still seems justified to consider the stratification of educational outcomes across social backgrounds as a puzzling macro-level *explanandum*.

In order to advance understanding of why the socioeconomic status of the family in which individuals live continues to exert such a pervasive influence on their educational choices, this paper introduces a new formal model of how individuals' educational preferences form. On a theoretical level, the model's novelty is that it frames educational preference formation as the result of both individual ability/benefit evaluations and peer-group pressures, thus cross-fertilising the sociological rational-choice approach to educational choices (Breen & Goldthorpe, 1997) with recent heterodox theoretical perspectives in economics regarding inequality (see, in particular, Durlauf, 1999a, 2002, 2006) and education (see, in particular, Akerlof, 1997; Akerlof & Kranton, 2002). On a methodological level, the paper's originality resides in its deduction of the macro-level consequences of the formal model by means of agent-based computational simulations (see Shoam & Leyton-Brown, 2009; Wooldridge, 2009). These computational regularities will, however, be systematically compared to French empirical survey data. An interface will thus be created between the regression-based approach that dominates quantitative studies of educational stratification and new advances in simulation methodology. Both the theoretical and the methodological development proposed can be justified in light of the current state of the literature on the explanation of educational stratification.

Since Boudon's (1974, see, in particular, pp. 29–31) pioneering study, rational-choice oriented explanations of the aggregate association between individuals' social backgrounds and their educational achievements have been regularly refined (see, in particular, Breen & Goldthorpe, 1997; Gambetta, 1987; Goldthorpe, 1996; Jonsson & Erikson, 2000). According to this analytical schema, class differentials in educational attainment arise from the composition of the strategies of rational actors who, as they proceed through the educational system, systematically evaluate their cognitive abilities, their probability of success (as a function of their ability), the benefits that they can obtain from education, and the direct and indirect education costs. The basic assumption is that the higher the social status of the actor's family, the higher the actor's ability; consequently, the

better the actor's perception of his/her probability of success, the higher his/her educational aspiration, and the lower the perceived costs of education. As a consequence, the members of higher social groups should succeed at a higher rate at each educational transition, thereby ending up with higher educational credentials.

However, after around two decades of statistical-based empirical tests (see, among others, Ballarino & Bernardi, 2001; Becker, 2003; Breen & Yaish, 2006; Cobalti, 1992; Davies, Heinesen, & Holm, 2002; Gabay-Egozi, Shavit, & Yaish, 2010; Hillmert & Jacob, 2003; Holm & Jaeger, 2008; Manzo, 2006; Mastekaasa, 2006; Need & de Jong, 2000; Raftery & Hout, 1993; Schizzerotto, 1997; Stocké, 2007; Van de Werfhorst & Hofstede, 2007), while several pieces of the theoretical framework have received empirical support – such as the systematic variation of ability, of perception of success probability and of educational aspirations across social groups (see, for instance, Becker, 2003; Need & de Jong, 2000; Stocké, 2007) – still unclear is the extent to which these micro-level factors really impact on individuals' educational decisions (see, for instance, Stocké, 2007; Van de Werfhorst & Hofstede, 2007; Gabay-Egozi et al., 2010). As recently acknowledged by Kronenberg and Kalter (2012), when one inspects the published empirical results, one of the most striking findings is that no matter what indicator of ability and perceived benefits and costs is adopted, the effect of social background on educational choices and attainment is still substantial. This is especially puzzling at the highest educational transitions, where student heterogeneity is lower because of selection at previous transitions (see, for instance, Mastekaasa, 2006).

While this might be related to inappropriate statistical model specifications and/or poor measures of the theoretical constructs of interest, one might also think, more radically, that the rational-choice approach to educational stratification does not take account of certain mechanisms which potentially mediate the link between social background and educational choices/outcomes. On this reasoning, a few empirical analyses have started to study two types of mechanism under-conceptualised within this research tradition: genetic heritability (see Lucchini, Della Stella, & Pisati, 2010) and social interactions in the form of peer effects (see, in particular, Jaeger, 2007; Morgan, 2005).

As announced, this paper focuses on social interactions. In particular, it computationally tests the hypothesis that the (French) empirical stratification of educational choices cannot be generated without assuming that homophilic dyadic interactions taking place within friendship networks sustain self-reinforcing mimetic educational behaviours that progressively

accentuate the existing social differentiation in ability and perception of education benefits.

This accounts for the methodology chosen in what follows to prove this statement, that is, agent-based computational simulations. The empirical quantification of the net effect of neighbourhood- and network-based social influences on individual outcomes has proved extremely difficult (see, respectively, [Mouw, 2006](#); [Sobel, 2006](#); [Sampson, Morenoff, & Gannon-Rowley, 2002](#); [Shalizi & Thomas, 2011](#); and [VanderWeele, 2011](#)). If two actors are related by a friendship tie and end up with a similar (educational) outcome, then this correlation may actually arise from three distinct phenomena: (1) the two actors are exposed to certain common factors, like school or teaching quality (see the concept of ‘ecological effect’; [Manski, 1993b](#), p. 31); (2) the two actors share some common social background characteristics, like parental income (see the concept of ‘contextual effect’; [Manski, 1993a](#), p. 532, [1993b](#), p. 31); (3) the two actors share some common individual characteristics, like an interest in school (see the concept of ‘correlated effect’; [Manski, 1993a](#), p. 533, [1993b](#), p. 31). If so, the similar (educational) outcome may arise not from the influence that the two actors exert on each other, which would constitute the interaction-based ‘endogenous effect’ in which one is interested, but from the potentially unmeasured shared factors that modify the probability of being friends and that of experiencing a certain (educational) outcome. Unfortunately, unless very restrictive conditions are introduced, the four effects cannot be empirically distinguished on the basis of nonexperimental observational data (for an overview, see [Durlauf & Loannides, 2010](#)), and, because of the pervasive problem of unmeasured heterogeneity, empirical estimations tend to be ‘biased’ ([Harding, Gennetian, Winship, Sanbonmatsu, & Kling, 2011](#)). Moreover, an additional complication concerns the direction of the causality. Does the average (educational) outcome among a given subset of friends truly affect their (educational) behaviour, or does the group-level factor simply reflect this behaviour? Again, as Manski’s pioneering contribution demonstrated, there is no simple econometric solution to ‘reflection problem’ – which becomes all the more difficult to solve when the aim is to estimate the several potential sources of the overall researched ‘endogenous effect’ (see [Cohen-Cole & Zanella, 2008](#)).

Agent-based computational modelling cannot help solve these estimation and identification problems on empirical grounds. At the theoretical level, however, they offer a unique opportunity for rigorous study of formal models containing neighbourhood- and/or network-based interdependences among individual behaviours. First of all, there is no unobserved

heterogeneity within an agent-based model. By construction, since the modeller defines them, all group- and individual-level variables defining agents' attributes are perfectly known. Similarly, because agents' spatial and network locations are defined by the modeller, the way in which agents are linked together and the composition of their local neighbourhood is also completely transparent. 'Ecological', 'contextual' and 'correlated' effects are thus completely controlled for. Moreover, they can be easily separated from 'endogenous' effects. By means of appropriate procedures to control for model stochasticity, it is in fact possible to re-run exactly the same simulation with and without the network-based social influence mechanisms (this counterfactual manipulation will be performed in section 'Computational results, empirical data and model dynamic'). The net effect of being in contact with a given set of agents *over and above* the effect associated with the distribution of individual-level characteristics can thus be isolated. Finally, the 'reflection' problem can be efficiently handled. Within an agent-based model, the sequence of events, the model scheduling, is defined by the modeller so that, although sometimes complicated, it is possible to establish what causes what. There is an additional attractiveness in using agent-based simulations to study the effect of social interaction on individuals' behaviour. As acknowledged by Goux and Maurin (2007), empirical data usually do not allow for study of individuals' closest neighbours, which represent the neighbourhoods that really matter for individuals' outcomes. Existing studies typically proxy interaction-based mechanisms with school-and/or neighbourhood-level aggregate variables (among analyses specifically addressing individual educational outcomes, see, for instance, Agirgad, 2011; Brännström, 2008; Fekjær & Birkelund, 2007). Agent-based models make it possible to overcome this limitation by enabling the creation of any sort of actor-to-actor network topology.

For these reasons, given the aim of the paper, that is, to test the hypothesis that the actual (French) distribution of educational choices across social groups cannot be generated unless one assumes that some network-based social influence mechanism is at work, agent-based modelling represents the best methodological choice. Instead of inferring the effect of social interactions from observation of their outcomes (for this criticism within economics, see Manski, 2000), agent-based computation simulations will enable me deductively to generate aggregate structures of numerical data that can be compared to empirical data structures, *given* an entirely specified set of behavioural rules and interacting patterns – the formal model proposed provides this set of elements. On the other hand, the counterfactual possibilities offered by the technique will make it possible to isolate

the specific contribution of the interaction-based mechanism postulated. The paper thus contains a ‘computational test’ of the formal model proposed, in that it proves, by growing educational stratification *in silico*, the conditions under which the actual (French) educational stratification can be best approximated (on the deductive nature of agent-based models, a point often poorly understood, see Epstein, 2006, ch. 1, pp. 10–12; on the concept of ‘constructive proofs’, see Borrill & Tesfatsion, 2010).

The article is organised as follows. The first section provides a description of the empirical data and sets out the research strategy underlying the paper. Second section presents the formal model aimed at explaining the empirical observations – a generalisation of this model is briefly discussed in Appendix B. Third section reports the simulation results and compares them to the empirical data. Fourth section evaluates the robustness of the simulation results concerning the network-based social mechanism postulated and comments on some counterintuitive results generated by the model with respect to the link between ‘weak ties’ and educational inequality. Fifth section discusses the major limitations of the analysis, and is followed by a short general conclusion.<sup>1</sup>

## DATA, VARIABLES AND RESEARCH STRATEGY

The empirical data on which I draw come from a large national survey, representative of the French population aged 18–65, carried out in 2003 by the INSEE, the French national bureau of statistics. In order to ensure that all individuals have attained their highest educational level, I consider here only male and female respondents aged 27–65 at the time of the interview for whom complete information on their own and at least one of their parents’ education achievements was available.

As regards variables, the dependent variable on which the paper focuses is the respondents’ highest educational level, whereas the independent variable is the highest educational level attained by respondents’ parents, which is measured as the highest educational level between the respondent’s father and mother (for the ‘dominance principle’ applied here, see Erikson & Goldthorpe, 1992, p. 238). In order to make the analysis more directly readable within the international literature, both variables have been recoded into the Casmin educational schema (see Brauns & Steinman, 1997; Müller, Lüttinger, König, & Karle, 1989; Müller & Karle, 1993; and, more recently, Breen, 2004, pp. 14–16). However, in order not to excessively increase the number of groups of artificial agents to be represented in the

formal model, I have adopted the following five-category version of the original schema: (1a) inadequately completed general education; (1bc) elementary education; (2ab) lower-secondary education; (2c) upper-secondary education and (3ab) tertiary education (for a similar five-category classification, see [Breen et al., 2010](#)).

Concerning the method, as argued in the introduction, the paper combines statistics and simulation. More precisely, the analysis follows the five-step research strategy recently summarised by [Hedström and Bearman \(2009, p. 16\)](#) according to which: ‘(1) we start with a clearly delineated social fact that is to be explained; (2) we formulate different hypotheses about relevant micro-level mechanisms; (3) we translate the theoretical hypotheses into computational models; (4) we simulate the models to derive the type of social facts that each micro-level mechanism brings about; (5) we compare the social facts generated by each model with the actually observed outcomes’ (for a similar, but three-step based, strategy, see also [Goldthorpe, 2001, p. 10](#); for an overall critical assessment of these developments, see [Manzo, 2007, 2010](#)).

In the following analysis, given the way in which independent and dependent variables are coded, the ‘delineated social fact to be explained’ consists of the aggregate patterns contained in Table 1, which cross-classifies the highest educational destinations attained by French men and women aged 27–65 in 2003 (columns) by the highest educational level reached by their parents (rows).

In order to describe both the absolute and the relative dimension of these patterns – that is to say, the amount of education that the members of a given educational background obtain and, on the other hand, their relative position, that is, their educational opportunity, within the educational hierarchy (see [Breen, 2004, ch. 2](#)) – the following statistics have been computed, the first three referring to the absolute aspect, the others to the relative one: the percentage of cases who are educationally immobile (‘ip’); the percentage of cases who are upwardly mobile (‘ump’); the ratio between the percentage of cases attaining the highest educational level within the highest and the lowest educational backgrounds (‘hlgr’); the generalised odds ratios of obtaining the highest educational level for the members of the highest educational background (‘hggor’); the generalised odds ratios of obtaining any educational credential for the members of the lowest educational background (‘lggor’) and the average of generalised odds ratios computed for each group (‘agor’).<sup>2</sup>

The statistical values reported in the bottom part of Table 1 thus show that educational inequality was substantial in France at the beginning of the



**Table 1.** Respondents' Highest Educational Level (Columns) by Their Parents' Highest Educational Level (Rows) – Row Percentages, Absolute Frequency (in Parenthesis) and Generalised Odds Ratios; France (2003), men and women aged 27–65.

	1a	1bc	2ab	2c	3ab	<i>N</i>
1a	44.22 (4,175) 8.98	29.00 (2,738) 4.02	15.48 (1,462) 0.81	5.61 (530) 0.27	5.69 (537) 0.13	(9,442)
1bc	16.23 (2,527) 1.22	28.03 (4,363) 2.47	26.98 (4,200) 1.25	13.93 (2,169) 0.73	14.83 (2,309) 0.36	(15,568)
2ab	9.51 (284) 0.58	11.38 (340) 0.66	26.01 (777) 1.30	20.96 (626) 1.52	32.14 (960) 1.33	(2,987)
2c	6.84 (147) 0.40	9.78 (210) 0.60	21.32 (458) 1.10	19.51 (419) 1.57	42.55 (914) 2.39	(2,148)
3ab	5.02 (169) 0.40	4.13 (139) 0.25	11.69 (394) 0.69	17.30 (583) 2.10	61.86 (2,084) 6.90	(3,369)
Absolute (1–3) and Relative (4–6) Educational Mobility Statistics						
	IP (1)	UMP (2)	HLGR (3)	AGOR (4)	HGGOR (5)	LGGOR (6)
	35.26	49.07	10.87	1.68	6.9	8.98

twenty-first century. More than one-third of the French respondents were indeed educationally immobile (see *ip*'s value); the offspring of the highest educational group reached the highest educational level about 10 times more frequently than did individuals from the lowest educational origin (see *hlgr*'s value); the overall competitive educational advantage for highest educational origin to obtain the highest educational level was about 7 (see *hggor*'s value), whereas the likelihood of respondents of lowest educational origin falling in the less desirable educational destination was about 9 (see *lggor*'s value), the two values thus express a quite intense polarisation of educational opportunity across social groups.

To explain these statistical facts, several hypotheses about the main explanatory factors that the literature on educational choices usually focuses on – that is, abilities, cost/preference trade-offs, subjective perceptions of success probability and peer-based social influences – will be formulated (Hedström and Bearman's step 2). These hypotheses about the relevant micro-level mechanisms will then be translated into a computational model (Hedström and Bearman's step 3) that will take the form of an agent-based model: that is to say, a computer program in which the decision of each numerical entity in the system about whether or not to make a given educational transition is driven by a set of entirely specified rules and interaction patterns (see section 'Computational modelling of educational preference formation'; Eqs. (1)–(4)). By iterating these rules until each agent reaches a stable educational level, the simulation of the model (Hedström and Bearman's step 4) makes it possible to trigger the process potentially associated with the hypothesised mechanisms so as to enable bottom-up deduction of the aggregate pattern that these mechanisms are able to generate (for the simulation algorithm, see section 'Computational modelling of educational preference formation'; Fig. 1). In this paper, the comparison of the simulated aggregate pattern with 'the actually observed fact' (Hedström and Bearman's step 5) takes the following form.

On the one hand, the formal model is evaluated on the basis of its capacity to reproduce the variety of features of the observed contingency table that I have just described, rather than a single aspect of the actual cross-tabulation (for a similar multi-statistics test, see, for instance, [Burke & Heiland, 2006](#), tab. 1). On the other hand, the formal model is evaluated on the basis of its capacity to reproduce a distribution of contingency tables rather than one single cross-tabulation. As often recommended, but less often done, distributions of outcomes matter more than single outcomes when assessing the explanatory and predictive power of a stochastic simulation model (see, for instance, [Stonedahl & Wilensky, 2010](#)).

To meet both requirements, I implemented the following procedure. As regards empirical data, I bootstrapped 100 samples of 5,000 cases from the original French sample and, for each sample, I computed the table, cross-classifying the highest educational destinations attained by French respondents with the highest educational level reached by their parents, and I computed the statistics described above on this cross-tabulation. As regards simulated data, for a given parameter structure, the simulation of the formal model was replicated 100 times and, for each replication, a simulated cross-tabulation was created. On each of them, I then computed the same statistics as computed on the empirical data. Thus, the comparison between empirical and simulation data finally took the form of an assessment of how overlap the nonparametric 95% confidence interval computed over the empirical/simulated series of values (for this approach, see [Law, 2007](#), pp. 269–271; on bootstrapped percentile confidence intervals, see [Davison & Hinkley, 1997](#), ch. 5).

As advised by step 5 of Hedström and Bearman’s research strategy, the mechanisms formalised by the model presented in the next section will be introduced sequentially. For each model variant, the comparison strategy that I have just described will be applied so that it is possible to assess the extent to which each (combination of) mechanism(s) is able to reproduce the qualitative structure of the empirical contingency table. As argued in the introduction, the fundamental goal of this incremental procedure is to establish in a deductive manner whether or not the structure and the level of educational inequality observed in France across educational backgrounds can be generated without assuming that some interaction-related mechanisms are at work.<sup>3</sup>

## **COMPUTATIONAL MODELLING OF EDUCATIONAL PREFERENCE FORMATION**

To test the generative power of several mechanisms potentially underlying the macro-level regularities described in the previous section (see Table 1), a population of numerical entities (hereafter called ‘artificial agents’ or, simply, ‘agents’) was programmed to make four sequential binary choices. In the present thematic context, each choice represents a decision about whether or not to enrol in a given educational level  $L$  (indexed from 1, the first educational transition, to 4, the last one), the sequential nature of these levels requiring that an agent can only move to the next choice if the

previous choice was positive. Since the formal model is intended to explain the relation between the group to which an actor belongs and his/her educational outcome, the artificial agents assumed to mimic the real actors are exogenously attributed to one of five groups  $g$  (indexed from 1, the highest group, to 5, the lowest one). In the present thematic context, the group to which the agent belongs represents his/her educational background. Each artificial group contains as many agents as respondents belonging to a given group in the French empirical sample.

As expressed by Eq. (1), the choice that an agent  $i$  belonging to group  $g$  is supposed to perform at each transition is assumed to be a monotonically but nonlinear increasing probabilistic function of the strength of the agent's preference  $P_{igL}$  for educational level  $L$  (with  $c$  being the centre parameter of the logistic curve set here to 1.5), so that the larger is  $P_{igL}$ , the higher the probability of the agent choosing educational level  $L$  – if not, s/he is given the opportunity to evaluate it again (the maximum number of permitted trials being three, see Fig. 1, step 2.2). Framing each decision as a stochastic choice means that the mechanisms relating agents' group to agents' educational preference that I shall postulate are not assumed to give full account of how individuals' educational preferences form.<sup>4</sup>

$$\Pr_{ig}(L = 1|L-1 = 1) = \frac{\exp(P_{igl}-c)}{1 + \exp(P_{igL}-c)}, \text{ with } P_{igL} = \text{Eq. (2)} \quad (1)$$

The analytical core of the formal model is the formation of  $P_{igL}$ , that is, the preference of agent  $i$  belonging to group  $g$  for educational level  $L$ . As expressed by Eq. (2), this preference is assumed to depend additively on four basic elements: (1) the agent's ability ( $A$  term); (2) the agent's perception of the pay-offs from the educational level compared with the perceived costs of obtaining it ( $B$  term); (3) the agent's perception of this benefit/cost balance as a function of his/her ability ( $\phi$  ( $A$ ) term); (4) the social influence exerted on the agent by the educational decisions of the agents with which s/he is in contact ( $SI$  term) (see Appendix B for a possible generalisation of the model).

$$P_{igL} = A_{ig} + \phi(A) \times B_{ig} + SI_{igL}, \text{ with } \phi(A) = \text{Eq. [3]} \text{ and } SI_{igL} = \text{Eq. [4]} \quad (2)$$

Thus, formally, the way in which I have modelled the agent's educational behaviour belongs to the family of binary choice models with social interactions as deeply analysed by Durlauf (1999b, 2001; for a less formal review, see also Rolfe, 2009). The main difference concerns the absence, in my formalisation, of any maximising functions at the agent level, which makes the model analytically less tractable but increases its realism.

Pioneering applications of this class of model can be found in Schelling (1971, 1973), Granovetter (1978), and Granovetter and Soong (1983, 1988).

### *The Ability Term ( $A$ )*

In psychology (see Herrnstein and Murray's (1994) controversial analysis; but, among others, see also Fischer et al.'s (1996) reaction), in economics (see Gintis, Bowles, & Osborne, 2001, 2002) and in sociology (see Sewell, Hauser, Springer, & Hauser, 2003), the social, biological-based differentiation among individuals' abilities has often been mobilised to account for the relationship between individuals' social backgrounds and their educational outcomes. Empirical evidence suggests that ability systematically varies across social groups, and that this differentiation arises and stabilises very early in an individual's cognitive development (see Duncan & Murnane, 2011, chs. 2–5). No matter how ability is measured, empirical studies show that it powerfully affects educational decisions (see, for instance, Breen & Yaish, 2006; Cheadle, 2008; Need & de Jong, 2000; Jonsson & Erikson, 2000, p. 350; Mastekaasa, 2006; Stocké, 2007, pp. 512, 515).

In the present model, agent's ability  $A_{ig}$  is assumed to represent any cognitive and personality traits that might matter during an individual's educational career. In order to take account of the above-mentioned empirical evidence about the social differentiation of ability, I follow Breen and Goldthorpe (1997) and assume that the distribution of agents' ability  $A_{ig}$  has common variance  $\sigma_A^2$  across groups of agents but different mean  $A_g$ . Different from Breen and Goldthorpe, however, here this distribution is taken to be log-normal rather than normal (for an overview of log-normal distribution, see Limpert, Stahel, & Abbt, 2001).<sup>5</sup> It is finally assumed that agents' ability is static across the four educational transitions that they must make within the artificial society (on the basis of the empirical evidence available, it is not easy to establish whether ability changes during schooling can reduce the ability gap between social groups: see, for instance, Falch & Massih, 2011; Winship & Korenman, 1997).

### *The Benefit Term ( $B$ )*

According to Eq. (2), the second main factor driving the formation of agents' educational preferences for a given educational level is  $B_{ig}$ , that is, the estimation that agent  $i$  belonging to group  $g$  makes the net benefit of the education investment.

Since Keller and Zavalloni's (1964) seminal article, rational-choice explanations of educational decisions have introduced the assumption that individuals' education evaluations depend on their relative positions on the social ladder in order to interpret the puzzling empirical observation that individuals from different social backgrounds tend to make different educational choices even when they have similar abilities (see Boudon, 1974, p. 28). Within this tradition, the quantitative analysis of educational inequalities is now framed in terms of primary ability-driven and secondary choice-driven effects (see Jackson, Erikson, Goldthorpe, & Yaish, 2007; Jackson, 2012). The latter are also supposed to depend on actors' perceptions of the costs of education, perceptions that are assumed to follow equally systematic social variations (see Boudon, 1974, p. 29; Breen & Goldthorpe, 1997).

From an empirical point of view, while still unclear the extent to which the social differentiation of education benefits is due to status-maintenance concerns, as Breen and Goldthorpe (1997) assume, and how intensely these concerns and cost perceptions only drive secondary effects (see Gabay-Egozi et al., 2010, and Stocké, 2007), it is empirically proven that both benefit and cost subjective perceptions of education tend to be more favourable, the higher the group to which the agent belongs (see, for instance, Becker, 2003, pp. 19–21; Need & de Jong, 2000, p. 88; Stocké, 2007, p. 512).

In order to take this empirical regularity into account, agents' perceived benefits of education  $B_{ig}$  are taken to be log-normally distributed with common variance  $\sigma_B^2$  across groups of agents, but with different mean  $B_g$ . For modelling parsimony, however, I do not follow the rational-choice approach in representing benefit and cost perceptions as two distinct quantities to be subtracted from each other (for an explicit representation of this kind, see, for instance, Jonsson & Erikson, 2000, p. 359). In the present model, the parameter  $B_{ig}$  must instead be interpreted as the final ratio established by the agent between his/her benefit and cost perceptions, with a value higher than 1 indicating that the agent's perceived benefits of education outweigh his/her perceived costs.

### *The Multiplicative Term between Ability and Benefit*

The  $\varphi$  (A) term of Eq. (2) shows that the role performed by the agent's benefit perceptions of education in the process of educational preference formation is assumed to be more subtle than its simple additive contribution *over and above* the effect of ability. Agents' abilities and perceived benefits

are supposed to interact with each other. This assumption comes from Jonsson and Erikson (2000, p. 359), who explicitly hypothesise that actors weigh their perception of the benefits of a given educational level against an estimation of their probability of success at that level (see also Breen, 1999). For the sake of simplicity, I follow Breen and Goldthorpe (1997) in assuming that this estimation can be represented directly in terms of ability (it should be noted, however, that Breen and Goldthorpe posit that ‘benefits’ and ‘ability’ affect choices only independently). Stocké’s (2008) empirical results suggest that this interaction may play a role in educational decisions.

Eq. (3) provides the specific way in which the present model represents this interaction. In particular, it is assumed that, when an agent’s ability is lower/higher than 0.5 (as clarified later, all agents’ variables are rescaled to vary between 0 and 1), his/her perception of education benefit is *reduced/increased* by an amount inversely proportional to this perception but (nonlinearly) directly proportional to his/her ability. Basically, the functional form thus states that, taking 0.5 as the threshold, the lower/higher the agent’s ability, the lower/higher the sanction/prize that modifies his/her initial benefit perception; however, at the same time, that the more positive the initial perception of benefits, the less the agent is sensitive to ability (the specific form also makes it possible to keep the term between 0 and 1, which, for reasons that will be clearer later, is desirable).

$$\begin{aligned} \text{if } A_{ig} \leq 0.5, \text{ then } \phi(A) &= B_{ig} \times (B_{ig})^{A_{ig}} \\ \text{if } A_{ig} > 0.5, \text{ then } \phi(A) &= B_{ig}^{(1-A_{ig})} \end{aligned} \quad (3)$$

Andrew and Hauser (2011) have recently cautioned against exaggerating the intensity of the interaction between actors’ ability and expectations, their data suggesting that, while students do indeed modify their educational expectations during their school careers as a function of their academically proven ability, they do so only in response to large changes in the latter. Despite its complexity, the functional form of the term  $\phi(A)$  is able to represent this empirical fact (numerical examples that prove this statement are available upon request).

### *The Social Influence Term (SI)*

Eq. (2) shows that there is a third major factor supposed here to drive the formation of an agent’s educational preference for a given educational level  $L$ : the amount of social influence (SI) exerted on the agent by the artificial

agents with whom s/he has direct relationships (hereafter, agent's neighbours,  $N$ ). In the present context, the agent's neighbours are assumed to be his/her closest friends, those persons who really matter for his/his attitude toward schooling. In particular, as expressed by Eq. (4), this influence is quantified in fractional terms as the number of the agent's direct neighbours who have chosen at instant  $t-1$  the educational level  $L$  that the agent is evaluating at time  $t$  divided by the agent's total number of neighbours.

$$SI_{igL} = \frac{N_i(L=1)_{t-1}}{N_i} \quad (4)$$

This term puts an important theoretical element back into a reason-based explanation of educational inequalities. Indeed, while [Breen and Goldthorpe's \(1997\)](#) seminal paper makes no mention of network-based social influence as a potential mechanism generating class differentials in educational choices, [Boudon's](#) original model explicitly admitted it, even though he did not formalise it – ‘not choosing a prestigious curriculum may represent a high social cost for a youngster from a middle-class family if most of his friends have chosen it; but choosing the same course may represent a high cost for a lower-class youngster if most of his friends have not’, – wrote [Boudon \(1974, p. 30\)](#). Discarding *a priori* interaction-based reasons of this kind seems unwise because, as actors' (and their parents') networks tend to be socially segregated (see [DiPrete, Gelman, McCormick, Teitler, & Zheng, 2011](#), for a recent survey-based study), higher- and lower-group actors are likely to obtain systematically different feedbacks with respect to schooling from their contacts – a fact that the Wisconsin research tradition has acknowledged since the 1960s (see [Haller & Woelfel, 1972](#); [Sewell et al., 2003](#)). To the best of my knowledge, only [Morgan \(2005, ch. 6\)](#) and [Jaeger \(2007, p. 474\)](#) have recently overtly suggested that network-based imitative behaviours should be (re)introduced into a rational-choice explanation of educational choices.

Eq. (4) generalises and formalises this idea. The SI term implies that the higher the proportion of choices for educational level  $L$  within the agent's close relational neighbourhood, the larger the impact on his/her probability of also choosing  $L$ . Hence, the SI term explicitly states that social interactions with closest friends matter for educational choices because they trigger ‘educational conformism’.

Why should this be so? The basic hypothesis underlying Eq. (4) is that of ‘mimetic interactions’ (for this concept in economics, see [Orléan, 1995](#)): that is to say, interactions in which actors have good reasons to imitate each other (for the concept of ‘rational imitation’, see [Hedström, 1998](#)). Cognitive- and normative-belief changes, as well as opportunity changes,



may be at work here. On a cognitive level, the larger the proportion of *ego*'s contacts that choose educational level *L*, the more cognitive salient educational level *L* can become for *ego*, thus increasing the probability that *ego* will also choose educational level *L* (see Harding et al., 2011).<sup>6</sup> On a normative level, the larger the proportion of *ego*'s contacts choosing educational level *L*, the higher the probability that *ego* will have to pay psychological costs in terms of individual and social identity if s/he makes a different choice (for a theory of educational choices entirely relying on this mechanism, see, in economics, Akerlof, 1997, and Akerlof & Kranton, 2002). With respect to opportunity, finally, the larger the proportion of *ego*'s contacts that choose educational level *L*, the higher the probability that *ego* will gain access to resources – information about the organisation of education, course notes, or material resources like transportation and housing – exploitable during school life if s/he makes a similar choice.

In economics, the basic intuition at the core of the so-called membership theory of inequalities (see Durlauf, 1999a, 2002, 2006) is that when mimetic interactions of this kind are at work among individuals with similar characteristics, individual outcomes depending on these characteristics will be reinforced, thus spreading more quickly and becoming even more resistant to change. In the analysis of economic inequality, the concept of 'poverty trap' refers to this self-reinforcing interaction-based dynamic (see Durlauf & Cohen-Cole, 2004).

To study the extent to which this process may matter for the genesis of educational inequalities, the term SI of Eq. (2), and hence the computation of Eq. (4), will rely on a specific network structure. In particular, each agent of the artificial population that I simulate is embedded in a network of artificial dyadic links whose proportion of in-group links – that is, links with agents belonging to the same group (and therefore with the same educational background) – is a network parameter that can be manipulated. Technically, I build on the structural properties of a 'small-world' topology (see Watts, 2004) to create varying degrees of educational homophily and heterophily within agents' ego-centred networks.<sup>7</sup>

### *The Simulation Algorithm and the Model Calibration Procedure*

Fig. 1 summarises the basic set of algorithms that I programmed to deduce by simulation the aggregate distribution of agents' educational choices as a function of the group to which they belong when the educational behaviour of each of them is driven by the formal model expressed by Eqs. (1)–(4).<sup>8</sup>

The computational sequence depicted in Fig. 1 shows that, after the initialisation stage (steps 1.1–1.3), the simulation enters a double-step loop in which agents are allowed for 52 iterations to change their mind about whether or not to choose educational level  $L$  (see step 2.1). On the basis of the identity one iteration = one week, this lapse of time was chosen to roughly mimic the year which usually lasts between two enrolling dates in the real educational system. At the end of this intermediate decision time, during which, when the social influence term of Eq. (2) is activated, agents continuously influence each other (see Eq. (4)), the educational choice of each agent is fixed (step 2.2). Only once all the agents have reached a stable educational destination, then, is computation made of the simulated cross-tabulation which expresses agents' highest educational choice as a function of the educational group to which they were attributed at the beginning of the simulation (step 2.3).

As said in section 'Data, variables and research strategy', for any specific set of parameter values, steps 1.1–2.3 are replicated 100 times (with 100 different seeds), each of these replications generating a slightly different simulated cross-tabulation. The statistics described in section 'Data, variables and research strategy' were computed on each simulated cross-tabulation and compared with the same statistics computed on each of the 100 French empirical cross-tabulations (bootstrapped, it will be recalled, from the French original sample). It should thus be clear that, except for the size of agent groups, the simulation algorithm does not contain any empirical information from the empirical data whose structure the formal model is intended to reproduce. Each simulated cross-tabulation is a purely numerical construction deduced from the behavioural rules contained in Eqs. (1)–(4) under a specific set of parameter values.

As regards determination of these parameter values, I followed the procedure usually referred to in simulation studies as 'model calibration' (see, for instance, [Railsback & Grimm, 2012](#), ch. 20; [Stonedahl & Wilensky, 2010](#)). For each specific combination of mechanisms introduced in the simulation, model parameters are stretched to minimise the 'distance' between simulated and empirical cross-tabulations – the 'distance' being quantified here by means of the dissimilarity index, which will be computed as half of the sum of the absolute difference between the simulated value and the observed value in every cell of the table divided by the sum of the frequencies in the table (see [Breen, 2004](#), p. 24). This tuning operation will concern only the means  $A_g$  and  $B_g$  of the log-normal distributions from which agents' ability and perceived benefits are drawn, that is, 10 out of the 16 independent parameters that may be manipulated (see Appendix A,

**BEGIN****1-INITIALIZATION****1.1-Initialization of the population**

CREATE 5000 Agents;  
 ATTRIBUTE agents randomly to one of the following  $g$  groups:  
 $g_1=3ab=990$ ,  $g_2=2c=380$ ,  $g_3=2ab=524$ ,  $g_4=1bc=1769$ ,  $g_5=1a=1418$ ;

**1.2-Initialization of agent's attributes**

SET agent's state "IN the educational race";  
 SET current educational transition "1=first transition=1bc";  
 SET choice for the current educational transition "not chosen";  
 SET number of failures at the current educational transition 0;  
 IF ability term (A) "on" (see Eq. 2).

SET agent's  $A_{ig} \sim$  Log-normal with mean =  $A_g$  and variance =  $\sigma_a^2$ ;

IF benefit term (B) "on" (see Eq. 2).

SET agent's  $B_{ig} \sim$  Log-normal with mean =  $B_g$  and variance =  $\sigma_b^2$ ;

**1.3-Initialization of agent's network**

IF social influence term (SI) "on" (see Eqs. 2 and 4).  
 CREATE a "small-world" network with:  
 1/ average degree =  $K$ ;  
 2/ probability of out-group links =  $p$ ;  
 3/ probability of being linked to an agent belonging to a specific group:  
 $P_{GSD=1} > P_{GSD=2} > P_{GSD=3} > P_{GSD=4}$  ( $GSD$ = group social distance);

**2-MODEL DYNAMIC**

SET iteration 0;

UNTIL any agent with state "IN the educational race":

**2.1-Intermediary Choices**

REPEAT 52 times:  
 For each agent with state "IN the educational race", sequentially but randomly invoked:  
 COMPUTE educational preference according to equation [2];  
 DETERMINE choice for the current educational transition according to equation [1];  
 SET iteration + 1;

**2.2-Final Choice**

IF iteration / 52 = integer,  
 DETERMINE choice for the current educational transition according to equation [1] ;  
 IF choice for the current educational transition "not chosen",  
 SET number of failures at the current educational transition + 1;  
 IF number of failures at the current educational transition > 3,  
 SET agent's state "OUT of the educational race" ;  
 IF choice for the current educational transition "chosen",  
 IF current educational transition "4=last transition=3ab",  
 SET agent's state "OUT of the educational race";  
 OTHERWISE,  
 SET current educational transition "current educational transition + 1" ;

**2.3-Choice Aggregation**

IF iteration / 52 = integer,  
 For each educational transition, COMPUTE the number of agents with choice = "chosen";  
 CROSS-TABULATE agents' education choices by agents' educational group  $g$

**END**

Fig. 1. Basic Steps in the Simulation Algorithm.

Table A.1). It should be pointed out that the result of this parameter stretching operation is not predetermined. Nothing ensures that one will end up with a good match between simulated and empirical data under a qualitatively realistic parameterisation. As stressed by [Railsback and Grimm \(2012, p. 256\)](#), the purpose of such a model calibration procedure is precisely to establish if ‘there is something wrong with the model so that it cannot be forced to match observations closely’. Thus, it would be inappropriate to consider the model calibration strategy adopted as tautological.<sup>9</sup>

## COMPUTATIONAL RESULTS, EMPIRICAL DATA AND MODEL DYNAMIC

As anticipated in section ‘Data, variables and research strategy’, in order to assess the specific explanatory contribution of the micro- and network-based mechanisms formalised by Eq. (2), the agent-based model was simulated by introducing these mechanisms sequentially. To be noted in this regard is that, before the agent’s educational preference was computed according to Eq. (2), each term entering the equation was rescaled to range between 0 and 1. This procedure ensures that the differences that might be observed between the simulations of two model variants reflect the net effect of the mechanism introduced rather than the differences in the range of values taken by the numerical variables used to express the mechanisms. As regards the order of mechanism introduction, quantitative studies on educational inequalities suggest the following (see, in particular, [Stocké, 2008](#)):

1. In order to assess how intensely ability distribution across agent groups segregates their educational choices, the model was first simulated with agents’ educational choices driven by ability only (i.e. term A of Eq. (2));
2. In order to establish whether the distribution of subjectively perceived benefits of education segregates educational choices across agent groups *over and above* ability distribution, the model was then simulated with agents’ educational choices driven by ability coupled with perceived benefits (i.e. terms A and B of Eq. (2));
3. In order to determine the extent to which the interaction between ability and perceived benefits contributes to segregating educational choices more than when these two factors work independently, the model was simulated thirdly with agents’ educational choices driven by ability, perceived benefits and the multiplicative term between the two (i.e. when the term  $\varphi(A)$ , see Eq. (3), is present);

4. In order to evaluate whether or not dyadic interactions between agents have any specific inequality effect *over and above* the three previous mechanisms, the model was finally simulated in its complete form, that is to say with agents' educational choices driven by ability, perceived benefits, the multiplicative term between the two and the term expressing the network-based social influences as formalised by Eq. (4).

Table 2 reports the set of statistics (see section 'Data, variables and research strategy') computed on the distribution of simulated cross-tabulations (whose average is reported in Appendix A, Table A.2) generated under the four model variants (the last row in Table 2 refers to the model discussed in Appendix B).

As a benchmark, let us first consider a model in which none of the mechanisms postulated is at work (see Table 2b). Unsurprisingly, since no source of inequalities across agent groups is present, no educational stratification arises. For instance, the average generalised odds ratio ('agor') expressing the overall level of educational fluidity fluctuates around 1 across replications, meaning that no agent group has a competitive educational advantage at any educational level. The specific numerical structure of the simulated educational outflows is also easy to explain (see Appendix A, Table A.2b). When none of the mechanisms postulated is at work, by construction the educational preference of each agent, no matter what group s/he belongs to, amounts to 0, which implies that each agent has a very low probability of choosing a given educational level (namely, about 0.18). The sequential nature of the four educational choices to be made, coupled with the fact that the agent can go through each educational transition no more than three times (see Fig. 1, step 2.2), automatically explains that, on average, about 55% of each agent group is not able to make the first transition. As testified by the value of the dissimilarity index ('di'), this simulated distribution is extremely far from the distribution observed in the empirical data.

When the agent's educational choices are driven only by his/her ability, a substantial improvement towards realism is accomplished (see Table 2c). If agents' abilities are differentiated across agent groups – in particular, the higher the group, the higher, on average, the agents' ability (see Appendix A, Fig. A.1, solid line) – ability differentiation is sufficient to generate some amount of stratification of educational choices across agents' educational backgrounds. For instance, agents belonging to the highest group attain the highest educational level between two and three times more frequently than do agents belonging to the lowest group (see Table 2c, column 'hlgr').

**Table 2.** Non-Parametric (0.025–0.975) Percentile Intervals of Absolute/Relative Educational Mobility Statistics (See Section ‘Data, Variables and Research Strategy’) Computed on Each of the 100 Empirical Bootstrapped Cross-Tabulations and, on the Other Hand, on Each of the 100 Simulated Cross-Tabulations Generated by the Simulation of Several Model Variants.

IP	UMP	HLGR	AGOR	HGGOR	LGGOR	DI
(a) Empirical French Data (bootstrapped cross-tabulations) 37.46–40.44	44.44–47.21	6.38–8.86	1.61–1.79	5.38–8.13	8.23–12.00	NA
(b) ‘Straw Man’ Model: All Mechanisms Are Absent 25.56–27.65	20.18–22.43	0.63–1.51	1–1.02	0.7–1.36	0.83–1.15	42–44.18
(c) Model Variant 1: Ability 26.45–28.18	35.39–37.44	2.07–3.05	1.03–1.05	1.17–1.56	1.64–2.08	24.57–26.72
(d) Model Variant 2: ‘Model Variant 1’ + Perceived Benefits 32.09–34.14	38.51–41.75	5.14–7.81	1.17–1.25	2.76–3.89	2.69–3.53	15.89–18.17
(e) Model Variant 3: ‘Model Variant 2’ + Ability/Benefits Interaction 33.68–35.6	38.23–41.34	5.97–8.96	1.22–1.32	3.42–5.4	2.77–3.74	14.81–17.17
(f) Model Variant 4: ‘Model Variant 3’ + Social Influence 34.22–36.32	43.09–46.65	6.86–9	1.3–1.48	5.44–10.09	3.27–4.28	11.04–13.28
(g) Model Variant 5: ‘Model Variant 4’ + Inflation Spiral Mechanisms (see Appendix B) 31.89–34.37	50.44–54.28	5.17–6.54	1.28–1.44	5.22–9.53	3.16–4.15	13.89–16.38

This value, like those of all other statistics computed on the simulated data, however, is far lower than the value computed for the empirical data. As testified by the value of the dissimilarity index ('di'), the overall distribution of educational choices across group of agents is still quite distant from the French actual distribution.

The model variant containing both ability and perceived benefits makes it possible to get closer to the empirical data. When both ability and benefit perceptions are systematically differentiated across agent groups – in particular, the higher the group, the higher, on average, the agent's ability/benefit perceptions (see Appendix A, Fig. A.1, dotted lines) – the simulation of the model produces a deeper stratification of educational choices compared to that generated by the model with agents' behaviour driven by ability only. The percentage of agents undertaking the highest education transition within the highest group compared to the lowest group is now considerable larger (compare Tables 2c and 2d, column 'hlgr'). Statistics expressing the relative positions of the highest and lowest groups of agents ('hggor' and 'lggor', respectively) also present higher values than in the previous scenario, indicating that systematic differentiation of benefit perceptions across agent groups contributes to deeply segregating educational choices among these groups.

This is an important theoretical result. Indeed, in line with [Breen and Goldthorpe's \(1997\)](#) hypothesis, the simulation demonstrates in a rigorously deductive way that a bottom-up dynamic in which educational choices are driven by socially differentiated perceptions of education benefits is able to accentuate group differentials in educational choices *over and above* 'primary effects' – controlled for here by including the differentiation of agents' ability in the simulation. As a meaningful numerical example – if one extracts agent-level data from the simulation being commented on here, and if one considers agents from the highest and the lowest groups with strictly comparable ability, namely, those whose ability ranges between 0.2 and 0.4 – it would appear that, when differentiated benefit perceptions are present, despite the same level of ability, the median educational level attained by agents from the highest group is the second or the third one (in 71 out of the 100 replications of the simulation), whereas the median educational level reached by agents belonging to the lowest educational level is the first one (in 94 out of 100 replications).

That said, the simulation of the model variant including ability and benefit perceptions also shows that, given the effect of ability, the inequality effect of socially differentiated perceptions of educational benefits is not sufficient to differentiate agents' educational choices as much as they are

differentiated within the French empirical data. Indeed, the dissimilarity index ('di') tells us that between about 16% and 18% of cases are still misclassified by the simulated model.

This gap is not substantially reduced when, given the ability and benefit differentiation across agent groups allowed in the previous model variant, the model is simulated by including the multiplicative term between agents' ability and perceptions of education benefits (see Eq. (3)). The Mathieu effect implied by this term – on average, agents with higher ability will also tend to have more positive perceptions of education benefit so that the two factors tend to reinforce each other – accentuates the stratification of educational choices among groups a little bit more. For instance, the overall competitive advantage of the highest agent group is larger than in the model in which ability and benefit perception act as independent mechanisms (see Table 2e, column 'hggor'). However, the value of all other statistics remains virtually unchanged, and the overall fit of the simulated cross-tabulations to the empirical data is only marginally improved – the range of the dissimilarity index largely overlaps with the range obtained under the previous model variant.

By contrast, a noticeable fit improvement appears when the last mechanism hypothesised is introduced, that is, the network-based inter-dependences among agents' educational choices (see Eq. (4)). As regards network parameterisation, the following choices were made: (1) each agent has, on average, four direct friends; (2) only 10% of all links present in the system relate agents belonging to different educational backgrounds; (3) these inter-group links are mainly of a short-distance type – that is, about 60% of inter-group links put in contact agents with educational backgrounds at distance 1 (meaning with agents from the group just above/below them). These choices aim to mimic characteristics of empirical social networks, namely, that the number of closest friends substantially affecting specific behaviours tends to be relatively small (for instance, McPherson, Smith-Lovin, & Brashears, 2006 found a size ranging between 2 and 3 for confident discussions networks; Christakis & Fowler, 2007, found a friendship average degree of around 1 with range from 0 to 8, see also Fowler & Christakis, 2008; Snijders & Steglich, 2013, state that friendship networks usually exhibit average degrees between 3 and 4), and, on the other hand, that both 'strong' and 'weak' friendship ties are overwhelmingly homophilic (on education-based mate selection, see Blossfeld, 2009, and Skopek, Schulz, & Blossfeld, 2011, for on-line dating settings; on the social segregation of 'weak' ties, see DiPrete et al., 2011; on homophily more generally, see McPherson, Smith-Lovin, & Cook, 2001).



Under this parameterisation (alternative ones will be studied in the next section), the introduction of the network-based social influence mechanism leads to aggregate results that are very much in line with empirical observations (notably, 3 out of 6 statistics almost completely overlap with their empirical counterparts). In particular, the presence of local dyadic interactions between agents increases the percentage of educationally immobile agents, the overall intensity of inequality of educational opportunity, and, quite remarkably, the competitive educational advantage of the highest agent group (see Table 2f, columns ‘ip’, ‘agor’ and ‘hggor’, respectively). These indicators are especially important because they testify to the capacity of the network-based mechanism to segregate educational choices to a larger extent than the three previous ‘atomistic’ model variants were able to do, thus bringing the simulated educational stratification closer to the strong level of inequality exhibited by the actual French data.

In order to see the net effect of local dyadic interactions among agents more directly, let us re-run the simulation under discussion with exactly the same distributions of agents’ ability and benefit perceptions (see Appendix A, Fig. A.1, dashed lines) but turning off the social influence term. Fig. 2 reports the result of this counterfactual by focusing on the

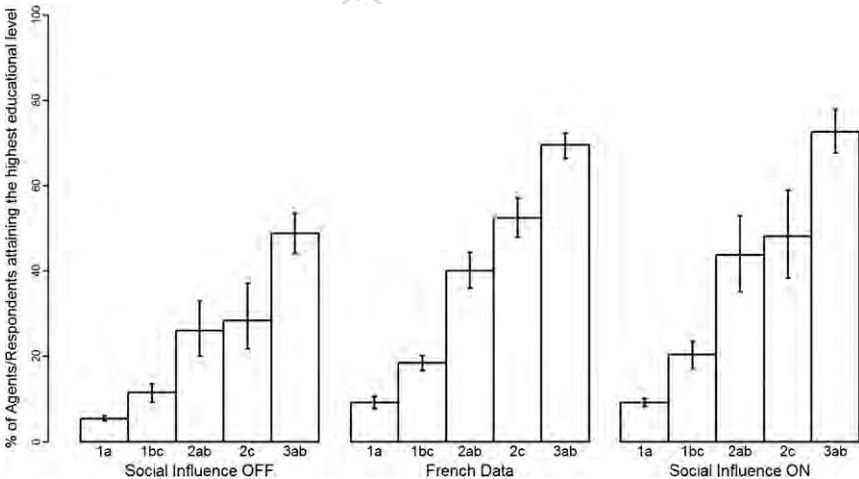


Fig. 2. Percentages of Agents Attaining the Highest Educational Level within the Five Agent Groups Averaged over 100 Replications of the Model (Bars Give the Values within Which Fall 95% of the Model Replications) When the Social Influence Term (SI) is Turned Off/On.

percentages of agents attaining the highest educational level within each of the five groups of agents.

Three facts appear. First, the differentiation of educational choices across agent groups is much more pronounced when the social influence term is present. Given that the differentiation of ability and benefit across groups is exactly the same under the two experimental conditions, this means that network-based interdependences among agents' educational choices have the capacity to progressively amplify the initial differentiation of ability and benefit perceptions. Second, this segregating effect of local dyadic interactions operates much more intensely, the higher the agent's group. Finally, no matter what group is considered, the empirical percentages observed in the French data fall within the range of the simulated percentages only when the social influence term is turned on, thereby suggesting that the actual deep stratification of educational choices cannot be fully explained by postulating only a socially differentiated structure of ability and perceived benefits.

Fig. 3 clarifies the source of this segregating effect sustained by local dyadic interactions between agents. For each educational level that agents must choose, the graphs report the proportion of agents' neighbours who

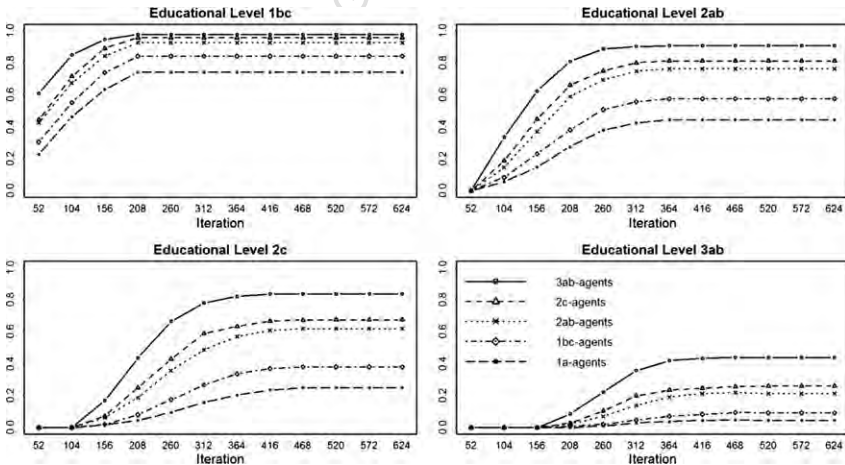


Fig. 3. Average Proportion of Agents' Neighbours Who Chose the Educational Level that the Agent is Evaluating at Instant  $t$  (Proportions Are Averaged over 100 Replications; Variability Across Replications Is Omitted to Make Graphs More Readable).

chose the educational level. To understand the model dynamic, this is the most appropriate piece of information because it is in this that the social influence term basically consists (see Eq. (4)).

The curves show that, as a consequence of the initial differentiation of ability and perceived benefits, the higher the agent's group, the richer his/her local neighbourhood in terms of contacts making a positive choice – which holds at each transition. When one considers that, under the present network parameterisation, agents tend to be linked mainly with agents having similar ability and benefit perceptions, it is clearly apparent that, at each transition, stronger support from the agent's neighbours will thus benefit precisely those agents who already have higher starting values on ability and perceptions of education benefits. As a result, the initial differentiation of ability and perceived benefits is dynamically deepened and widened by means of a cumulative process (see Merton, 1968, pp. 606, 610), in which the amplification mechanism (see Boudon, 1979, pp. 156–157) – the social multiplier if one prefers (see Durlauf, 2006; Durlauf & Cohen-Cole, 2004) – is based here on the socially segregated composition of agents' ego-centred dyadic networks.

While, as argued, the presence of this amplifying process fuelled by local dyadic interactions between agents is necessary to get closer to the actual cross-sectional educational stratification observed in France, to conclude it should be acknowledged that the simulated data generated under the full model do not fit perfectly with the empirical observations. Between about 11% and 13% of cases are still misclassified even when dynamic network-based interdependences among agents' educational choices are included in the simulation (see Table 2f, column 'di'). A careful comparison of the simulated and empirical cross-tabulations shows that this gap is mainly due to the fact that the current model parameterisation tends to allow too many agents belonging to the lowest educational group to go beyond the first educational transition while not leading enough agents belonging to the middle groups beyond the same point (see Appendix A, compare Tables A.2a and A.2f). Among the statistics computed on the simulated data, this under/overproduction is reflected by the value of the generalised odds ratio summarising the relative position of the lowest group of agents. Compared to the empirical value, the simulated one, although it is the best among the four model variants simulated, is still not high enough, which suggests that the full model is too 'generous' with respect to agents belonging to the lowest group.

To correct for this problem, it would suffice to introduce an additional differentiation in agents' abilities and/or perceived benefits with the aim of

representing the existence of specific barriers/advantages at some educational transition for some group of agents. While one may find theoretical/empirical justifications for such modifications, I will not follow this strategy for two reasons. On the one hand, this strategy would imply that the mechanisms postulated are believed to be the only ones at work, which is a position that I would not endorse. Conceptualising additional substantive mechanisms may thus be a strategy to reduce the gap between simulated and empirical data that is more reasonable than introducing fine-tuning *post-hoc* modifications of the existing mechanisms (on this point, see Appendix B). On the other hand, the proximity reached between simulated and empirical data seems quite remarkable given the relatively limited number of parameters that have been manipulated. In particular, ability/perceived benefits were assumed to be differentiated only across groups of agents but not across educational levels. This means that the formal model proposed was able to approximate all the qualitative features of an empirical complex structure contained in a  $5 \times 5$  cross-tabulation by manipulating only 10 parameters out of the 16 independent parameters that might have been manipulated (see Appendix A, Table A.1). In terms of model parsimony, this is a significant result.

## SENSITIVITY ANALYSIS OF THE SOCIAL INFLUENCE MECHANISM

By means of computer simulations, the previous section aimed to prove that a network-based social influence mechanism must be postulated so as to match closely the stratification of educational choices across educational backgrounds observed in the French data. In particular, the amplifying process triggered by this mechanism is necessary to reproduce the large gap at the highest educational levels between the lowest and the highest social groups. This result was produced under a specific network parameterisation (see Appendix A, Table A.1), which, as said above, attempted to mimic some basic empirical regularities of real friendship networks. The present section assesses the robustness of this result against alternative network parameterisations.

In this respect, the first parameter manipulation to be performed concerns the average degree of the network in which agents are embedded. Fig. 4 focuses on the percentage of agents attaining the highest educational level within each group and plots it as a function of increasing average network degrees, 4 being the value adopted in the simulation previously performed to

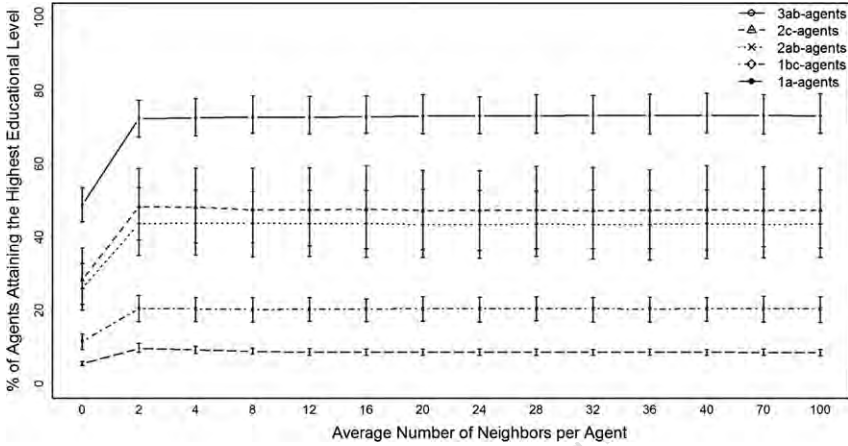


Fig. 4. Percentages of Agents Attaining the Highest Educational Level within Each of the Five Groups of Agents Averaged over 100 Replications of the Model under Different Values of the Parameter  $K$ , that is the Average Number of Neighbours with Whom Each Agent is in Contact (Bars Give the Values within Which Fall 95% of the Model Replications).

approximate the French educational stratification. As the low probability of inter-group link formation is unchanged, this parameter manipulation thus mainly amounts to progressively increasing the number of direct contacts that each agent establishes within his/her own group. In terms of Granovetter’s (1973, 1983) distinction between ‘strong’ and ‘weak’ ties, it is mainly the amount of the agent’s ‘strong’ in-group ties that it is made variable here.

The main result is that the model’s aggregate behaviour is remarkably stable, and hence robust to the modification of the average number of contacts that each agent is allowed to establish (mainly) within the group to which s/he belongs. The curves show that, no matter what group of agents is considered, the only significant difference observed is between the situations in which no contact at all exists – this is simply the model without the social influence term – and those where a few neighbours are allowed to influence the agent’s educational choices. Even extremely high (and unrealistic – it is becoming well known that relevant contacts are relatively limited even within web-based social networks and exchanges: see Easley & Kleinberg, 2010, pp. 54–58; Skopek et al., 2011, p. 185) network degrees leave the percentage of agents attaining the highest educational level virtually

unchanged. Thus, in the present model, ‘strong’ in-group ties do not necessarily have undesirable effects. On the contrary, when there are few of them, they sustain agents’ educational choices; when they are too numerous, at worst, they do not significantly change agents’ behaviour.

The way in which the network-based social influence mechanism is formalised here explains this result (see Eq. (4)). In fact, the fractional form of this term implies that, for a given distribution of agents’ ability and perceived benefits, increasing the size of the agent’s neighbourhood will not affect his/her educational choices because numerical changes in the numerator, that is, the number of contacts choosing the educational level that the agent is evaluating, will be neutralised by similar size changes in the denominator, that is, the total number of the agent’s neighbours. Under increasing neighbourhood sizes, when anything else is unchanged in the model parameter setting, both quantities tend to increase to a similar extent, thus producing virtually no effect at the aggregate level (the analysis that numerically proves this statement is available upon request).<sup>10</sup>

Substantially, this result suggests that if, in the reality, actors were influenced by the proportion of their contacts making a certain choice rather than by their absolute number, being exposed to a large fraction of contacts may have, at aggregate level, similar effects as being exposed to a few neighbours. In the case of educational choices, this seems realistic. If local dyadic interactions matter, as assumed earlier, because interacting with others enables actors to share cognitive and material resources as well as to establish contact with specific identity models, a given actor may share, say, car, housing, course notes and normative models with a few fellow students but, because of time and cognitive limitations, not with tens of them. If so, only a relatively small size of the actor’s friendship network will be really effective.

All other things being equal, modifying the agents’ probability, and hence the amount of the inter-group links, should generate more visible effects. In this case, in fact, for a given number of neighbours, more or fewer contacts with agents outside an agent’s group should imply a modification of the educational composition of the agent’s neighbourhood. The social influence exerted on the agent should thus be qualitatively different, and this should impact to some extent on the agent’s educational choices.

Fig. 5 focuses on the percentage of agents attaining the highest educational level within each group and plots it as a function of values of the probability of establishing contacts outside an agent’s group ranging from 0 to 1, which amounts to moving progressively from five disconnected regular networks where the probability of inter-group linkage is 0 – under

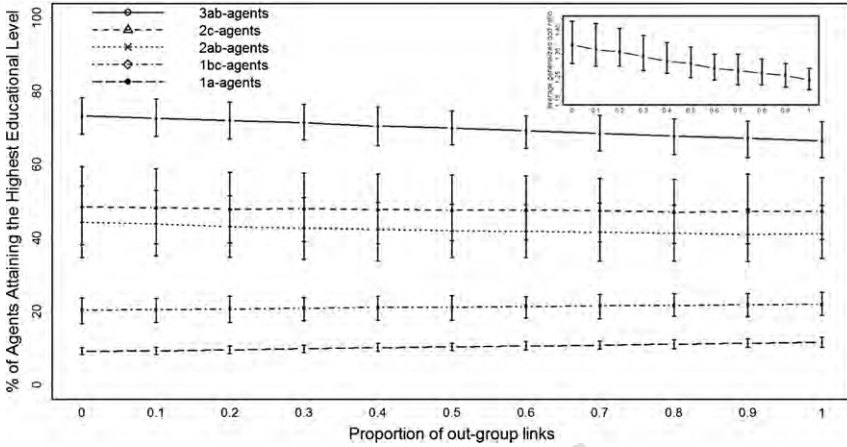


Fig. 5. Percentages of Agents Attaining the Highest Educational Level within the Five Groups of Agents Averaged over 100 Replications of the Model Under Different Values of the Parameter  $p$ , that is the Probability that an Agent’s In-Group Link is Disconnected and Rewired to an Agent Belonging to a Different Group (Bars Give the Values within Which Fall 95% of the Model Replications). The Small Right-Corner Graph Plots the Average Generalised Odds Ratio Computed on the Simulated Cross-Tabulations – the Closer the Coefficient to 1, the Less Inequality of Educational Opportunity is Present in the Table (Central Points and Bars Have the Same Interpretation as Above). To Be Noted Is that the Average Network Degree Is Kept at 4, that is the Value Adopted to Approximate the French Educational Stratification.

this condition, agents only have in-group links – to a single random network where the probability of inter-group linkage is 1 – here agents only have out-group links (for the case of a one-group population, see Watts, 1999, pp. 503–509). In terms of Granovetter’s (1973, 1983) distinction between ‘strong’ and ‘weak’ ties, this experimental setting thus amounts to manipulating the agent’s proportion of ‘weak’ out-group ties.

Two main results ensue. On the one hand, for a large range of increasing heterophily level around the low value adopted to approximate the actual French educational stratification, that is, 0.1, the aggregate behaviour of the model is virtually insensitive to changes in the probability of creating inter-group links. Similarly to what was observed with respect to modifications of the average network, the results presented in the previous section are thus quite robust to alternative network parameterisations. On the other hand, if one considers the entire range of the probability of inter-group link

formation – thus including in the analysis a quite unrealistic level of heterophily, values higher than 0.5 meaning that out-group links tend to become more frequent than in-group links – the curves plotted in Fig. 5 show that the gap between highest and lowest groups at the highest educational transition tends to be reduced.

The upper right-corner subplot in Fig. 5 testifies that this effect is not localised to the distribution of choices at the highest educational level. It shows, in fact, that if one computes the average generalised odds ratio on the simulated cross-tabulations generated under increasing levels of ego-centred network heterophily, then the value of this synthetic coefficient clearly exhibits a decreasing tendency, suggesting that the overall amount of inequality of educational opportunity is reduced by allowing agents to make contacts with an increasing number of agents with educational backgrounds different from their own. In the present model, in line with Granovetter’s original argument, ‘weak ties’ thus generate desirable consequences at the aggregate level, namely, more educational fluidity. However, contrary to the idea that ‘weak ties’ tend to generate positive effects *in general*, closer inspection of the curves plotted in Fig. 5 shows that, while the lowest group agents tend more frequently to reach the highest educational level when the proportion of inter-group links increases, the contrary holds for the highest group agents, while the in-between agent groups do not exhibit any particular trend.<sup>11</sup>

This asymmetry arises from the way in which the network linking agents is created within the artificial society studied here. On the one hand, as shown by Eq. (4), dyadic links are not weighted, which implies that the educational choice of an upper-group agent matters for the educational choice of a lower-group agent exactly to the same extent as the choice of the latter matters for the choice of the former. On the other hand, since the network-building algorithm requires the network degree to be kept constant, each inter-group link that has been created replaces one in-group link (see footnote 7). Under these conditions, when the out-group links tend to put into contact agents (mainly) belonging to groups at a one-step distance (see Appendix A, Table A.1), agents at the very bottom of the group hierarchy will benefit from being more and more in contact with agents just above them, who tend to make slightly more positive choices than they do. Agents in the middle of the group hierarchy will be stretched by counterbalancing tendencies, agents above them making slightly more positive choices whose positive influences will be neutralised by agents below them making educational choices slightly less positive than theirs. The educational outcomes of highest group agents can instead only become



worse, their neighbourhood becoming progressively filled with agents below them who make less positive educational choices than they do.

This result thus illustrates, in the context of a theoretical model of a specific relevant social behaviour, that is, educational choices, the counter-intuitive results obtained by Centola, Eguiluz, and Macy (2007) and Centola and Macy (2007) using strictly comparable network topology, who demonstrated that the ‘strength of weak ties’ does not hold in general. Under certain conditions, namely when, for some or other reason, a given behaviour is ‘costly, risky or controversial’, the support of those in the best position to make the choice – that is, in the present case, agents with the same educational background – is crucial for adopting the behaviour. When heterophily increases beyond a certain level, within the artificial society studied here, highest groups agents can no longer profit from relational proximity with agents with a similar capacity to make costly educational choices, thus losing part of their competitive advantage at the highest educational transition.

## DISCUSSION

Section ‘Computational modelling of educational preference formation’ developed a generative formal model of the macro-level structure of educational inequality which frames educational choices as the result of both subjective ability/benefit evaluations and peer-group pressures, thus enriching rational-choice explanations of educational choices in sociology through theoretical insights deriving from heterodox theoretical perspectives in economics of inequality and of education.

By means of agent-based computational simulations, the macro-level consequences of this model have been tested against French empirical data. By introducing the hypothesised mechanisms sequentially and by quantifying the proximity between the simulated and the empirical educational stratifications for each combination of mechanisms, section ‘Computational results, empirical data and model dynamic’ has deductively proved that ability and subjective perceptions of education benefits, no matter how intensely differentiated across agent groups, are not sufficient *on their own* to generate the actual stratification of educational choices across educational backgrounds existing in France at the beginning of the twenty-first century. By computational counterfactual manipulations, it has been proved that network-based interdependences among educational choices contribute, *over and above* the differentiation of ability and of benefit perceptions, to the

genesis of educational stratification by reinforcing, hence accentuating, the segregation of the educational choices that agents make on the basis of purely private ability/benefit calculations.

It should now be emphasised that the results presented in section ‘Computational results, empirical data and model dynamic’ (see Table 1) allow, although imperfectly, assessment of the relative generative power of each of the mechanisms postulated. In particular, close inspection of the variations of the dissimilarity index across the model variants that have been simulated suggests that (1) the differentiation of ability across agent groups leads to the largest reduction, that is, around 17 percentage points, in the dissimilarity index value compared to the baseline model in which none of the mechanism hypothesised was at work; (2) the differentiation of education benefit perceptions across agent groups is the second most powerful factor, leading to an around 9 additional percentage-points reduction of the dissimilarity index compared to the ability-based model variant; (3) the network-based social influence mechanism comes in the third position, yielding around 4 more additional percentage-points in the dissimilarity index reduction compared to the model variant in which ability, benefit and the multiplicative term between these two factors were at work – the multiplicative term between ability and benefit coming last with only about 1 percentage-point of dissimilarity index reduction.

Although the paper has demonstrated that the network-based social influence mechanism is necessary to match the empirical data closely, it does not lead to the conclusion that all social interactions matter. Quite the contrary, the ‘explanatory’ hierarchy established on the basis of the dissimilarity index reduction is very much in line with statistically based empirical studies on educational inequalities which suggest that, while ability/cognitive skills systematically exhibit very high predictive power (see, for instance, Gabay-Egozi et al., 2010; Stocké, 2007), peer effects, when statistically significant, tend to be relatively weak (for an overall assessment, see, for instance, Breen & Jonsson, 2005, p. 229). In this respect, the main contribution of the present simulation-based study is its demonstration that, although their quantitatively average net effect might be modest, dyadic social interactions should not be discarded on a theoretical level because, without them, the actual level of educational inequality could not be accounted for. While the network-based social influence mechanism cannot generate *on its own* educational choice differentials across social backgrounds, the analyses reported here prove that it is necessary to assume that the mechanism is at work to account for the actual deep social segregation of educational choices across social groups, especially when the gap between

highest and lowest social groups at the highest educational levels is considered.

Section ‘Sensitivity analysis of the social influence mechanism’ has shown that this result is robust over a wide range of different, but still realistic, network parameterisations, with respect to both the local density of the network in which agents are embedded and the amount of heterophily allowed within the network. However, it is important to be aware of the specific way in which the network-based social influence mechanism has been conceptualised here.

In this regard, the following simplifications warrant especial attention. First of all, the network linking artificial agents is assumed to be static during the simulation – the educational states of agents’ neighbours thus dynamically change but the set of neighbours does not. Second, each tie is assumed to be symmetric and is not weighted – dyadic influences are thus systematically bidirectional and of equal intensity. Third, agents’ sensitivity to neighbours’ educational choices is not modelled – agents are thus homogenous with respect to how they react to social influence. Finally, the agent’s network does not contain any spatial/geographical dimension.

Without a doubt, enriching the model with more conceptually refined solutions on each of these aspects would lead to stimulating theoretical insights. Assuming, for instance, that friendship networks evolve during an agent’s school career – that is, allowing lower-group agents who succeed to become progressively more frequently in contact with higher-group agents – would be an interesting theoretical development. Introducing asymmetric and weighted ties – that is, assuming that a higher-group agent is less affected by the choice of a lower-group neighbour than the latter is affected by the choice of the former – would make the model more realistic. Embedding friendship networks in spatial neighbourhoods – that is, limiting the probability that geographically distant agents interact with each other – would also increase the model’s applicability. My expectation, however, is that unless the first modification is given an unrealistically high (numerical) power, these modifications can only strengthen the main result produced by the model studied here: that is, network-based interdependences among educational choices tend to reinforce and accentuate the differentiation across social groups in individuals’ ability- and benefit-based calculations, thus helping to explain the actual high level of social segregation of educational choices.

Moreover, it should not be overlooked that additional hypotheses, no matter how theoretically realistic they may appear, imply more parameter values to be initialised, which, from the standpoint of an empirically

oriented variant of the simulation methodology, means in turn more empirical data to be collected in the hope of eventually being able to parameterise the model by means of real-world data.

This, in my opinion, is the main limitation of the present study. While the macro-level consequences of each variant of the formal model have been systematically matched against aggregate individual-level survey data, the parameter values on which each micro-level mechanism relies only come from arbitrarily chosen numerical values. Although these choices can be justified, and although their aggregate consequences have been proved to be realistic, so proving the generative sufficiency of the mechanisms at hand (see Epstein, 2006, chs. 1–2), I believe that it would be highly appreciable to input agents' ability and education benefit perceptions on the basis of representative individual survey data. This, I maintain, is the further step most needed for the present work. Besides providing a full example of a truly empirically calibrated agent-based model (on this concept, see Hedström, 2005, ch. 6), introducing such micro-level empirical bases into model parameters would also enable more solid assessment of the relative explanatory power of the theoretical mechanisms that the model assumes to be at work. After all, this is probably what formal modelling – here in its agent-based computational variant – is fundamentally meant to do: point out relevant parameters for which richer and more detailed empirical information is needed (see Breen, 2009; White, 2000).

## CONCLUSION

Overall, although imperfect, the formal model at the centre of the paper and the research strategy adopted to study it seem to open promising avenues for future research. On a theoretical level, the model has the interest of linking the macro-level structure of educational inequalities with the way in which group belonging biases the heterogeneity of actors in ability, preferences and social contacts: a factor that has been under-investigated within the sociological rational-choice approach to educational choices. The article thus suggests an explanatory factor that may also help to enhance understanding of the long-term evolution of educational inequalities. As pointed out by Breen et al. (2010, pp. 1515–1516), in fact, studies of educational inequalities now need to explain the mix of temporal reduction and stability of class differentials in educational attainment. Paying closer attention to friendship networks may help. As suggested by Blossfeld and Timm (2003a), educational homogamy induces network segregation. There are strong theoretical

reasons to expect that the expansion of educational systems, and hence increasing duration in school, leads to a high level of educational homogamy across cohorts (for some confirmatory empirical evidence, see Blossfeld & Timm, 2003b). As a consequence, one may expect in turn that social segregation in friendship network at best does not change and, at worst, increases across generations (some signs of an increasing network homogeneity along several socio-demographic traits, including education, were discovered by McPherson et al., 2006, pp. 361–362, 371). Network homophily may thus contribute to explaining the observed mix of change and stability of educational inequality in that stable or even increasing levels of social homophily may counterbalance (or attenuate) the equalising effects of long-term potentially converging benefit and cost perceptions of education across social classes, as well as the effects of educational policies aimed at reducing performance gaps across social groups. On a methodological level, the paper's contribution seems to be twofold. On the one hand, it contributes to the literature on neighbourhood effects by suggesting that agent-based computational simulations can be fruitfully exploited to analyse formal models containing network-based mechanisms. On the other hand, it reinforces the presence of this technique within the sociology of stratification and social mobility and proves that agent-based models are powerful tools with which to design complex sets of hypotheses linking structures, actors and networks, and to test the macro-level consequences of these hypotheses. This test having been performed against empirical data, the paper also suggests the interest of reinforcing an interface among formal theoretical modelling, the quantitative analysis of empirical data and computational techniques – a research strategy which is still too rarely followed in the quantitative analysis of social stratification (for a notable exception, see Bruch & Mare, 2006). Thus, despite the limitations explicitly discussed in the previous section, I would be more inclined to keep on cooking rather than abandon this multi-faceted research strategy.

## **ACKNOWLEDGEMENTS**

Preparatory work for this article was presented at the Sociology Seminar Series (Nuffield College, University of Oxford), at the GSADI Ph.D. Seminar Series (Universitat Autònoma de Barcelona), at the Analytical-Quantitative Sociology Seminar (University of Oslo) and at the Sociology Colloquium Series (MZES, University of Mannheim). I am very grateful to the participants at these seminars for their comments and suggestions. Carlo

Barone is the person who has influenced me most in writing the final version of the article: I would like to thank him for the time he devoted to questioning my conception of simulation and of educational choices. I have also greatly benefited from discussing the article with Clemens Kroneberg, as well as from responding to Andreas Wimmer, Flaminio Squazzoni, Pablo Jensen and Torkild Lyngstad, who carefully commented on a previous version of the manuscript. The comments of two anonymous referees have also been of invaluable help, as well as Gunn Birkelund's regular encouragements. I am greatly indebted to Louis-André Vallet for providing me with his SAS syntax to recode French educational and occupational information into the Casmin schema. I obtained the French dataset from the Réseau Quetelet, which I thank for its efficiency. Last but not least, I wish to express my gratitude to Adrian Belton for revising my English. The usual disclaimers apply.

## NOTES

1. The present paper extends [Manzo \(2009\)](#) in several respects: (1) the formal model presented in section 'Computational modelling of educational preference formation' is designed to generate new, and more recent, empirical data; (2) the *explanandum* is different, in that these data concern the statistical association between individuals' educations and the educations of their parents, rather than the statistical association between individuals' educations and the social class of their parents; (3) the formal model is extended to represent five social groups instead of four; (4) the formal model is implemented in NetLogo instead of Java (whose original code was due to Frédéric Amblard). Compared to [Manzo \(2011\)](#), the paper takes some additional steps further: (1) the formal model is now a truly probabilistic model; (2) the formal model is now able to accommodate a representation of the effect of ability on educational choices; (3) the number of parameters adopted to represent actors' subjective benefits and costs is dramatically reduced from 100 to 10; (4) the simulated agents' scheduling has been greatly simplified, all agents being now sequentially updated at each iteration.

2. [Cobalti \(1989\)](#) proposed using generalized odds ratios as measures of association to describe the relative aspect of social mobility in easily interpretable terms. He later extended this proposal to analysis of the inequality of educational opportunities (see [Cobalti, 1992](#), pp. 139–142). Formally, generalized odds ratios are simply the geometric means of all the 'basic sets' forming a given cross-tabulation (see [Goodman, 1969](#)). As demonstrated by [Kaufman and Schervish \(1987, p. 233\)](#), there is a direct link between log-linear models and generalized odds ratios: for a two-way cross-tabulation, the generalized odds ratios for a given cell can be computed by raising the corresponding multiplicative parameter of the saturated model at power  $(l \times c)/(l-1 \times c-1)$ , where  $l$  and  $c$  are the number of the table's rows

and columns. I built on this relation to compute the coefficients reported in Table 2. Log-linear model estimations were performed with the functions 'loglm' and 'gnm' (for the 'uniform difference' model, see footnote 2) contained in the 'MASS' and 'gnm' R packages, respectively.

3. By exclusively focusing on the aggregate cross-tabulation between individuals' educational background and their educational achievement (see Pfeffer, 2008), I am following the numerous empirical analyses of social and educational mobility that have proved that the basic structure of the cross-sectional association between the two variables is largely invariant across several otherwise important socio-demographic factors, mainly sex and geographical areas of residence. I myself tested this hypothesis for the French data that I focus on here by performing a log-linear analysis of the three-way cross-tabulations of respondents' highest educational levels by their parents' education and sex. The results of this analysis (available upon request) clearly suggested that almost all the variation of respondents' educational attainments was due to their educational backgrounds. The 'constant association' model, in fact, absorbs more than 99% of the residuals produced by the baseline model – here the 'conditional independence' model, which unrealistically postulates that respondents' education only co-varies with gender. By contrast, the 'uniform difference' model (Xie, 1992), which posits that the educational–origin–destination association differs between men and women by a multiplicative factor  $\beta$ , only adds about 0.01% to the variance explained by the 'constant association' model, and the dissimilarity index is virtually unchanged (for a recent review of these statistical models, see Breen, 2004, ch. 2). Thus, the extent of the variations of educational outflows and opportunities across genders does not seem large enough to justify the introduction of specific hypotheses linking families' educational strategies to offspring's sex into the explanatory formal model that I discuss in the next section (for a recent comparative analysis of gender differences in educational inequality which would justify this simplification, see Breen et al., 2009).

4. The choice of  $c = 1.5$  is due to the fact that the model contains three terms (see Eq. 2), each of which will range between 0 and 1, so that the theoretical minimum value for  $P_{igL}$  will be 0, whereas the maximum will be 3. Centring the logistic on 1.5 rather than leaving it centred on 0 thus allows the three terms to fall within a large range of the  $x$ -axis, thereby avoiding high probabilities of making the choice even when no mechanisms are present (i.e. when, by construction,  $P_{igL} = 0$ ).

5. Despite the frequent use of the normal distribution to approximate the distribution of ability, it is long-standing argument in economics that 'there is little reason to assume that ability is in fact normally distributed' (see Mayer, 1960; more recently, see Koerselman, 2011). From a descriptive point of view, assuming a log-normal distribution allows better account to be taken of the frequently observed skewed form of the distribution of ability. From a numerical point of view, this choice also avoids the problem of negative values of ability for which a clear substantive interpretation is not immediately apparent.

6. This interaction-based exposure effect has proven to be relevant for such diverse choices as whether or not to commit suicide (Hedström, Liu, & Nordvik, 2008), whether or not to get divorced (Åberg, 2009), whether or not to pay taxes (Hedström & Ibarra, 2010).

7. To create the artificial network, the specific algorithm that I adopted extends the algorithm proposed by Watts and Strogatz (1998; see also Watts, 1999, pp. 503–506, 524) to the situation where several groups of agents are present and the rewiring process must consequently apply to the creation of links among groups. It can be summarized as follows: (1) first, a regular network of symmetric and non-weighted links is created within each of the five group of agents, each of the agent having  $K$  links; (2) then, each link is considered and, with probability  $p$ , is rewired outside the agent's group; (3) finally, the educational group of the potential out-group neighbour is determined according to the following probabilistic criterion: the more distant the focal agent's and potential neighbour's educational backgrounds, the lower the probability that a link between the two agents will be created. Note that  $K$ , the average number of links among agents set in step 1 is kept constant when rewiring steps 2 and 3 are realized. The rationale behind this choice is that it enables evaluation of the relative impact on the outcome of interest (educational choices, here) of the local density of the network or of the fraction of long ties (the out-group links, here) (see Centola & Macy, 2007, p. 711) – this experimental manipulation will be performed in section “Sensitivity analysis of the social influence mechanism”. Note also that I adopt the simplest concept of distance among agents' educational backgrounds, namely the absolute value of the difference between educational group indexes (ranging from 1, the most advantaged educational background, to 5, the most disadvantaged one).

8. The program is written in NetLogo 5.0 (see Tisue & Wilensky, 2004a, b). Currently, Railsback and Grimm (2012) constitute the best introduction to Netlogo as a programming language for agent-based model building. To appreciate the power Netlogo has achieved, compare Lytinen and Railsback (2012) to Railsback, Lytinen, & Jackson (2006).

9. From a logical point of view, this parameter tuning operation is equivalent to what is done in statistical modelling when, given some identification constraints, one looks for parameter values which minimize the distance between fitted and observed values. As correctly pointed out by Snijders and Steglich (2013), however, unlike statistical modelling, agent-based modelling is not meant for causal inference. As a consequence, parameter search procedures in agent-based modelling do not need to rely on assumptions about variable and error distributions because parameters are not supposed to provide efficient and robust estimations of average effects generalized to the population, but only to express theoretically meaningful and realistic relations among (some of the) variables on which algorithms rely.

10. It should be noted that this numerical phenomenon implied by the fractional form of the social influence term adopted here is part of a deeper phenomenon related to the two ways in which threshold effects more generally can be formalized. As pointed out by Centola and Macy (2007, p. 711), ‘fractional’ thresholds, that is thresholds based on the proportion of an agent's neighbours doing something instead of the absolute number of neighbours, implies that both ‘adopters’ and ‘nonadopters’ influence the agent's behaviour. Under the ‘fractional’ threshold, when the agent's neighbourhood size increases, it is thus also possible that, if the fraction of adopters is limited, the amount of influence undergone by the focal agent decreases instead of increasing because the fraction of ‘nonadopters’ will weigh more. Although tiny, a trace of this counterintuitive effect is present in Fig. 4, namely



among agents belonging to the lowest group for whom, given their average level of ability/perceived benefits, highest educational choices are not so frequent. For them, a slight decreasing trend is visible as the average number of agents with whom they are in contact increases. Given variability across replications, however, it would not be justified to exaggerate the importance of this effect in the present model.

11. All the trends described would be even more marked if one simulated the model under increasing values of the inter-group link probability but applying an alternative, although far less realistic, configuration of these links on the basis of which the more distant the focal agent's and potential neighbour's educational backgrounds, the higher the probability that the out-group link will be created (such a modification amounts to manipulating parameters  $PG_{GSD}$ , see Appendix A, Table A.1). This configuration is thus the exact contrary of the one I have adopted throughout the paper, in that the majority of out-group links would in this case be between maximally socially distant agents rather than between agents at a social distance equal to 1 (results are available upon request).

## REFERENCES

- Åberg, Y. (2009). The contagiousness of divorce. In P. Hedström & P. Bearman (Eds.), *The Oxford handbook of analytical sociology* (pp. 342–364). Oxford: Oxford University Press.
- Agirdag, O., Van Houtte, M., & Avermaet, P. V. (2011). Why does the ethnic and socio-economic composition of schools influence math achievement? The role of sense of futility and futility culture. *European Sociological Review*. doi:10.1093/esr/jcq070.
- Akerlof, G. (1997). Social distance and social decisions. *Econometrica*, 65(5), 1005–1027.
- Akerlof, G. E., & Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4), 1167–1201.
- Albouy, V., & Wanecq, T. (2003). Les inégalités sociales d'accès aux grandes écoles. *Economie et Statistique*, 361, 27–52.
- Andrew, M., & Hauser, R. M. (2011). Adoption? Adaptation? Evaluating the formation of educational expectations. *Social Forces*, 90(2), 497–520.
- Ballarino, G., & Bernardi, F. (2001). Uso di dati time-budget per lo studio delle risorse familiari: Capitale sociale e culturale dei genitori dei bambini in età scolare in Italia. *Quaderni di Sociologia*, 2, 7–36.
- Becker, R. (2003). Educational expansion and persistent inequality of education. Utilizing subjective expected utility theory to explain increasing participation rates in upper secondary school in the federal republic of Germany. *European Sociological Review*, 19(1), 1–24.
- Blossfeld, H.-P. (2009). Educational assortative marriage in comparative perspective. *Annual Review of Sociology*, 35, 513–530.
- Blossfeld, H.-P., & Timm, A. (2003a). Educational systems as marriage markets in modern societies: A conceptual framework. In H.-P. Blossfeld & A. Timm (Eds.), *Who marries whom? Educational systems as marriage markets in modern societies* (pp. 9–18). Dordrecht: Kluwer Academic Publishers.
- Blossfeld, H.-P., & Timm, A. (2003b). Assortative mating in cross-national comparison: A summary of results and conclusions. In H.-P. Blossfeld & A. Timm (Eds.), *Who*

- marries whom? Educational systems as marriage markets in modern societies* (pp. 331–342). Dordrecht: Kluwer Academic Publishers.
- Borrill, P., & Tesfatsion, L. (2010). *Agent-based modeling: The right mathematics for the social sciences?* Working Paper No. 10023. Iowa State University, Department of Economics.
- Boudon, R. (1974). *Education, opportunity, and social inequality*. New York: Wiley & Sons.
- Boudon, R. (1979). *La logique du social*. Paris: Presses Universitaires de France.
- Brännström, L. (2008). Making their mark: The effects of neighbourhood and upper secondary school on educational achievement. *European Sociological Review*, 24(4), 463–478.
- Brauns, H., & Steinman, S. (1997). *Educational reforms in France, West-Germany, the United Kingdom and Hungary: Updating the Casmin educational classification*. Working Papers 21. Mannheim.
- Breen, R. (1999). Beliefs, rational choice and Bayesian learning. *Rationality and Society*, 11(4), 463–480.
- Breen, R. (Ed.). (2004). *Social mobility in Europe*. Oxford: Oxford University Press.
- Breen, R. (2009). Formal theory in the social sciences. In P. Hedström & B. Wittrock (Eds.), *Frontiers of sociology* (pp. 209–230). Leiden: Brill.
- Breen, R., & Goldthorpe, J. (1997). Explaining educational differentials: Towards a formal rational choice theory. *Rationality and Society*, 9(3), 275–305.
- Breen, R., & Jonsson, J. O. (2005). Inequality of opportunity in comparative perspective: Recent research on educational attainment and social mobility. *Annual Review of Sociology*, 31, 223–243.
- Breen, R., Luijckx, R., Müller, W., & Pollak, R. (2009). Nonpersistent inequality in educational attainment: Evidence from eight European countries. *American Journal of Sociology*, 114(5), 1475–1521.
- Breen, R., Luijckx, R., Müller, W., & Pollak, R. (2010). Long-term trends in educational inequality in Europe: Class inequalities and gender differences. *European Sociological Review*, 26(1), 31–48.
- Breen, R., & Yaish, M. (2006). Testing the Breen-Goldthorpe model of educational decision making. In L. Stephen, D. B. Morgan, D. Grusky & G. S. Fields (Eds.), *Frontiers in social and economic mobility*. Stanford, CA: Stanford University Press.
- Bruch, E., & Mare, R. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology*, 112(3), 667–709.
- Burke, M. A., & Heiland, F. (2006). The strength of social interactions and obesity among women. In F. C. Billari, T. Fent, A. Prskawetz & J. Scheffran (Eds.), *Agent-based computational modelling* (pp. 117–137). Heidelberg: Physica-Verlag HD.
- Centola, D., Eguiluz, V. M., & Macy, M. W. (2007). Cascade dynamics of complex propagation. *Physica A*, 374, 449–456.
- Centola, D., & Macy, M. W. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3), 702–734.
- Cheadle, J. E. (2008). Educational investment, family context, and children's math and reading growth from kindergarten through third grade. *Sociology of Education*, 81(1), 1–31.
- Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4), 370–379.
- Cobalti, A. (1989). A relative mobility table. A modest proposal. *Quality and Quantity*, 23(2), 205–220.
- Cobalti, A. (1992). Origine sociale e livello di istruzione: Un modello. *Polis*, 6(1), 117–145.

- Cohen-Cole, E., & Zanella, G. (2008). Unpacking social interactions. *Economic Inquiry*, 46(1), 19–24.
- Collins, R. (1979). *The credential society*. New York: Academic Press.
- Davies, R., Heinesen, E., & Holm, A. (2002). The relative risk aversion hypothesis of educational choice. *Journal of Population Economics*, 15(4), 683–713.
- Davison, A., & Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.
- DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology*, 116(4), 1234–1283.
- Duncan, G. J., & Murnane, R. (2011). *Whither opportunity? Rising inequality, schools, and children's life chances*. New York: Russell Sage.
- Durlauf, S. (1999a). The memberships theory of inequality: Ideas and implications. In E. Brezis & P. Temin (Eds.), *Elites, minorities, and economic growth*. Amsterdam: North Holland.
- Durlauf, S. (1999b). How statistical mechanics contribute to study of science. *Proceedings of the National Academy of Sciences*, 96(19), 10582–10584.
- Durlauf, S. (2001). A framework for the study of individual behaviour and social interactions. *Sociological Methodology*, 31(1), 47–87.
- Durlauf, S. (2002). The memberships theory of poverty: The role of group affiliations in determining socioeconomic outcomes. In S. Danziger & R. Haveman (Eds.), *Understanding poverty in America*. Cambridge: Harvard University Press.
- Durlauf, S. (2006). Groups, social influences, and inequality: A memberships theory perspective on poverty traps. In S. Bowles, S. Durlauf & K. Hoff (Eds.), *Poverty traps*. Princeton, NJ: Princeton University Press.
- Durlauf, S., & Cohen-Cole, E. (2004). Social interactions models. In K. Lempf-Leonard (Ed.), *Encyclopedia of social measurement*. Academic Press.
- Durlauf, S. N., & Ioannides, Y. M. M. (2010). Social interactions. *Annual Review of Economics*, 2, 451–478.
- Duru-Bellat, M., Kieffer, A., & Reimer, D. (2011). Les inégalités d'accès à l'enseignement supérieur: Le rôle des filières et des spécialités. *Une Comparaison entre l'Allemagne de l'Ouest et la France, Économie et Statistique* (433–434), 3–22.
- Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge: Cambridge University Press.
- Epstein, J. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton, NJ: Princeton University Press.
- Erikson, R., & Goldthorpe, J. (1992). *The constant flux: A study of class mobility in industrial society*. Oxford: Clarendon Press.
- Falch, T., & Massih, S. S. (2011). The effect of education on cognitive ability. *Economic Inquiry*, 49(3), 838–856.
- Fekjær, S. N., & Birkelund, G. (2007). Does the ethnic composition of upper secondary schools influence educational achievement and attainment? A multilevel analysis of the Norwegian case. *European Sociological Review*, 23(3), 309–323.
- Fischer, C. S., Hout, M., Jankowski, M. S., Lucas, S. R., Swidler, A., & Vos, K. (1996). *Inequality by design: Cracking the bell curve myth*. Princeton, NJ: Princeton University Press.
- Fowler, J. H., & Christakis, N. A. (2008). The dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study. *British Medical Journal*, 337, a2338.

- Gabay-Egozi, L., Shavit, Y., & Yaish, M. (2010). Curricular choice: A test of a rational choice model of education. *European Sociological Review*, 26(4), 447–463.
- Gambetta, D. (1987). *Where they pushed or did they jump? Individual decision mechanisms in education*. Cambridge: Cambridge University Press.
- Gintis, H., & Bowles, S. (2002). Intergenerational inequality. *Journal of Economic Perspectives*, 16(3), 3–30.
- Gintis, H., Bowles, S., & Osborne, M. (2001). Incentive-enhancing preferences: Personality, behavior, and earnings. *American Economic Review*, 91(2), 155–158.
- Goldthorpe, J. (2001). Causation, statistics, and sociology. *European Sociological Review*, 17(1), 1–20.
- Goldthorpe, J. H. (1996). Class analysis and the reorientation of class theory: The case of persisting differentials in education attainment. *The British Journal of Sociology*, 47(3), 481–505.
- Goodman, L. A. (1969). How to ransack social mobility tables and other kinds of cross classification tables. *American Journal of Sociology*, 75(1), 1–40.
- Goux, D., & Maurin, E. (2007). Close neighbours matter: Neighbourhood effects on early performance at school. *The Economic Journal*, 117(523), 1193–1215.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6), 1420–1443.
- Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1, 201–233.
- Granovetter, M. (1988). Threshold models of diversity: Chinese restaurants, residential segregation and the spiral of silence. *Sociological Methodology*, 18, 69–104.
- Granovetter, M., & Soong, R. (1983). Threshold models of diffusion and collective behavior. *Journal of Mathematical Sociology*, 9, 165–179.
- Haller, A. O., & Woelfel, J. (1972). Significant others and their expectations: Concepts and instruments to measure interpersonal influence on status aspirations. *Rural Sociology*, 37(4), 591–622.
- Harding, D., Gennetian, L., Winship, C., Sanbonmatsu, L., & Kling, J. (2011). Unpacking neighborhood influences on education outcomes: Setting the stage for future research. In G. J. Duncan & R. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 277–299). New York: Russell Sage.
- Hedstrom, P. (1998). Rational imitation. In P. Hedstrom & R. Swedberg (Eds.), *Social mechanisms. An analytical approach to social theory*. Cambridge: Cambridge University Press.
- Hedstrom, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge: Cambridge University Press.
- Hedström, P., & Bearman, P. (2009). What is analytical sociology all about? An introductory essay. In P. Hedström & P. Bearman (Eds.), *The Oxford handbook of analytical sociology* (pp. 3–24). Oxford: Oxford University Press.
- Hedström, P., & Ibarra, R. (2010). On the contagiousness of non-contagious behavior: The case of tax avoidance and tax evasion. In H. Joas & B. Klein (Eds.), *The benefit of broad horizons: Intellectual and institutional preconditions for a global social science* (pp. 315–336). Leiden: Brill.
- Hedström, P., Liu, K.-Y., & Nordvik, M. (2008). Interaction domains and suicides: A population-based panel study of suicides in the Stockholm metropolitan area, 1991–1999. *Social Forces*, 87(2), 713–740.

- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in America life*. New York: Free Press.
- Hillmert, S., & Jacob, M. (2003). Social inequality in higher education. Is vocational training a pathway leading to or away from university? *European Sociological Review*, 19(3), 319–334.
- Holm, A., & Jaeger, M. M. (2008). Does relative risk aversion explain educational inequality? A dynamic choice approach. *Research in Social Stratification and Mobility*, 26(3), 199–219.
- Ichou, M., & Vallet, L.-A. (2011). Do all roads lead to inequality? Trends in French upper secondary school analysed with four longitudinal surveys. *Oxford Review of Education*, 37(2), 167–194.
- Jackson, M. (Ed.). (2012). *Determined to succeed? Performance, choice and education*. Stanford, CA: Stanford University Press.
- Jackson, M., Erikson, R., Goldthorpe, J. H., & Yaish, M. (2007). Primary and secondary effects in class differentials in educational attainment: The transition to A-Level courses in England and Wales. *Acta Sociologica*, 50(3), 211–229.
- Jaeger, M. M. (2007). Economic and social returns to educational choices. Extending the utility function. *Rationality and Society*, 19(4), 451–483.
- Jonsson, J. O., & Erikson, R. (2000). Understanding educational inequality: The Swedish experience. *L'Année Sociologique*, 50(2), 345–382.
- Kaufman, R. L., & Schervish, P. G. (1987). Variations on a theme. More uses of odds ratios to interpret log-linear parameters. *Sociological Methods and Research*, 16(2), 218–255.
- Keller, S., & Zavalloni, M. (1964). Ambition and social class: A respecification. *Social Forces*, 43(1), 58–70.
- Koerselman, K. (2011). *Bias from the use of mean-based methods on test scores*, Swedish Institute for Social Research (SOFI). Working Paper 1/2011. Stockholm University.
- Kroneberg, C., & Kalter, F. (2012). Rational choice theory and empirical research. Methodological and theoretical contributions in Europe. *Annual Review of Sociology*, 38, 73–92.
- Law, A. M. (2007). *Simulation modeling and analysis*. New York: McGraw-Hill.
- Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *Bio Science*, 51(5), 341–352.
- Lucchini, M., Della Stella, S., & Pisati, M. (2010). *The weight of the genetic and environmental dimensions in the inter-generational transmission of educational success*. *European Sociological Review*. doi:10.1093/esr/jcr067.
- Lytinen, S. L., & Railsback, S. F. (2012). The evolution of agent-based simulation platforms: A review of NetLogo 5.0 and ReLogo. Proceedings of the fourth international symposium on agent-based modeling and simulation (21st European Meeting on Cybernetics and Systems Research [EMCSR 2012]). Vienna, Austria, April 2012.
- Manski, C. (1993a). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3), 531–542.
- Manski, C. (1993b). Identification problems in social sciences. *Sociological Methodology*, 23, 1–56.
- Manski, C. (2000). Economic analysis of social interaction. *Journal of Economic Perspectives*, 14(3), 115–136.
- Manzo, G. (2006). Generative mechanisms and multivariate statistical analysis. Modeling educational opportunity inequality by Multi-Matrix Log-Linear topological model: Contributions and limits. *Quality and Quantity*, 40(5), 721–758.

- Manzo, G. (2007). Variables, mechanisms, and simulations: Can the three methods be synthesized? A critical analysis of the literature. *Revue Française de Sociologie – An Annual English Selection*, 48(Suppl.), 35–71.
- Manzo, G. (2009). *La spirale des inégalités. Choix scolaires en France et en Italie au XX siècle*. Paris: Presses de l'Université Paris-Sorbonne.
- Manzo, G. (2010). Analytical sociology and its critics. *European Journal of Sociology*, 51(1), 129–170.
- Manzo, G. (2011). *Educational choices and educational traps. Towards an integration between computational and statistical modelling in the sociology of social stratification*. GEMASS Working Papers, GeWop 2011, no. 1.
- Mastekaasa, A. (2006). Educational transitions at graduate level: Social origins and enrolment in PhD programmes in Norway. *Acta Sociologica*, 49(4), 437–453.
- Mayer, T. (1960). The distribution of ability and earnings. *The Review of Economics and Statistics*, 42(2), 189–195.
- McPherson, J., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social isolation in America: Changes in core discussion networks over two decades. *American Sociological Review*, 71, 353–375.
- Merton, R. K. (1968). The Matthew effect in science. The reward and communication systems of science are considered. *Science*, 159(810), 55–63.
- Morgan, S. L. (2005). *On the edge of commitment: Educational attainment and race in the united states*. Stanford, CA: Stanford University Press.
- Mouw, T. (2006). Estimating the causal effect of social capital: A review of recent research. *Annual Review of Sociology*, 32, 79–102.
- Müller, W., & Karle, W. (1993). Social selection in educational systems in Europe. *European Sociological Review*, 9(1), 1–23.
- Müller, W., Lüttinger, P., König, W., & Karle, W. (1989). Class and education in industrial nations. *International Journal of Sociology*, 19(3), 3–39.
- Need, A., & de Jong, U. (2000). Educational differentials in the Netherlands: Testing rational choice theory. *Rationality and Society*, 13, 71–98.
- Orléan, A. (1995). Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion. *Journal of Economic Behavior and Organization*, 28(2), 257–274.
- Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review*, 24(5), 543–565.
- Raftery, A. E., & Hout, M. (1993). Maximally maintained inequality: Expansion, reform and opportunity in Irish education, 1921–1975. *Sociology of Education*, 66, 41–62.
- Railsback, S. F., & Grimm, V. (2012). *Agent-based and individual-based modeling: A practical introduction*. Princeton, CA: Princeton University Press.
- Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82, 609–623.
- Reynolds, J. R., & Johnson, M. K. (2011). Change in the stratification of educational expectations and their realization. *Social Forces*, 90(1), 85–110.
- Rolfe, M. (2009). Conditional choice. In P. Hedström & P. Bearman (Eds.), *The Oxford handbook of analytical sociology* (pp. 419–447). Oxford: Oxford University Press.

- Sampson, R., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing neighbourhood effects: Social processes and new directions in research. *Annual Review of Sociology*, 28, 443–478.
- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1, 143–186.
- Schelling, T. C. (1973). Hockey helmets, concealed weapons and daylight saving. A study of binary choices with externalities. *Journal of Conflict Resolution*, 17(3), 381–428.
- Schizzerotto, A. (1997). Perché in Italia ci sono pochi diplomati e pochi laureati? Vincoli strutturali e decisioni razionali degli attori come cause della contenuta espansione della scolarità superiore. *Polis*, 11(3), 345–365.
- Selz, M., & Vallet, L.-A. (2006). La démocratisation de l'enseignement et son paradoxe apparent (pp. *Données Sociales, La Société Française*, INSEE, Paris.
- Sewell, W. H., Hauser, R. M., Springer, K. W., & Hauser, T. S. (2003). As we age: A review of the Wisconsin longitudinal study, 1957–2001. *Research in Social Stratification and Mobility*, 20, 3–111.
- Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, 40(3), 211–239.
- Shavit, Y., & Blossfeld, H.-P. (Eds.). (1993). *Boulder, CO*. Westview Press.
- Shoam, Y., & Leyton-Brown, K. (2009). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge: Cambridge University Press.
- Skopek, J., Schulz, F., & Blossfeld, H.-P. (2011). Who contacts whom? Educational homophily in online mate selection. *European Sociological Review*, 27(2), 180–195.
- Snijders, T. A. B., & Steglich, C. E. G. (2013, forthcoming). Representing micro-macro linkages by actor-based dynamic networks models, *Sociological Methods and Research*.
- Sobel, M. E. (2006). Spatial concentration and social stratification: Does the clustering of disadvantage 'Beget' bad outcomes? In S. Bowles, S. N. Durlauf & K. Hoff (Eds.), *Poverty traps* (pp. 204–229). New York: Russell Sage Foundation.
- Stocké, V. (2007). Explaining educational decision and effects of families social class position: An empirical test of the Breen–Goldthorpe model of educational attainment. *European Sociological Review*, 23(4), 505–519.
- Stocké, V. (2008). *Educational decisions as rational choice? Testing the Erikson-Jonsson model*. Working Papers 504, Universität Mannheim.
- Stonedahl, F., & Wilensky, U. (2010). *Evolutionary robustness checking in the artificial Anasazi model*. Proceedings of the AAAI fall symposium on complex adaptive systems: Resilience, robustness, and evolvability. November 11–13, 2010. Arlington, VA.
- Thélot, C., & Vallet, L.-A. (2000). La réduction des inégalités sociales devant l'école depuis le début du siècle. *Économie et Statistique*, 334, 3–32.
- Tisue, S., & Wilensky, U. (2004a). NetLogo: Design and implementation of a multi-agent modeling environment. Evanston, IL, Center for Connected Learning and Computer-Based Modeling, Northwestern University. <http://ccl.northwestern.edu/papers/>
- Tisue, S., & Wilensky, U. (2004b). NetLogo: A simple environment for modeling complexity. Evanston, IL, Center for Connected Learning and Computer-Based Modeling, Northwestern University. <http://ccl.northwestern.edu/papers/>
- Vallet, L. A., & Selz, M. (2007). Evolution historique de l'inégalité des chances devant l'école: Des méthodes et des résultats revisités. *Education & Formation*, 74, 65–74.

- Van de Werfhorst, H. G. (2009). Credential inflation and educational strategies: A comparison of the United States and the Netherlands. *Research in Social Stratification and Mobility*, 27, 269–284.
- Van de Werfhorst, H. G., & Hofstede, S. (2007). Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared. *The British Journal of Sociology*, 58, 391–415.
- VanderWeele, T. J. (2011). Sensitivity analysis for contagion effects in social networks. *Sociological Methods & Research*, 40(2), 240–255.
- Watts, D. J. (1999). Networks, dynamics, and the Small-World phenomenon. *American Journal of Sociology*, 105(2), 493–527.
- Watts, D. J. (2004). The new science of networks. *Annual Review of Sociology*, 30, 243–270.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘Small-world’ networks. *Nature*, 393, 440–442.
- White, H. (2000). Parameterize: Notes on mathematical modeling for sociology. *Sociological Theory*, 18(3), 505–509.
- Winship, C., & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ in the bell curve. In B. Devlin, S. E. Fienberg, D. P. Resnick & K. Roeder (Eds.). *Intelligence, genes and success: Scientists respond to the Bell curve* (pp. 215–234, ch. 10). Springer-Verlag.
- Wooldridge, M. (2009). *An introduction to multi agent systems*. John Wiley and Sons, Ltd.
- Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review*, 57(3), 380–395.



## APPENDIX A

**Table A.1.** Number, State and Initialisation Values of the Model Parameters Concerned by the Simulation.

Parameter	Number of Independent Parameters	State and Initialisation Values
Mean of the log-normal distribution of agents' ability ( $A_g$ )	5	Manipulated
Variance of the log-normal distribution of agents' ability ( $\sigma_A^2$ )	1	Set to 0.25
Mean of the log-normal distribution of agents' perceived benefits ( $B_g$ )	5	Manipulated
Variance of the log-normal distribution of agents' perceived benefits ( $\sigma_B^2$ )	1	Set to 0.25
Network average degree ( $K$ )	1	Set to 4
Probability of out-group links ( $p$ )	1	Set to 0.1
Probability of out-group links as function of the 'Social' distance among groups ( $P_{GSD} \in \{1,2,3,4\}$ )	3	Set to 0.6, 0.25, 0.10, 0.05
Maximum number of times the educational level $L$ can be evaluated $F$	1	Set to 3

**Table A.2.** Percentage of Actors/Agents Reaching a Given Educational Level (Columns) within a Given Educational Background/Agent Group (Rows) – French Real-Data Cross-Tabulation (Average of 100 Bootstrapped Cross-Tabulations) and Simulated Cross-Tabulations Generated by Each Model Variant Discussed in Section ‘Computational Results, Empirical Data and Model Dynamic’ (Average Over 100 Replications).

	1a	1bc	2ab	2c	3ab	N
(a) French Empirical Data						
1a	48.90	20.38	14.33	7.21	9.17	1,418
1bc	15.34	24.39	25.53	16.22	18.51	1,769
2ab	7.66	8.96	23.80	19.60	39.98	524
2c	5.63	5.73	16.80	19.53	52.31	380
3ab	4.76	2.79	8.13	14.81	69.51	909
(b) ‘Straw Man’ Model: All Mechanisms Are Absent						
1a-agents	54.69	24.69	11.31	5.11	4.20	1,418
1bc-agents	54.67	24.88	11.19	5.14	4.12	1,769
2ab-agents	54.80	24.54	11.30	5.05	4.31	524
2c-agents	54.79	24.46	11.27	5.20	4.27	380
3ab-agents	54.38	25.03	11.22	5.11	4.26	909
(c) Model Variant 1: Ability						
1a-agents	44.73	24.43	13.26	8.42	9.15	1,418
1bc-agents	33.95	22.71	14.22	9.34	19.78	1,769
2ab-agents	29.47	21.57	13.16	11.32	24.47	524
2c-agents	31.10	21.56	14.17	12.56	20.61	380
3ab-agents	34.31	19.29	13.89	10.26	22.23	909
(d) Model Variant 2: ‘Model Variant 1’ + Perceived Benefits						
1a-agents	45.01	24.37	13.23	8.33	9.06	1,418
1bc-agents	34.51	22.82	14.14	9.25	19.28	1,769
2ab-agents	17.23	14.43	12.79	10.17	45.38	524
2c-agents	15.52	14.20	11.28	10.95	48.04	380
3ab-agents	14.36	10.31	11.34	7.62	56.36	909
(e) Model Variant 3: ‘Model Variant 2’ + Ability/Benefits Interaction						
1a-agents	45.52	24.43	13.07	8.21	8.77	1,418
1bc-agents	36.01	23.15	13.96	9.03	17.85	1,769
2ab-agents	17.06	14.35	12.25	9.92	46.42	524
2c-agents	14.24	12.64	10.59	9.93	52.60	380
3ab-agents	11.45	8.59	9.60	6.79	63.56	909
(f) Model Variant 4: ‘Model Variant 3’ + Social Influence						
1a-agents	37.67	26.26	16.41	10.45	9.21	1,418
1bc-agents	25.79	24.58	17.20	12.02	20.40	1,769
2ab-agents	12.73	14.98	14.19	14.37	43.73	524
2c-agents	11.29	13.09	12.84	14.63	48.14	380
3ab-agents	5.64	5.89	7.11	8.79	72.57	909
(g) Model Variant 5: ‘Model Variant 4’ + Inflation Spiral Mechanisms (see Appendix B)						
1a-agents	31.88	24.91	17.61	11.89	13.71	1,418
1bc-agents	18.11	20.95	16.52	13.13	31.29	1,769
2ab-agents	8.34	11.46	12.10	12.24	55.86	524
2c-agents	7.88	9.75	10.68	11.86	59.83	380
3ab-agents	4.10	4.60	5.25	6.17	79.88	909

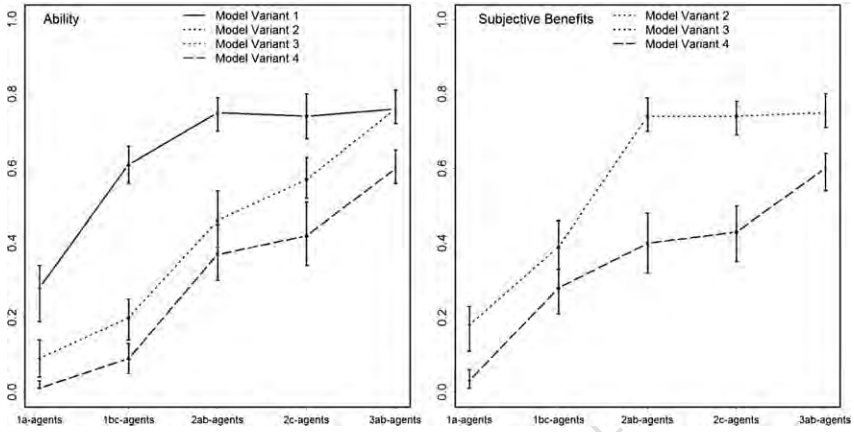


Fig. A.1. Effective Mean Values (Averaged over 100 Replications) of Agents' Ability (Left) and Perceived Benefits (Right) as a Function of Agents' Group for Each Model Variant (Bars Give Values within Which Fall 95% of Model Replication). Example: Under Model Variant 1, the Mean of Ability Distribution for Agents Belonging to the Lowest Group (1a-Agents) Ranges across Replications between  $\approx 0.20$  and  $\approx 0.37$ .

(C) Emerald Group Publishing Limited

## APPENDIX B

As I have stressed in the discussion section, the formal model presented here contains several simplifying assumptions. In particular, with regard to the way in which actors evaluate the net benefits of a given educational level, both its pay-offs on the job market and the actor's subjective evaluation of these pay-offs from the point of view of the social group to which s/he belongs are conflated into a single parameter,  $B_{ig}$  (see section 'Computational modelling of educational preference formation'; Eq. (2)).

On a conceptual level, however, separating these two components would be appreciable. Such analytical decomposition would allow more direct representation of the inflation spiral in which educational choices tend to be trapped in the long run. While an explicit representation of this process is absent from Breen and Goldthorpe's (1997) formal model, Goldthorpe's (1996) first discursive outline of the model explicitly acknowledged that when education is considered a positional good, its value decreases as more people obtain a given educational qualification, thus progressively modifying actors' subjective incentives to do their best to go beyond that qualification if they want to maintain their relative position in the social hierarchy (see already Boudon, 1974, ch. 8; for a historical outline of education inflation dynamics, see Collins, 1979). Recent empirical analyses have shown that individuals' expectations/ambitions about going beyond the upper secondary level increase in the long run (see Reynolds & Johnson, 2011) and that the inflation trend of higher educational certificates induces increasing proportions of all social classes to try and obtain them (see Van de Werfhorst, 2009).

Eq. (B.1) slightly modifies Eq. (2) (see section 'Computational modelling of educational preference formation') in order to accommodate in a simple way an explicit representation of this process of educational inflation and its effect on actors' beliefs. To this end, the two following terms are introduced into the original Eq. (2): (a) the term  $OR_{Lt}$  – that is, the pay-off from educational level  $L$  on the job market at the instant  $t$  (OR stands for objective returns) and (2) the term  $NC_{Lt}$  – that is, the overall number of choices that a given educational level  $L$  has received at the instant  $t$  from all agents living in the artificial society ( $N$  being the total number of artificial agents). These two terms are then made to interact in a specific way.

On the one hand, the diffusion of educational choices at the aggregate level is assumed to affect the agent's educational choice by progressively reducing the objective return  $OR_{Lt}$  of educational level  $L$ , thereby indirectly depressing the agent's evaluation of educational level  $L$ . However, on the

other hand, it is also assumed that artificial agents actively react to this inflation dynamic of educational levels by iteratively readjusting their initial subjective perception  $B$  of the benefits of the educational level  $L$ : that is, the larger the proportion of favourable choices for  $L$ , the more perceived benefits are upwardly reevaluated.

$$P_{igL} = A_{ig} + \frac{1}{\ln(NC_{Lt-n})} OR_{Lt} + (\phi(A) \times B_{ig})^{1 - \frac{NC_{Lt-n}}{N}} + SI_{igL} \quad (B.1)$$

Thus, the main difference between the model studied in the article and the generalisation presented here is the introduction of a second source of externalities among agents' educational choices, which instead of being based on the educational choices performed by agents' closest contacts, derive from the educational behaviour of the population of agents considered as a whole ( $NC$ ). As the detailed simulation algorithm clarifies (see Fig. 1, step 2.2), however, since the total number of choices  $NC_{Lt}$  in favour of each educational level is computed, and put back into the agent's education behavioural function [B.1], every 52 iterations, the two forms of social externalities do not operate on the same temporal scales. While the network-based social influence is continuously at work, the population-based feedbacks triggered by the diffusion of education levels are activated only discretely.

According to Goldthorpe's (1996) verbal analysis, if dynamic belief updates of this kind were at work under the pressure of a process of education devaluation, one should expect to find that, *ceteris paribus*, class differentials in educational choices persist because the educational race becomes increasingly intense for every social group.

In order to test this hypothesis, exactly the same set of simulations that enabled best approximation of the educational stratification empirically observed in France – that is, the model containing all the mechanisms postulated (see section 'Computational results, empirical data and model dynamic'; Table 2f) – were re-run by introducing the two additional mechanisms formalised by Eq. (B.1). To initialise the four parameters of which the term  $OR$  consists, I took the proportion of French respondents reaching the service class at the end of their occupational careers (for the CASMIN occupational schema adopted here, see Breen, 2004, p. 12) among those reaching one of the five educational levels – thus representing the objective return associated with a given educational level as the capacity of this level to ensure a place among the most advantageous social positions.

The results reported in Table 2g (see section 'Computational results, empirical data and model dynamic') confirm this expectation. The statistics

computed on the simulated data show that, when artificial agents are assumed to constantly upwardly reevaluate their initial appreciation of education benefits as a function of the devaluation of educational level, while the overall proportion of agents reaching the highest educational level increases (the proportion of upwardly agents is higher than in the model without the 'inflation' spiral; see also Appendix A, Table A.2g), the overall level of inequality of educational opportunity, as measured by the average generalised odds ratios, is virtually unchanged and so are the coefficients expressing the relative positions of the highest and the lowest group of agents (see columns 'agor', 'hggor', 'lggor', respectively).

That said, I would stress the exploratory value of the test performed. Rigorous study of the generative power of the two mechanisms representing the inflation spiral and its effect on individuals' belief updating would require testing the computational model against a distribution of empirical cross-tabulations covering a series of cohorts instead of using a cross-sectional cross-tabulation. This would certainly be a step forward in developing the present study.