European Sociological Review, 2017, Vol. 33, No. 1, 124–136
doi: 10.1093/esr/jcw054
Advance Access Publication Date: 21 December 2016
Original Article

Late School Tracking, Less Class Bias in Educational Decision-Making? The Uncertainty Reduction Mechanism and Its Experimental Testing

Joël Berger[1,2,*] and Benita Combet[3,*]

[1]Department of Sociology, Utrecht University, The Netherlands, [2]Institute of Sociology, University of Zurich, Switzerland and [3]Institut des Sciences Sociales, LIVES—Swiss National Centre of Competence in Research, University of Lausanne, Switzerland

*Corresponding author. Email: j.berger@uu.nl; benita.combet@unil.ch
Note: Shared first authorship (the ordering of names has been determined randomly).

Submitted March 2016; revised October 2016; accepted October 2016

Abstract

A long-standing hypothesis in the sociology of education is that the timing of ability tracking impacts the inequality of educational opportunity. While earlier studies mainly focused on how early tracking impacts the primary effect of social origin (systematic performance differences due to social background), the impact of early tracking on the secondary effect of social origin (class-specific educational decision-making) has been neglected. Recently, the idea has been put forward that late tracking decreases uncertainty in educational decision-making, thus enabling more rational decision-making by lower-class individuals. Extending this idea, we propose the uncertainty reduction mechanism (URM) as a theoretical foundation; this mechanism can be derived from a decision-theoretic model on educational decision-making based on prospect theory. Moreover, we perform a first empirical test of the URM by means of a computerized laboratory experiment. The evidence is in line with the model predictions and the results support the intuition that a postponement of tracking could reduce the negative bias of lower-class individuals in educational decision-making, thereby reducing educational inequality with respect to social background.

Introduction

Inequality of educational opportunity between social classes is a phenomenon known in virtually all societies (Breen and Jonsson, 2005). Students from higher social classes not only perform better on average than those from lower social classes (primary effect of social origin, see Boudon, 1974), but also when controlling for performance, those from higher social classes have a higher probability of continuing with school after compulsory schooling and have a higher probability of attending school tracks with more demanding curricula (secondary effect of social origin; e.g. various contributions in Jackson, 2013). As a result, those from higher social classes obtain higher educational credentials, which tend to yield higher labour-market returns (Becker, 1964).

Nevertheless, the connection between social background and educational attainment varies considerably across countries, suggesting that national educational policies could impact this connection (Van de Werfhorst and Mijs, 2010). A recent study shows that there is much more variation in the magnitude of secondary effects of social origin than in primary effects (Jackson and Jonsson, 2013). As an explanation, the authors 'expect transitions taken at older ages to be less susceptible to secondary effects because the amount of information available to students about their own abilities and their chances of success at higher levels of education must only increase as their exposure to formal education increases' (Jackson and Jonsson, 2013: p. 332).

In other words, Jackson and Jonsson pick up a long-standing hypothesis in sociology of education which suggests that late tracking reduces the influence of social background on educational attainment.[1] Moreover, they suggest that the timing of tracking not only impacts primary effects but also secondary effects; namely, the postponement of tracking reduces uncertainty and thus enables a more rational decision to be made about whether to invest in higher education.

While the hypothesis that late tracking reduces uncertainty has some intuitive appeal, it is not derived from a theory nor is it grounded in empirical evidence. Also, it remains unclear why specifically individuals from lower social classes in comparison to those from higher classes should profit from a reduction in uncertainty—which necessarily must be the case if uncertainty reduction indeed reduces the secondary effects of social origin. In other words, any explanation of why late tracking reduces inequality must consider why the potential factor affects students from a lower social class differently than students from a higher social class (in statistical terms, this means an interaction effect between social class and timing of tracking).

In this article, we specify the uncertainty reduction mechanism (URM) in a model for educational decision-making based on prospect theory (Tversky and Kahneman, 1979, 1981), which was originally put forward by Page (2005). The model suggests that specifically high-performing students from lower social classes should profit from a late timing of tracking, while it should impact the decisions of students from higher social classes to a lesser degree. Consequently, a late timing of tracking should decrease secondary effects and thus educational inequality with respect to social origin.

We additionally provide a first test of the URM by means of a laboratory experiment. Such an experiment is well suited to a first test of a theory since it makes it possible to test whether a hypothetical causal mechanism does exist and is in play at least under ideal conditions (Loomes, 1989; Hey, 1991). It should be noted that we do not extrapolate conclusions from the laboratory to the real world. Rather, we test experimentally one out of several possible theoretical mechanisms that can potentially explain behaviour in the field.[2] For the question at hand, a laboratory experiment is specifically well suited because it allows us to isolate the causal effect of one single aspect of educational systems (the timing of decision) on decision-making. In contrast, using observational data almost always implies the possibility of a spurious correlation (be it cross-sectional or over time; Shadish, Cook and Campbell, 2002). Specifically, research on the impact of educational systems is error-prone for the following reasons. On the one hand, the relevant aspects of educational systems are usually correlated (for instance, there are hardly any educational systems with a high number of tracks and a late onset of tracking). As such, there is usually not enough variance to disentangle the effects of the different potentially relevant features of educational systems. On the other hand, the design of educational systems is also correlated with other potentially influential factors, such as welfare policy (e.g. there is no social-democratic welfare state with an educational system characterized by a high number of tracks and an early timing of tracking; c.f. Goldthorpe, 2000; Jackson and Jonsson, 2013).

Theoretical Framework

We specify the URM in a prospect theory framework. In the first part, we outline how class-specific biases in educational decision-making derive from an individual's social position in the hierarchy of social stratification. In the second part, we show how such biases can be moderated by uncertainty—a product of the timing of tracking. In the third part, we derive testable hypotheses.

Status Maintenance and Class-Specific Risk Preferences

Based on a seminal contribution by Boudon (1974), class-specific decision-making is thought to be one of the main reasons for class-specific differentials in participation in higher education. More specifically, when people decide whether to invest in higher education, they are primarily concerned with the so-called 'status maintenance motive'; that is, with maintaining their socio-economic status. Put differently, the higher an individual's socio-economic status, the more education he or she requires for status maintenance. For example, the offspring of a higher-class background (e.g. a higher-grade professional) must successfully obtain a tertiary educational credential to maintain his or her social class position, while a working-class individual does not need tertiary education but only vocational training to maintain his or her social class position. Models based on the idea of status maintenance have been put forward by Erikson and Jonsson (1996) and Breen and Goldthorpe (1997) amongst others.

In a similar vein, Page (2005) derives the motive for status maintenance from a general theory of human decision-making, namely, prospect theory (Kahneman and Tversky, 1979). Prospect theory is based on the idea that, when individuals make a decision, they evaluate the potential outcomes in comparison to a reference point—for example, the status quo. A central feature of this theory is that people react more extremely to a negative deviation from the reference point than to a positive deviation of the same degree (e.g. Novemsky and Kahneman, 2005; Abdellaoui, Bleichrodt and Paraschiv, 2007; Tom *et al.*, 2007). Another of the theory's central features is that it claims that actors decide in a risk-avoiding manner when facing a potential gain and in a risk-seeking manner when trying to avoid a potential loss (Tversky and Kahneman 1979, 1981)—an empirically well-established behavioural pattern (see Kühberger, 1998 for a meta-analysis). The Page model integrates the motive for status maintenance into a prospect theory framework, assuming that an individual's class position is his or her reference point when taking an educational decision. In short, social class is associated with whether students are in a loss frame or a gain frame. Higher classes are in a loss frame—they need to pursue longer school careers to avoid a loss of social status relative to their parents. Lower social classes are in a gain frame, as they gain status if they spend longer in schooling than their parents.

For analytical reasons, we assume the existence of two social classes only, the *lower class* and the *higher class*. We also assume the existence of only two educational tracks, *basic education* (preparing for vocational training and an early labour-market entry) and *higher education* (as a requirement for tertiary education). We further assume that lower-class individuals can reproduce their social status when completing basic education, while upper-class individuals need to successfully complete higher education to reproduce theirs.

Having completed basic education, an actor faces the decision of whether to invest in higher education. For all actors, the benefit B_{HS} of successfully completing higher education (i.e. the corresponding labour-market returns in the forms of money and prestige) is greater than the benefit of basic education, B_B. However, completing higher education requires a certain level of performance and therefore there is a risk of failure. The benefit from an unsuccessful attempt to invest in higher education (B_{HF}) is the worst outcome possible. On the one hand, an unsuccessful actor suffers a loss in terms of direct costs, opportunity cost, time and psychological costs (i.e. a negative impact on his or her self-esteem). On the other hand, incomplete higher education could be interpreted as a negative signal in the labour market. Denoting the probability of successfully completing higher education as P and the converse probability of failure as $1–P$, the educational investment decision can be conceived as a choice between an alternative with one certain outcome B_B (basic education) and an alternative with two uncertain outcomes (a successful or unsuccessful investment in higher education, B_{HS} and B_{HF} respectively). The latter can be represented by a lottery (B_{HS}, P; B_{HF}, $1–P$), with $B_{HS} > B_B > B_{HF}$.

Recall that in a prospect theory framework, actors evaluate the alternatives in their choice sets differently depending on their reference points. Technically, this is represented by a value function V, which is assumed to be S-shaped: convex in the domain of losses and concave in the domain of gains. Hence, the alternatives can be stated as $V(B_B)$ and $V(B_{HS}, P; B_{HF}, 1–P)$, respectively. Higher-class members overvalue the lottery relative to the certain outcome, while the opposite holds for the lower class. Therefore, higher-class members, facing a possible loss, are risk seeking; lower-class members, facing a possible gain, are risk averse (for empirical evidence that social class position is associated with risk preferences, see Breen, van de Werfhorst and Jaeger, 2014).

Page *et al.* (2007) implemented the discussed binary decision situation in a laboratory experiment. Participants could decide to invest in a lottery, in which the probability of success depends on their performance. As predicted, they found that individuals invest in a lottery with a higher probability when trying to avoid a loss compared to when perceiving the same decision in a gain frame.

Timing of Tracking and Uncertainty

How could the education system impact class-specific educational decision-making? As pointed out by Jackson and Jonsson (2013), in educational decision-making, a crucial factor is uncertainty and a major source of uncertainty is performance uncertainty (e.g. Breen and Goldthorpe, 1997; Breen, van de Werfhorst and Jaeger, 2014). For a successful completion of higher education, a certain performance level has to be maintained over the time span of schooling, and at any given point in time, an actor cannot be certain that he or she will be able to uphold that level of performance in the future. When deciding whether to invest in higher education, an actor thus generates a subjective probability of successfully completing higher education P, which primarily depends on his or her performance. When estimating his or her probability of successfully completing higher education, an actor's past performance level is thus likely to be the central source of information.

It is obvious that the earlier the timing of tracking (i.e. the timing of the decision on whether to attend higher education), the less information about his or her performance—on which an actor's estimate of his or her success probability P is based on—an actor has at hand. In an education system where the decision for higher education has to be made early (e.g. after 4–6 years, as in Germany or Switzerland), this P-value is based on only half the amount of data on the actor's performance, as compared to when the same decision has to be made after completion of compulsory school (e.g. in Finland or Sweden). Moreover, the earlier a prediction is made, the longer the time span it must hold. Note that this does not mean that on average, the success probability P is greater in a school system such as the Finish or Swedish one. Rather, the prediction of one's success probability P is sounder in a system with late tracking. To clarify this point, it is useful to distinguish between *risk* (uncertainty about an outcome; e.g. the lower the success probability P, the greater the risk of failure when deciding for higher education) and *ambiguity* (uncertainty about a probability; e.g. the less information there is to estimate P, the greater the ambiguity in P). Here we use the term *uncertainty* as an umbrella term for both risk and ambiguity (see Camerer and Weber, 1992 for an overview of different conceptions of uncertainty, risk and ambiguity). A simple model of ambiguity is proposed by Kahn and Sarin (1988), who suggest that the less reliably a success probability can be estimated, the more decision-makers discount it ('ambiguity aversion'):

$$\pi(P) = \mathrm{E}(P) – \lambda\sigma^2. \quad (1)$$

The expectation $\mathrm{E}(P)$ is an actor's best estimator of his or her probability of success and σ^2 is the variance in this expectation. The less information the actor has at hand when estimating P, the greater the variance (ambiguity) σ^2 and thus the less reliable the estimate of P. An actor will discount his or her subjective probability P more with greater ambiguity σ^2 and greater individual ambiguity aversion λ.

But in what way could ambiguity resulting from an early timing of tracking impact the decisions of lower-class individuals differently from the decisions of higher-class individuals? It has been shown experimentally that individuals in a gain frame (lower-class individuals) are not only risk averse (they dislike low P-values), but are also more ambiguity averse (they dislike large variance in P-values) than those in a loss frame (upper-class individuals) (Cohen, Jaffray and Said, 1985; Einhorn and Hogarth, 1986; Hogarth and Einhorn, 1990). In other words, on average, λ is greater for those deciding in a gain frame than for those deciding in a loss frame.

We thus rewrite the above lottery as:

$$\mathrm{V}(B_{\mathrm{HS}}, \pi[P]; B_{\mathrm{HF}}, 1 – \pi[P]). \quad (2)$$

To summarize, the earlier the decision for higher education and thus the greater the ambiguity, the more actors discount their subjective success probability P. Put differently, when the decision is postponed, the negative bias in the success probabilities diminishes. But this effect will not affect the decision of all actors equally. One the one hand, ambiguity aversion is stronger for those in a gain frame (those from the lower class) than for those in a loss frame (those from the higher class). On the other hand, the poorly performing actors will hardly change their educational decision when their success probability increases on a low level. Specifically, high-performing individuals with an overly negative view of their success probability (i.e. those in a gain frame) will continue with education when ambiguity is reduced.

Having stated all the effects that we assume to be in play when actors decide whether to invest in higher education (whether to continue with Level 2 in the laboratory experiment), we can now state our hypotheses:

H1: *High-performing individuals in a gain frame decide with a higher probability to continue with Level 2 when the timing of the decision is late rather than early.*

We expect this behaviour because later tracking reduces ambiguity in the subjectively estimated success probability, and individuals in a gain frame (i.e. from the lower social class) are specifically averse towards ambiguity. However, only high-performing individuals with a sound objective success probability will adapt their decision when their negative bias is reduced; those with a genuinely bad performance will hardly change their mind:

H2: *Low-performing individuals in a gain frame do not have a higher probability of deciding to continue with Level 2 when the decision has to be taken late.*

H1 is a precondition for the URM to indeed reduce educational inequality. Only if a late timing of the educational decision increases the probability of lower-class individuals taking part in higher education, can educational inequalities be reduced. Additionally, a second condition has to be met: only if a late timing of the educational decision increases the probability for higher-class individuals to a lesser degree than the probability for the lower class, is educational inequality actually reduced:

H3: The probability to continue with Level 2 does not differ between the early and the late-decision treatment for individuals in a loss frame, independently of their performance.

H3 derives from the assumption that individuals in a loss frame are less concerned with their performance or success probability than those in a gain frame since the former are less ambiguity averse than the latter. Therefore, the timing of the decision should not, or should only slightly, impact decisions in the loss frame treatment.

Should, as predicted, the skilled individuals in a gain frame have a higher probability of continuing with Level 2 in the late-decision treatment (H1), while generally individuals in a loss frame are only slightly affected by the timing of the decision (H3), it is a consequence that the disparities in decision-making of both frames are reduced, at least when focusing on the high-performing individuals only:

H4: Among the high-performing, the difference in the proportion of individuals in a loss frame that continue with Level 2 and the proportion of high-performing individuals in a gain frame that continue is smaller when the timing of decision is late rather than early.

Should the ambiguity reduction brought about by a late timing of the decision be strong enough, we should find the empirical patterns stated in hypotheses H1 and H4 not only for the high-performing individuals, but also for the whole sample:

$H1_{overall}$: Generally, individuals in a gain frame have a higher probability of deciding to continue with Level 2 when the timing of the decision is late rather than early.

$H4_{overall}$: Generally, the difference in the proportion of individuals in a loss frame that continue with Level 2 and the proportion of individuals in a gain frame that continue is smaller when the timing of decision is late rather than early.

Experimental Design

Subjects, Design and Procedures

Our aim is not to draw an inference from the laboratory experiment with respect to the real world. Rather, we aim to test whether the URM is, at least, observable under ideal conditions. Our experimental design is based on Page *et al.* (2007); however, we adapted it to our purposes and thus created the simplest design possible for testing the URM hypothesis. As such, the design must account for the factors the theory identifies as relevant: frame (gain vs. loss) and timing of tracking (early vs. late). In addition, a performance task is needed, since the theory states that a longer time span of estimating their performance allows decision-makers to make more reliable estimates of the probability of success (i.e. their performance level). We refrained from implementing several transitions and several tracks that would have made the design somewhat more realistic, following a methodological rule that the simplest design that allows a postulated causal effect to be isolated should be used. The structure of the design is discussed more thoroughly in the Online Supplementary Material (OSM) (Section 2), where we also provide a table that compares aspects of the design with their real-world analogues.

The experiment was implemented using *z-Tree* (Fischbacher, 2007) and conducted at the Decision Science Laboratory at ETH Zurich. Participants were students from ETH Zurich and the University of Zurich. All participants ($N = 165$, 51% female) received a show-up fee of CHF 15 (approx. 14 Euro at the time of experimentation). Additionally, they could win up to CHF 22 depending on their individual performance and their decision. After completing the experimental task, the participants completed a questionnaire. Screenshots of the experimental instructions and tasks, as well as more information on the questionnaire, can be found in the OSM (Sections 4 and 6).

Taken together, we have a 2x2 between-subjects design with the factors being the timing of the decision and the frame, as summarized in Table 1.

In the experiment, the participants solved 12 sets of anagrams (an anagram is a jumble of letters that, when ordered properly, results in a word); each set consisted of eight anagrams. These anagrams had to be solved under increasing time pressure, starting with 120 seconds in the first set and decreasing in steps of 2 seconds per set. The increasing time pressure mitigates practice effects and represents an aspect of educational systems; namely, that the difficulty of the curriculum increases. As such, increasing time pressure increases ambiguity. Four (*early-decision condition*) or eight (*late-*

Table 1. Overview of factors and case numbers ($N = 165$)

		Frame	
		Gain	Loss
Timing of decision	Early	40	41
	Late	42	42

decision condition) sets of anagrams constituted Level 1. After Level 1 was completed, the participants received feedback about the mean number of anagrams they had correctly solved in Level 1. An average of four anagrams per set had to be reached to pass Level 1. Subjects who did not manage to pass received the lowest payoff δ. Those participants who passed could decide whether to continue. When they decided not to continue, they received an intermediate payoff β. When they continued and, again, achieved at least an average of four anagrams correctly solved, they received a high payoff α. When they continued and failed to reach the required threshold, they received a low payoff γ (therefore $\alpha > \beta > \gamma > \delta$). Participants knew that when they did not participate, they would solve anagrams as well, but that in this case, their performance would not affect their payoff in Level 2.

As in the study by Page and colleagues (2007), we induced a loss frame or a gain frame in that the payoffs were presented either as a loss (*loss frame condition*) or as a gain (*gain frame condition*). Specifically, participants in a loss frame started with CHF 22 (α) and could lose everything but CHF 6 (δ). Conversely, participants in a gain frame started with 6 (δ) and could win up to CHF 16, so that they would also have CHF 22 in the end (α).[3] The payoffs presented as gains are listed in Figure 1; the payoffs presented as losses are listed in Figure 2. Figure 1 represents the early-decision treatment; Figure 2 represents the late-decision treatment. These two figures, together with the other two figures, representing the other two conditions, were also provided to the participants. The originals were written in German (translation by the authors; see OSM for an overview of all four figures).

Performance

We use one direct measure of performance and additionally include two proxies to increase robustness. First, we measured performance as the average number of anagrams solved in Level 1 (*ability*). We defined high-ability individuals as those in the upper quartile and low ability individuals as those in the three lower quartiles. Secondly, the participants were asked whether, and if so, how often, they train in solving anagrams at home, for example, by playing Scrabble (*training at home*). They were defined as trained when they performed an activity related to solving anagrams at least once a year (a more detailed measurement was not possible due to the skewed distribution of the variable). Training at home is a proxy for performance in that it can be expected that the trained will tend to perform better in solving anagrams (on average, they solved half an anagram more per set, $P = 0.002$; regression table not shown). Thirdly, we implemented experimental training for half of the participants (*experimental training*) because it is well known that measuring behaviour with retrospective questions is only a rough approximation of real behaviour (Bertrand and Mullainathan, 2001). In each session, half the participants had the opportunity to solve eight anagrams under time pressure, immediately before the actual experiment started (training). Meanwhile, the other half of the participants solved math problems (no training). Participants were allocated randomly to be either trained or not.

Table 2 lists the distribution of the high-performing and low-performing participants for the three measures.

Experimental Evidence

Of all 165 participants, 137 (83 percent) passed Level 1. Out of these 137 participants, 36.5 percent (30.3 percent of the overall sample) decided to continue (to attend higher education).

We analyse the data with logistic regression models, where the dependent variable indicates if participants decided to continue with Level 2 (1 = yes, 0 = no). The analyses of the decision on whether to continue are based on the 137 participants who passed Level 1. We did not include the participants who failed to complete Level 1 because they could not decide whether to continue. We use one-sided tests of the estimates if we had a directed hypothesis (as stated in the hypotheses section); in the other cases, we use two-sided tests.

Before testing our hypotheses, we analyse the main effects of the experimental factors ‘frame’ and ‘timing of decision’ as well as our three measures for skill. A strong main effect of framing is a precondition to finding a potential mediating effect of timing of the decision, while the theory remains silent regarding a main effect of timing. As listed in Table 3, Model 1, we find a strong and statistically significant ($P = 0.028$) main effect of the frame treatment factor: participants in a loss frame continue with a 17.7 percentage point higher probability

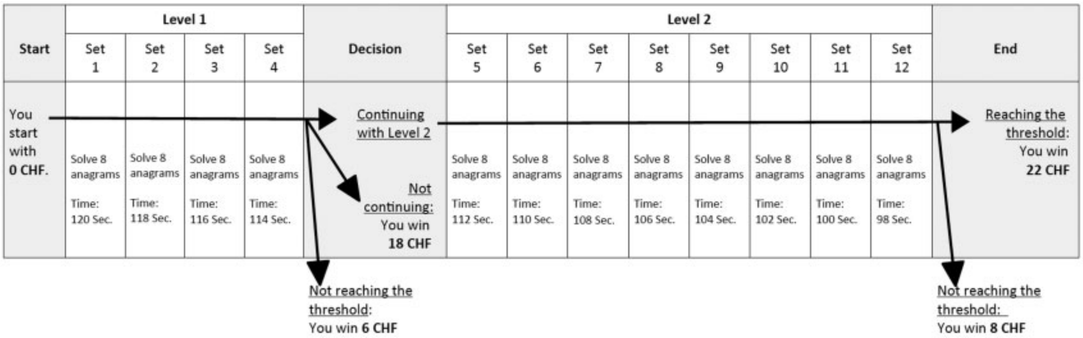

Figure 1. Early decision/gain frame condition

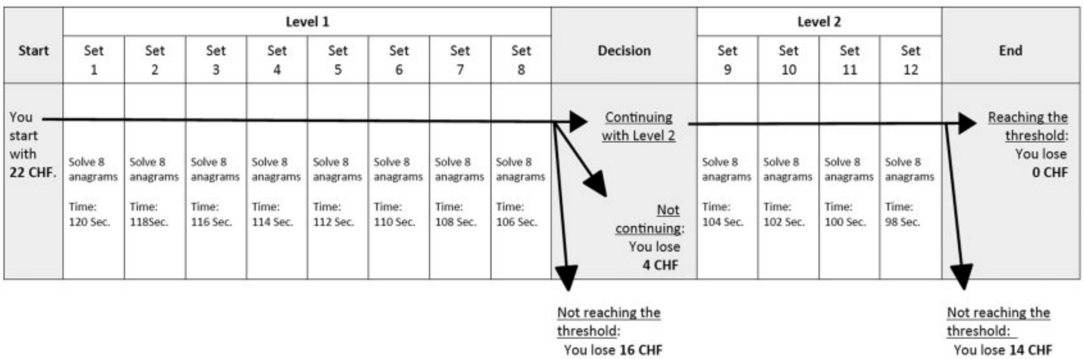

Figure 2. Late decision/loss frame condition

Table 2. Overview of factors and case numbers ($N = 165$)

		Number of anagrams solved L1	Frame	
			Gain	Loss
		Performance		
Timing of decision	Early	$Nr \geq Q_{0.75}$	14	11
		$Nr < Q_{0.75}$	26	30
	Late	$Nr \geq Q_{0.75}$	9	12
		$Nr < Q_{0.75}$	33	30
		Training at home		
Timing of decision	Early	Yes	10	10
		No	30	31
	Late	Yes	10	16
		No	32	26
		Experimental training		
Timing of decision	Early	Yes	23	24
		No	17	17
	Late	Yes	22	22
		No	20	20

than those in a gain frame. This framing effect is considerably stronger than the one found in the original study (Page, Garboua and Montmarquette, 2007). In the original study, there were three decisions and the participants in a loss frame exhibited a higher probability, of approximately 0 (first decision) to 9 (third decision) percentage points, of continuing than those in the gain frame. Concerning the second treatment factor, we do not find a significant main effect of timing of the decision ($P = 0.36$), which means that roughly the same numbers of individuals choose to continue in both experimental conditions. Considering the influence of the performances measures on the probability to continue, we see moderate to strong and statistically significant effects of 'number of anagrams solved' and 'experimental training' but no statistically significant effect for 'training at home'.

Table 4 lists predictive margins for all the combinations of the two treatment factors (frame and timing of decision). The first column lists the overall results. For each of the performance measures, the results are then

Table 3. Logit models of the main treatment effects, including performance. Dependent variable: Probability to continue, individuals who passed Level 1

	M1	M2	M3	M4	M5
Loss frame	0.177*	0.163*	0.170*	0.177*	0.164*
(Ref: Gain frame)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)
Late decision	0.026	0.071	0.071	0.029	0.084
(Ref: Early decision)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Number of anagrams solved L1		0.188***			0.194***
		(0.04)			(0.04)
Training at home			0.117		0.045
(Ref: yes)			(0.09)		(0.08)
Experimental training				0.125+	0.152*
(Ref: yes)				(0.08)	(0.08)
Pseudo R2	0.027	0.127	0.037	0.040	0.151

Notes: Average marginal effects/discrete change effects reported.
$N = 137$, standard errors in parentheses.
$^{+}P < 0.1$; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

Table 4. Probability to continue, predictive margins

		Overall	Number of anagrams solved L1		Training at home		Experimental training	
			Nr. < $Q_{0.75}$	Nr. ≥ $Q_{0.75}$	No	Yes	No	Yes
Early decision	Gain frame	0.249 *(0.074)*	0.164 *(0.084)*	0.364 *(0.126)*	0.351 *(0.088)*	0.202 *(0.126)*	0.221 *(0.112)*	0.267 *(0.099)*
	Loss frame	0.457 *(0.084)*	0.315 *(0.097)*	0.742 *(0.129)*	0.345 *(0.097)*	0.665 *(0.195)*	0.321 *(0.125)*	0.557 *(0.113)*
Late decision	Gain frame	0.302 *(0.078)*	0.168 *(0.076)*	0.661 *(0.161)*	0.208 *(0.082)*	0.624 *(0.176)*	0.272 *(0.112)*	0.328 *(0.108)*
	Loss frame	0.440 *(0.079)*	0.452 *(0.098)*	0.419 *(0.142)*	0.491 *(0.105)*	0.352 *(0.127)*	0.372 *(0.119)*	0.490 *(0.105)*

Notes: Controlled for sessions, standard errors in parentheses.

listed separately for the low performers and the high performers. Tables A1 and A2 list the overall results of Wald tests for statistical differences between the predictive margins.

The first hypothesis states that the late-decision treatment should increase the willingness to continue with Level 2 among the high-performing participants in a gain frame because ambiguity is reduced in the late-decision treatment. Indeed, among the high-ability participants (solving anagrams) in a gain frame, the predicted probability of continuing is almost twice as high in the late-decision treatment (66.1%) as compared to the early-decision treatment (36.4%, $P = 0.087$, one-sided test). A similar effect is found for training at home (20.2% vs. 62.4%, $P = 0.074$, one-sided test) but not for experimental training (26.7% vs. 32.8%, $P = 0.339$, one-sided test). Experimental training turned out to affect decision-making only weakly ($P = 0.128$, see Table 3, Model 4) and to not affect ability at all (beta = 0.185, $P = 0.300$, regression table not shown)—solving eight anagrams seems to add little to pre-existing skills. Taken together, the evidence supports H1. It also becomes evident that although the effects are large, they often reach statistical significance only on the 10% level. This suggests that the statistical power is relatively low. Still, since the effect predicted by H1 can be found two out of three specifications of performance, these results clearly support H1.

We also find evidence for H2, which states that low-performing participants in a gain frame do not have a higher probability of deciding to continue education with Level 2 in the late-decision treatment. Even if, due to the late decision, ambiguity is reduced and thus the negative bias of the subjects in the gain frame is diminished, those with a genuinely low success probability will not continue with a higher probability. As predicted, in all three specifications of performance, the gain frame/early treatment condition does not differ significantly from the gain frame/late treatment condition (ability in solving anagrams: $P = 0.4851$, one-sided test; training at home: $P = 0.3251$, one-sided test; experimental training: $P = 0.3752$, one-sided test).

In comparison to participants in a gain frame, participants in a loss frame are predicted to be less concerned with their performance and success probability since they are risk seeking. They are also assumed to be less affected by ambiguity than those in a gain frame. According to H3, we would therefore expect that their probability of continuing does not differ much between the early and the late tracking treatment, independently of their performance level. We find evidence supporting H3. Neglecting performance level, we do not find that the probability of continuing differs between participants in the loss frame of the early- and the late-decision treatment ($P = 0.889$, two-sided test). Accounting for performance level, we find only a significant difference for high-performing participants in the loss frame (ability in solving anagrams: $P = 0.1204$, two-sided test; training at home: $P = 0.0617$, two-sided test; experimental training: $P = 0.6647$, two-sided test). Interestingly, the high performers in a loss frame show a lower tendency to continue in the late-decision treatment in comparison to the early-decision treatment, which is not predicted by the theory but does not affect our main findings either. In summary, we can conclude that the timing of the decision does not impact decisions in the loss frame treatment, with the exception mentioned above. This means that ambiguity aversion does not have much impact on the loss frame treatment, as predicted.

Given H1, H2, and H3, it follows that high-performing participants in a gain frame have a higher probability of continuing with Level 2 in the late-decision system, while high-performing participants in a loss frame are only slightly affected by the timing of the decision (H4). Although difference-in-difference tests would be the appropriate statistical approach, we do not have enough statistical power to conduct such tests. Still, what we do find for all three specifications of performance is a statistical difference between high-performing participants in a gain frame and in a loss frame in the early-decision treatment, and an absence of such a difference in the late-decision treatment (early-decision treatment: ability in solving anagrams: $P = 0.0312$, training at home: $P = 0.0091$, experimental training: $P = 0.0341$; late-decision treatment: ability in solving anagrams: $P = 0.142$, training at home: $P = 0.2086$, experimental training: $P = 0.1477$. One-sided tests reported). Thus we conclude that the data supports H4.

Should the ambiguity reduction brought about by a late decision be strong enough, we should find the empirical patterns stated in hypotheses H1 and H4 for the whole sample as well. However, while we find the predicted patterns, the effects are only weak. As stated in $H1_{overall}$, participants in a gain frame continue slightly more often in the late-decision treatment than in the early-decision treatment (difference $= 0.053$, $P = 0.3100$, one-sided test). Considering $H4_{overall}$, the overall ambiguity reduction effect is rather weak (difference-in-difference test, $P = 0.3200$, one-sided test). This is because, as discussed above, this effect works only for high performers in a gain frame but not for the low performers, which mitigates the effect in the overall sample.

In a nutshell, the empirical pattern as a whole concurs well with the hypotheses, and some of the effects are reasonably strong, even though not all reach conventional levels of statistical significance. However, one should consider that, although the overall case number of 137 participants (counting only those who manage to complete Level 1) is quite large, when performing subgroup analyses, the case numbers reduce. It is thus unsurprising that not all of the predicted effects reach statistical significance. However, we use three different specifications of performance and the results are fairly robust across specifications.[4]

Discussion

A long-standing hypothesis in the sociology of education is that late school tracking reduces educational inequality with respect to social background. A possible explanation was put forward by Jackson and Jonsson (2013), namely, that late tracking reduces uncertainty because the earlier a decision whether to participate in higher education has to be made, the less information actors have at their disposal when estimating their probability of successfully completing higher education (the URM). Given that individuals from lower social classes are more averse to uncertainty (i.e. ambiguity) than those from higher social classes they invest in higher education with a higher probability when decisions are postponed. Meanwhile, timing of tracking impacts those from a higher social class to a lesser degree. As a result, the disparity in educational decision-making—the secondary effect of social origin—between the social classes should be reduced when decisions are made late.

This should be specifically pronounced for high-performing students of lower social classes because those with a genuinely low performance would also abstain from investing in further education in the absence of a negative bias in their probability of success.

To test this hypothesis, we conducted a computerized laboratory experiment, based on an experimental framework suggested by Page *et al.* (2007). Under this protocol, individuals solve anagrams under time pressure. The risk-seeking preferences of 'high-social class' individuals are induced by putting them in a loss frame, while the risk-averse preferences of 'low-social class' individuals are induced by putting them in a gain frame. After a certain number of anagrams are solved, the participants decide whether or not to continue. Only when participants continue and solve enough anagrams can they win (gain frame) or avoid losing (loss frame) the full amount of money. We replicated the Page experiment, additionally varying the number of anagrams participants solved before deciding whether to continue to induce the early- and late-decision situation. As predicted, high-performing participants in a gain frame are more inclined to continue when decisions are made late rather than early (i.e. when they can solve a greater number of anagrams before deciding and thus performance uncertainty is reduced), while those in a loss frame are hardly affected by the timing of the decision. As a result, among the high-performing participants, the decisions between the two framing treatments (gain frame vs. loss frame) differ when they are made early, while they do not differ when they are made late. As predicted, these effects do not come into play for low-performing individuals in a loss frame and thus, in the overall sample, the effects are too weak to be statistically significant. We can thus conclude that the URM cannot only theoretically explain how early tracking could reduce class-specific biases in educational decision-making, but is also corroborated by the results from the experiment.

A limitation of this article is that the proposed mechanism via URM is not the only one that could plausibly explain inequality-reducing effects of education systems. Other (mutually not exclusive) explanations are myopia (Lucas, 2001) or time discounting preferences (Breen, van de Werfhorst and Jaeger, 2014). Here we have focused on one of these possible explanations, and leave it to future research to investigate further explanations.

Future research should replicate our experiment, possibly using larger case numbers. Simultaneously, further features of educational systems could be implemented. For example, in the current study, participants with a very low performance dropped out of 'school' and therefore they could not decide whether to continue. A straightforward extension would be to leave the decision on whether to continue open to individuals who did not reach the predetermined performance threshold. This would not only increase the case number but also mimic an educational system in which teacher recommendations are not binding. Further, we assessed performance by measuring the number of anagrams solved, using a verbal question (training at home), and we implemented experimental training to enhance performance. However, experimental training turned out to have only a weak impact on real performance. We thus suggest experimentally implementing performance by varying the difficulty of the experimental task in a future study (e.g. the maximum number of anagrams to be solved could be lower for high performers).

Notes

1 A brief literature review of previous research concerning social class-specific educational inequality due to variations in the start of tracking can be found in the OSM.

2 We are aware that there might exist several other possible mechanisms that could influence students' decision-making besides the reduction in uncertainty. One additional mechanism which could explain this interaction is that parents' influence decreases at later transitions, which would especially enable students from a lower social class to make decisions by reference to other sources besides their parents, like their peers. One can assume that this applies to a lesser degree to students from a higher social class due to their (and their parents') loss aversion (e.g. Mare 1980). However, we want to stress that our contribution is analytical rather than descriptive. As such, we do not claim that we deliver (and experimentally test) a comprehensive theory of class differentials in educational decision-making and their interplay with the timing of decisions. Rather, we try to isolate one single mechanism. In our opinion this is necessary before attempting to build a comprehensive theory involving several complementing (or even counteracting) mechanisms.

3 The payoffs per stage could in principle also interact with the frame because the payoffs per set (but not the payoff per stage) differ between the early and the late-decision treatment (4 CHF for four or eight sets of anagrams, respectively). To completely rule this alternative explanation out, one would have to replicate the experiment

with equal payoffs per stage (and different total payoffs). However, we argue that the explanation provided in the article is more plausible because, as the results show, the interaction effect between frame and timing of decisions for high performers is exactly the pattern predicted by the URM, while the alternative explanation would predict a main effect (more individuals would continue in the late-decision treatment) rather than an interaction between frame and timing of decision.

4 Further, we also conducted sensitivity analyses with other thresholds to define low- and high-ability performers in solving anagrams and the results are reasonably robust for hypotheses 1–3, while they are less robust for hypothesis 4 but still fairly convincing given the overall evidence. For further information, see OSM, Section 5.

Acknowledgements

We thank five anonymous reviewers, Hartmut Esser, Ben Jann, Walter Müller, the participants of the Rational Choice Sociology Workshop 2015 in Venice, and the participants of the sociology colloquium, University of Bern in autumn 2015, for their helpful comments. Special thanks to Fabiana Koller for her excellent assistance with developing the experimental design. The experiment was designed and conducted by Berger and Combet, Berger developed the theoretical framework, while Combet provided the statistical calculations.

Funding

This work was supported by the Swiss National Science Foundation [grant number P0BEP1_148904].

Supplementary Data

Supplementary data are available at *ESR* online.

References

Abdellaoui, M., Bleichrodt, H. and Paraschiv, C. (2007). Loss aversion under prospect theory: a parameter-free measurement. *Management Science*, **53**, 1659–1674.

Becker, G. (1964). *Human Capital*. New York: Columbia University Press.

Bertrand, M. and Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *The American Economic Review*, **91**, 67–72.

Boudon, R. (1974). *Education, Opportunity, and Social Inequality. Changing Prospect in Western Society*. Neuwied: Luchterhand.

Breen, R. and Goldthorpe, J. (1997). Explaining educational differentials. Towards a formal rational action theory. *Rationality and Society*, **9**, 275–305.

Breen, R. and 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., van de Werfhorst, H. G. and Jaeger, M. M. (2014). Deciding under doubt. A theory of risk aversion, time discounting preferences, and educational decision making. *European Sociological Review*, **30**, 258–270.

Camerer, C. and Weber, M. (1992). Recent developments in modeling preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty*, **5**, 325–370.

Cohen, M., Jaffray, J. Y. and Said, T. (1985). Individual behavior under risk and under uncertainty: an experimental study. *Theory and Decision*, **18**, 203–228.

Einhorn, H. J. and Hogarth, R. M. (1986). Decision making under ambiguity. In Hogarth R. M., and Reder M. W. (Eds.), *Rational Choice. The Contrast Between Economics and Psychology*. Chicago: University of Chicago Press.

Erikson, E. and Jonsson, J. O. (1996). Explaining class inequality in education: the Swedish test case. In Eriksson R. and Jonsson J. O. (eds.), *Can Education be Equalized? The Swedish Test Case in Comparative Perspective*. Boulder: Westview Press, 1–63.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **2**, 171–178.

Goldthorpe, J. H. (2000). *On Sociology. Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.

Hey, J. D. (1991). *Experiments in Economics*. Oxford: Blackwell.

Hogarth, R. M. and Einhorn, H. J. (1990). Venture theory: a model of decision weights. *Management Science*, **36**, 780–803.

Jackson, M. (2013). *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford: Stanford University Press.

Jackson, M. and Jonsson, J. O. (2013). Inequality of opportunity across countries. In Jackson M. (Ed.), *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford: Stanford University Press, pp 306–337.

Kahn, B. E. and Sarin, R. K. (1988). Modeling ambiguity in decision under uncertainty. *Journal of Consumer Research*, **15**, 265–272.

Kahneman, D. and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, **47**, 263–291.

Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organizational Behavior and Human Decision Processes*, **75**, 23–55.

Loomes, G. (1989). Experimental economics. In Hey J. D. (Ed.), *Current Issues in Microeconomics*. New York: St. Martin's Press, pp. 152–178.

Lucas, S. R. (2001). Effectively maintained inequality: education transitions, track mobility, and social background effects. *American Journal of Sociology*, **106**, 1642–1690.

Novemsky, N. and Kahneman, D. (2005). The boundaries of loss aversion. *Journal of Marketing Research*, **XLII**, 119–128.

Page, L. (2005). Des inégalités sociales aux inégalités scolaires. *Choix Éducatifs Et Prospect Theory. Revue Économique*, **56**, 615–624.

Page, L., Garboua, L. L. and Montmarquette, C. (2007). Aspiration levels and educational choices: an experimental study. *Economics of Education Review*, **26**, 748–758.

Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont: Wadsworth.

Tom, S. M. *et al*. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, **315**, 515–518.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, **211**, 453–458.

Van de Werfhorst, H. G. and Mijs, J. J. B. (2010). Achievement inequality and the institutional structure of educational systems. A comparative perspective. *Annual Review of Sociology*, **36**, 407–428.

Joël Berger has studied Sociology and Education in Bern and is now a visiting researcher in Utrecht and a postdoctoral researcher in Zurich. His main fields of interest lie in competition and social inequalities, social cooperation, and experimental sociology. Recent publications: The Sanctioning Dilemma (*European Sociological Review*, 2016, with W. Przepiorka); Norm Enforcement in the City revisited (*Rationality and Society*, 2016, with D. Hevenstone); The Logic of Relative Frustration (*European Sociological Review*, 2015, with A. Diekmann).

Benita Combet has studied Sociology, Science of Religion and Islamic and Middle Eastern Studies at University of Bern and is now senior researcher at University Lausanne with the NCCR LIVES. Her main research interests are social inequalities with a focus on inequality in education, social mobility and gender inequality in the labour market.

Appendix

Table A1. *P*-values of Wald tests for statistical differences between the predictive margins in Table 4

Performance	Difference tests between… (frame/decision)	Hypotheses	*P*-values
Overall	Gain/early vs. loss/early	$H4_{overall}$	*0.0370*
	Gain/late vs. loss/late	$H4_{overall}$	*0.1114*
	Gain/early vs. gain/late	$H1_{overall}$	*0.3100*
	Loss/early vs. loss/late	H3	0.8890

Note: Directed hypotheses were tested one-sided (the corresponding *P*-values are printed in italics).

Table A2. *P*-values of Wald tests for statistical differences between the predictive margins in Table 4

Performance	Difference tests between… (frame/decision)	Hypotheses	*P*-values skills measured as…		
			Solving anagrams	Training at home	Experimental training
Low	Gain/early vs. loss/early		0.2656	0.6107	0.5550
	Gain/late vs. loss/late		0.0371	0.0478	0.5432
	Gain/early vs. gain/late	H2	*0.4851*	*0.3251*	*0.3752*
	Loss/early vs. loss/late	(H3)	0.3265	0.2726	0.7699
High	Gain/early vs. loss/early	H4	*0.0312*	*0.0091*	*0.0341*
	Gain/late vs. loss/late	H4	*0.1420*	*0.2086*	*0.1477*
	Gain/early vs. gain/late	H1	*0.0865*	*0.0739*	*0.3388*
	Loss/early vs. loss/late	(H3)	0.1204	0.0617	0.6647

Note: Directed hypotheses were tested one-sided (the corresponding *P*-values are printed in italics).