Citation for published version:
Mead, R, Hejmadi, M & Hurst, LD 2018, 'Scientific aptitude better explains poor responses to teaching of evolution than psychological conflicts', Nature Ecology & Evolution, vol. 2, no. 2, pp. 388-394. https://doi.org/10.1038/s41559-017-0442-x

DOI:
10.1038/s41559-017-0442-x

Publication date:
2018

Document Version
Peer reviewed version

Link to publication

This is the Authors' Accepted Manuscript of an article published in final form at: https://dx.doi.org/10.1038/s41559-017-0442-x

Scientific aptitude better explains poor responses to teaching of evolution than psychological conflicts

REBECCA MEAD[1], MOMNA HEJMADI[1], LAURENCE D. HURST[1]

The Milner Centre for Evolution, Department of Biology and Biochemistry, University Of Bath, Bath BA2 7AY, UK

EMAIL: L.D.HURST@BATH.AC.UK

It is considered a "myth" that non-acceptance of scientific consensus on emotive topics is owing to difficulties processing scientific information and is, instead, owing to belief-associated psychological conflicts, the strongest non-acceptors being highly educated. Do these results from adults explain variation in response to school-level teaching? We studied a cohort of UK secondary school students (ages 14-16) and assessed their acceptance and understanding of evolution. In addition, to address their aptitude for science we assessed their understanding of genetics and their teacher-derived assessment of science aptitude. As both models predict, students with low initial evolution acceptance scores showed lower increase in evolution understanding. Contra to conventional wisdom, this effect is better explained by lack of aptitude: before teaching, students with low acceptance had lower understanding of both evolution and of genetics; the low acceptance students sat disproportionately in the foundation (rather than higher) science classes; low acceptance students showed lower increments in genetics understanding; student gain in evolution understanding correlated positively with gain in genetics understanding. We find no evidence either for a role for psychological conflict in determining response to teaching or that strong rejectors are more commonly higher ability. From qualitative data we hypothesise that religious students can avoid psychological conflict by adopting a compatibilist attitude. We conclude that there exist students recalcitrant to the teaching of science (as currently taught) and that these students are more likely to not accept the scientific consensus. Optimizing methods to teach the recalcitrant students is an important avenue for research.

Why do some people reject the scientific consensus on certain subjects (e.g. vaccines, evolution and climate change)? Convention holds that strongly held beliefs about a subject, religiously or politically motivated, prohibit effective understanding[1] of that subject owing to psychological conflicts. This can be owing to cognitive dissonance[2,3], a desire to hold the same beliefs as those with whom we have ties[4], or avoidance of damage to perception of self worth[5,6]. When such denial or selective adsorption of the evidence[7] is commonplace, establishment of an unbiased understanding[4] is likely to be difficult. Such effects could explain the negative relationship between religion and scientific literacy[1]. That prior non-empirical world-views (i.e. beliefs) colour the processing of information that conflicts with that world-view is not, however, unique to one demographic: US Democrats believing that the surge in Iraq didn't work don't process well evidence to the contrary, while Republicans don't fairly process climate change evidence[5].

An alternative possibility is that the rejection of scientific consensus reflects a general inability to process complex arguments and evidence, or a deficit in knowledge[8-11]. However, at least in adults, the most vehement science-deniers tend to be highly educated[12-14]. Indeed, the notion that those who don't accept the scientific understanding are those who struggle to understand the science, has been described as one of the "myths[15]" of public understanding of science. However, Pew research, for example, report that as regards the question of whether humans are the product of evolution, an increasing proportion of individuals agree with the scientific view as science education attainment levels increase[16]. This could, however, reflect avoidance of science education owing to psychological conflicts.

Understanding of the relative roles of poor understanding and psychological conflicts comes in large part from studies on adults (and predominantly in the US). Do these results transfer to the secondary school classroom (in the UK)? In classroom teaching of science there will be students who are less able to process scientific information as usually taught[8]. Is this associated with low acceptance as well? Conversely, might prior beliefs restrict learning outcomes and might this be especially acute for some academically more able students[12-14]? That one can teach the mechanics of evolution to students, whether they accept or reject the scientific consensus[17], suggests that the prior beliefs need not always be a hurdle. N.B. for the distinction between belief and acceptance[18] see Supplementary note 1.

Does then aptitude or psychological conflict best predict student responses to teaching of contentious subjects? We address this issue in the specific context of the teaching of evolution to a UK-based cohort (number of students=1227, number of classes = 70) of secondary school children (ages 14-16). The schools were derived from both the state and private system and comprised a large breadth of social, religious and economic demographics[19]. Teachers were blinded to the aim of this study. For further details see Methods and prior paper[19].

Evolution as a subject is known to be difficult to teach for multiple reasons[2,8,20-22]. Two aspects are important in the current context. First, the concepts within evolution are hard and abstract[8]. Second, some people have a prior non-acceptance of the scientific view of evolution[2,17]. Non-acceptance includes both those actively rejecting the scientific consensus and those undecided.

Both psychological conflict[2,20] and aptitude models predict that a student's improvement in understanding of evolution through instruction would be predicted by their degree of acceptance of evolution prior to teaching. The aptitude model proposes that some students struggle with science, possibly owing to poor logical reasoning skills[23,24], and so are confused about evolution, a confusion that results in both poor understanding and poor acceptance[25-27]. This model thus also predicts that the ability to improve understanding of a less emotive but related subject will also be predicted by the acceptance of evolution prior to tuition. By contrast, the psychological conflict model predicts that a student's prior rejection of the scientific view of evolution should not predict their ability to understand the less emotive subject.

"Less emotive" in this context can mean one of two things: either that the subject matter is uncontentious or that any debate is uncorrelated with the possible belief-based foundations of the non-acceptance of evolution. Here we employ basic genetics understanding (DNA, mutation, Mendelism etc.) as that less emotive but related subject. Fundamental genetics is a good comparator, it being uncontentious, abstract but still a close intellectual relative of evolution. Further, aspects of genetics considered to be contentious (notably genetic modification) are uncorrelated with political or religious belief systems[28]. The same is not true for evolution, climate change, stem cell biology or the big bang, where religious/political stance correlate[28].

Here then we ask whether a student's degree of acceptance of evolution, prior to being taught evolution, predicts their pre- and post- teaching understanding of evolution alone or genetics as well. We also ask whether teacher-assessed general science ability predicts pre-teaching acceptance of evolution. If evidence from adults translates to school children, the conflict model also predicts that amongst the highest ability students there exists a discrete and larger subpopulation of low acceptors[4,12-14]. We thus ask whether acceptance levels prior to teaching accord with science ability as classified by teachers, and whether in higher ability classes we see evidence for an especially large subpopulation of low acceptors.

RESULTS

Students with low prior acceptance of evolution have lower prior understanding of evolution and of genetics

Consistent with both models, students with low prior acceptance of evolution have lower understanding of evolution prior to formal teaching (rho = 0.22, $P=9 \times 10^{-15}$; Fig 1). The aptitude model in addition predicts that pre-teaching acceptance will also predict pre-teaching understanding of genetics. This is indeed the case (rho=0.43, $P<2.2 \times 10^{-16}$). Similarly, a lower understanding of genetics is correlated with a lower understanding of evolution (rho=0.20, $P=3 \times 10^{-12}$).

Students with low prior acceptance of evolution are more common in foundation science classes

The general aptitude model also predicts that the students with low acceptance of evolution prior to teaching will be those of lower "ability"[4]. Classes were stratified (by teachers) into those doing foundation level GCSE versus those of higher ability (this being across all sciences, not just biology). Higher ability students indeed have a higher acceptance of evolution prior to teaching (Mann Whitney U test, $P=9 \times 10^{-11}$; Fig 2). Similarly, the higher ability students had higher genetics understanding (Mann Whitney U test, $P=1.7 \times 10^{-15}$) and evolution understanding (Mann Whitney U test, P=0.03), prior to teaching. These results suggest that the students with lower acceptance of evolution may tend to be lower ability students with lesser understanding of science more generally, of which evolution understanding is but one component. This contrasts, for example, with evidence regarding climate change denial in adults[4].

No evidence for a larger subpopulation of rejectors in the higher ability classes

While the above indicates that higher ability students tend to accept evolution more than foundation students, if some academically more capable people are more likely to adopt strong anti-science positions (as is the conventional wisdom[4,12-14]), we expect to see evidence of more strongly divided opinions in the higher ability class. Divided opinions should be reflected in a tendency to bimodality in the distribution of acceptance scores and a higher frequency of low acceptors.

An efficient measure of deviation from unimodality is the dip[29] score (lower dip scores are more unimodal). As the score is sensitive to sample size, we subsample from the 1055 higher ability students a random 172 students (the size of the foundation population). The median dip score of 10,000 random subsamples is 0.0407 (95% CI 0.0348-0.0509), identical to the dip score of the foundation class. After teaching, the unimodality of acceptance scores is also not significantly different (post-teaching median dip of higher class subsamples = 0.0407, 95% CI 0.0343-0.0494, dip of foundation class= 0.0349). The higher and foundation classes thus have the same (negligible) deviation from unimodality.

The frequency of evolution rejectors is also no different between higher and foundation classes. The percentage of students with a preteaching acceptance score ≤ 32 (the cut-off for "low" acceptance[30]) are the very similar in the higher and foundation classes (1.04% in higher ability, 1.16% in the foundation classes: Fisher's exact test, $P=0.70$, odds ratio $=0.90$). The same applies after teaching (1.2% in high ability, 1.16% in foundation class: Fisher's exact test, $P=1$, odds ratio $=1.06$). We conclude that we find no evidence for a greater polarization in acceptance, or for a greater frequency of strong evolution rejectors, when ability is high.

Students with low acceptance of evolution before teaching respond poorly to evolution teaching

A prediction of both models is that the students with initial lower acceptance of evolution are less receptive to evolution teaching. The fact that students with low acceptance also have lower ability and lower understanding prior to teaching introduces a statistical difficulty, in so much as, owing to a ceiling effect, a student's preteaching score in evolution understanding by necessity is negatively correlated with their absolute change in score: rho=-0.53, $P < 2 \times 10^{-16}$. We correct the change in understanding of evolution scores by considering the residuals of the loess regression of change in understanding of evolution versus preteaching understanding scores (Fig 3). These residuals scores do not correlate with preteaching understanding of evolution scores (rho=-0.018, P=0.52) and thus may be considered a normalised measure of response to evolution teaching. As expected of a discriminating measure, these residuals are higher for students in the higher ability class (Mann Whitney U test, P=0.013; median higher ability = -0.076; median foundation ability =-0.17).

Employing this normalised measure, we find that low initial acceptance predicts a poorer response to teaching (rho=0.17, $P=4.6 \times 10^{-9}$; Fig 4). Previously we showed that students taught genetics before evolution respond better than those taught evolution then genetics[19]. Does a student's initial acceptance level predict responses in both cohorts? We find that it does and similarly is observed in the higher ability classes and the foundation classes (Table 1). In a multivariate analysis in which normalized increase in evolution understanding is predicted by pre-teaching evolution acceptance, teaching order and ability, we find that all but ability are significant predictors (preteaching acceptance, estimate=0.02, $P=5 \times 10^{-7}$; order, estimate= 0.33, $P=1.5 \times 10^{-5}$; Ability, estimate=0.16 P=0.11, adjusted R^2=0.041).

The acceptance-gain correlation is robust to class effects

In the above analyses, we are considering all students in all classes *en masse*. Do we find that controlling for possible class, cohort or teacher effects we still find that pretesting acceptance levels predict the normalised increment in evolution understanding? We find that the correlation seen *en masse* is seen also within classes (Supplementary Table 1), supporting the hypothesis that students with low prior acceptance also have lower normalized gain in understanding of evolution, even when just compared against their class mates. In addition, this result indicates that differences in the time interval between pre- and post- testing do not explain the acceptance-gain correlation.

Poor response of low acceptance students is not specific to evolution understanding

Is the poor response to teaching of evolution of low acceptance students associated with a low responsiveness to teaching of science more generally or evolution in particular? To address this, we ask whether a student's preteaching acceptance of evolution predicts their response to the teaching of genetics. We consider the residuals of the loess regression of change in genetics score predicted by initial genetics score (which are not correlated with preteaching genetic scores: rho=-0.005, P=0.87) and consider these a normalized response to teaching of genetics. This response to the teaching of genetics is also predicted by the prior acceptance of evolution (rho=0.15, P=6.4 x 10^{-8}; Fig 5). The effect is seen when controlling for between-class effects (Supplementary Table S2). It is also seen for students doing genetics first (rho=0.16, P=4 x 10^{-6}) and those doing evolution first (rho = 0.10, P=0.04), for those in the higher ability group (rho = 0.11, P=0.0002) and those in the foundation group (rho=0.29, P=0.0001). In a multivariate analysis, ability (estimate 0.75, P=0.012), pre-teaching acceptance (estimate 0.079, P=1 x 10^{-6}) and order (estimate 1.07, P=5.5 x 10^{-7}) are all significant predictors of the normalised improvement in genetics understanding. Addition of the normalised change in evolution understanding shows it too to be a predictor (estimate 0.47, P=4.9 x 10-90), with adjusted R^2 = 0.075 (all predictors remain significant).

Consistent with the recalcitrance of students who don't accept evolution being owing to them having a general difficulty in learning about science, students who make larger gains in understanding evolution make larger gains in understanding genetics (rho=0.2, P=5.3 x 10^{-13}). This is also seen when we consider the correlation on a class-by-class basis (Supplementary Table 3).

No evidence for a role for psychological conflict

Above we have considered an extreme version of the psychological conflict model in which gain in genetics understanding is predicted to have no correlation with preteaching acceptance of evolution. A more nuanced model supposes that the relationship between preteaching evolution acceptance and normalized gain in evolution understanding has a steeper slope than that between preteaching evolution acceptance and normalized gain in genetics understanding. Only at the limit, if conflict were never an issue, would the latter slope be zero. A viable normalized metric of Relative Conflict Strength (RCS) can be:

RCS = [slope of evolution response – slope of genetics response]/evolution response

where, for direct comparability, the slopes are derived from a regression of the data (normalized genetics improvement, normalized evolution improvement, preteaching acceptance) expressed in deviation from mean in standard deviation units i.e. Z scores. Strikingly, a unit difference in standard deviation in preteaching acceptance scores translates to an identical (to two significant figures) 0.16 s.d. increment in both normalised genetics and normalised evolution understanding, thus giving RCS of zero. Note that the slope is nonetheless quite shallow.

Analysis of the correlations supports a similar conclusion. The correlation between preteaching acceptance and evolution gain is rho =0.167, while for genetics this is 0.154. The difference between these two is not significantly different (*P*=0.73, NPMCS). Partial correlation tests support the same conclusion (Supplementary results 1). Teaching order also has no effect (Supplementary results 2), arguing against possible cognitive conflicts being carried over when evolution is taught first.

These results all suggest that psychological conflict has little to no involvement in the poor response to evolution teaching in low accepting students and that the aptitude model is more viable.

No evidence for teacher non-acceptance or poor understanding

Student experience can also be conditioned on teacher non-acceptance[21] or reluctance to teach evolution[31]. While above we have controlled for by-class effects, it is helpful also to recognize that in our UK based setting we found no evidence that teacher non-acceptance was a serious issue. We find that 96% of 123 teachers are classified as accepting of evolution, 3% are unsure and 1% would be classified as rejectors. We also find little or no evidence for poor teacher understanding. Most teachers were specialist biology teachers with 72% having a degree in a biology-related subject. Their understanding of evolution and of genetics was fairly uniformly high. Over 65% of teachers answered all questions on evolution correctly and over 70% in genetics. The core concepts of evolution were well understood, with 79% of teachers recognising that evolution involves genetic changes in time. However, on more nuanced aspects there was some small degree of confusion. A notable minority (11%) considered that evolution involves the change from simple to complex organisms and there was confusion as to when life first appeared on earth.

DISCUSSION

Here we have considered two models regarding the possible causes of failure to accept scientific consensus. In the psychological conflict model, prior belief of the lack of correctness of the scientific explanation preconditions people to being unable to fairly process information pertinent to that emotive issue. In the alternative model, the prior non-acceptance is part of a nexus of low aptitude. In contrast to conventional wisdom[4,12,13,15], we find evidence strongly supporting the aptitude model and no evidence to support the conflict model, even in its more nuanced form. Moreover, and again in contrast to the accepted view[4,12-14], we find no evidence that strong rejectors are predominantly of higher educational attainment. Our results thus suggest that it is not a "myth"[15] that non-acceptance of scientific consensus is connected to knowledge and aptitude.

Why don't we see evidence for psychological conflict?

Why might we not be seeing evidence that psychological conflicts condition student learning? One possibility is that there is no conflict, the other is that conflicts are being avoided. An absence of conflict could come about if young pre-college students' attitudes/beliefs are yet to be fully resolved. This could explain why other studies, employing adults, find that cognitive conflicts, e.g. on climate change denial (e.g. [5]) and vaccine denial (e.g. [32]), are important. Adults will have had longer to embed their belief systems into a more coherent framework (e.g. a conspiracy theory view, see[33], but see also correction[34] and critique[35]). If the problem is a clash between evidence and an embedded belief system, then we might expect the more plastic developing belief systems of young adults to be less of an impediment to learning. Whether this is true for highly proscribed religious-based assertions about evolution is, however, less clear. Nonetheless, it may well be true that psychological conflicts explain much vehement science denial in adults, while at the same time science aptitude plays a deeper role in the developing brain.

In this context, an important caveat to our study is that it was performed on UK school students. The general level of acceptance of evolution is here high. The MATE tool[36] recommends to classify a person as accepting of evolution if they have a score of 46 or more. Under this classification, 78% accept prior to teaching, increasing to 85% after teaching, with only ~1% falling in the "reject" classification, the others being undecideds. This compares with the general adult population in the US where only 65% of respondents agree with the statement that humans evolved over time and 31% believe that humans have existed in their present form since the beginning of time[16]. Assuming a pressure to believe what an in-group believe[4], the pressure to accept the scientific consensus on evolution in the UK school context, even for religious students, is most likely stronger than in the US school system (or comparable low acceptance countries e.g. Turkey).

Might there be some value in the notion that students can avoid psychological conflict? To provide hypotheses to explain why conflict was not evident we assembled qualitative data via focus groups (N=76 students). These suggests that conflict may be being avoided by religious students in particular adopting a compatibilist intellectual stance, wherein acceptance of both religion and evolution is considered viable (Supplementary results 3, Supplementary figure 1). This possibility is worth further research, not least because it suggests simple interventions to help religious students learn about evolution.

A further possibility is that our measure of conflict-free academic progression is misleading. We have presumed that genetics is a suitable non-emotive control subject. Importantly, genetic modification issues are not more emotive to individuals non-accepting of evolution for belief based reasons[28]. More particularly, the material taught and examined under genetics is largely non-emotive. Moreover, any notion that opposition to GM crops explains why those not accepting of evolution show similar increments in genetics and evolution understanding, fails to explain why low acceptance students performed less well than accepting students prior to teaching in both evolution and genetics knowledge tests and why they were classified by their teachers into foundation ability sciences classes, where this reflected their performance in all core sciences.

Limits to generalizability

That our study was UK based, as we suggest, limits the generalizability of our study. Further, the use of genetics as a comparator limits our ability to generalize the results too far. Nonetheless, that low general science ability (classified by teachers) predicts poor response to teaching (Mann Whitney U test, normalized genetics response by ability, P=0.0015; normalised evolution response by ability P=0.013), suggests that the foundation ability students are not responding well to science teaching as currently practiced. It remains to be seen whether preteaching acceptance of evolution predicts a response to teaching of science subjects that are not biological and to subjects that are not scientific at all.

Implications

What implications does our study have? We have found poor response to teaching of non-accepting students is better explained by aptitude than by psychological conflicts. Is there much that can be done for those of lower aptitude? We previously showed that teaching genetics before evolution is an efficient mechanism to improve evolution understanding at no cost to genetics understanding, and that the genetics-first approach was the only ordering that enables an increase in evolution understanding in foundation classes. Optimization of teaching strategy for different aptitudes (as done in mathematics education) is worthy of research. Identification of learning styles ((auditory, visual kinaesthetic *etc*.) of those of low aptitude may well also help. Current evidence suggests that visual (graph based) presentation of information[5], rather than textual presentation may help many, especially visual learners.

The results here also suggest that focusing on acceptance *per se* is not helpful, as this may be more a consequence of the nexus of low scientific aptitude, rather than the cause of poor learning outcome. This thus reinforces the notion that teachers should teach the science and not focus on belief systems[17]. This comes with two caveats. First, it might be that for religious students, conflict may be avoided by encouraging a compatibilist position, but this remains to be tested. Second, a robust understanding of the difference between evidence-based and belief-centred assertions of understanding may be crucial for helping students understand the difference between science and non-science. In this context emphasis in the classroom on evidence-based acceptance of evolution, rather than “belief” in evolution may be a subtle but important route[18].

METHODS

Methods for this paper are identical to those we recently reported for our study of teaching order[19]. Here therefore we provide an overview of these methods and advise the reader to consult the prior paper and its supplements for fuller detail.

Ethical considerations and data protection

Ethical guidelines as prescribed by The British Educational Research Education[37] have been followed. Particular consideration has been taken when working with school students, and approaches that place any undue burden on participants have been avoided. Research through questionnaires and focus groups has taken place within students' schools and have involved students' usual science teachers so as to minimise undue intrusion. For consent forms see [19].

Student questionnaire

Quantitative data were collected through a student questionnaire to determine acceptance of evolution and understanding of genetics and evolution. This was devised for GCSE-level students (14–16-year-olds) who study evolution and genetics as part of their science GCSE science course. An advantage of analysis of this age group is that order effects may well be most easily detected if there has been little or no priming. While primary school children in the UK are presently expected to be taught basic genetics and evolution on the national curriculum, this is a recent introduction and the cohort we analysed did not have this exposure. Indeed, this academic stage was chosen as it is currently the first, and perhaps only, period at which students have to learn about evolution. This cohort is not self-selecting in the way that a higher academic stage might be. For example, students aged from 16−18 and studying for a Biology A-Level qualification will already have achieved a reasonable standard of academic achievement in science to enrol in this, and presumably have an interest for biology, or would not have chosen to study the subject further. Therefore, in choosing to study GCSE-level students, this research has involved a wide variety of students, in terms of academic ability and interest in evolution and science.

For evolution acceptance, evolution understanding and genetics understanding the tests were performed pretest – prior to learning both genetics and evolution and post-test –

immediately after learning of both topics. We consider here only data where a given student answered all pre and all post tests.

The questionnaire consists of 25 questions: 13 focus on acceptance of evolution (Section A), 6 on genetics knowledge (Section B), and 6 on evolution knowledge (Section C). None of the questions involve extended writing and are all variations of the multiple-choice question. These types of questions were chosen for their practicalities: to aid student completion time, to avoid instances of not being able to understand transcriptions, to allow for quantitative analysis of data, and that this method is commonly used in similar studies (e.g., [21]).

At all stages of the questionnaire development, including a pilot study, evolution and education experts were consulted from the University of Bath along with practising teachers. The questionnaire was designed with time constraints in mind: teachers consulted during its development were insistent that the questionnaire must be short enough so that its completion would not considerably reduce their lesson time. Ten to 15 minutes was considered an appropriate length. The final questionnaires are presented in our prior paper[19].

Evolution acceptance. Section A assesses students' opinions towards evolution and consists of 13 Likert Scale items. These were based largely on the Measure of Acceptance of the Theory of Evolution (MATE), which was developed to assess biology teachers' acceptance of evolution[38] and later, undergraduate students' acceptance of evolution [36]. The original MATE instrument consists of 20 items spread disproportionately across 6 subsections of evolutionary concepts or aspects. It was decided that this was too long for school students. Appropriate questions were chosen based on their relevance to these different aspects of evolution and their accessibility to school-aged students. Given that the MATE has been developed and tested predominately on teachers and undergraduate students (e.g., [21,39]), some modifications to the language used were needed. Where necessary, statements were reworded to make them more understandable. Two items were also based on Lovely and Konderick's study[40] into undergraduate opinions of evolution. This section was found to be reliable through internal consistency checks (alpha 0.82, G6 0.83).

Acceptance categorisation. Scores for individual items are measured on a scale of 1 to 5, corresponding to "very high acceptance," "high acceptance," "undecided," "low acceptance," or "very low acceptance" of evolution. Students receive a total score of between 13 and 65 (a higher score represents a higher acceptance of evolution). We treat each score as a quantitative variable, rather than a discrete one.

Genetics knowledge. Section B consists of 6 questions which focus on knowledge of genetics. This includes variations on questions from recent GCSE exams, questionnaires used in the Genetics Literacy Assessment Instrument (GLAI) for undergraduates [41], and questions from [42] in their study of school students' understanding of genetics. Two of these questions involve choosing or ordering key words from lists provided, and one question involves ticking boxes. These types of questions were chosen to gain greater insight into students' ideas on living organisms and genetics and to add variety to the questionnaire for students. This section was found to be reliable through internal consistency checks (alpha 0.77, G6 0.82).

Evolution knowledge. Section C focuses on evolution knowledge and consists of 6 questions. This section includes a variety of different aspects of evolution, including natural selection and geological time. Most of these were variations of questions used by Rutledge and Warden[21] in their research into acceptance and understanding of evolution among high school biology teachers. Additionally, a number of questions were devised with the assistance of evolution experts. Each question was scored equally with a section total out of 6. This section was found to be less reliable through internal consistency checks (alpha 0.25, G6 0.22) but this probably reflects the low number of questions and the fact that each question was testing a different issue (hence high cross correlation is not desirable).

Testing regime. Students were given the same questionnaire before and after teaching (for which there is precedent, see e.g. [30]). While this has the notable disadvantage that the students may be primed, thus obviating any analysis of absolute gains in understanding, by controlling the questions we remove a potential noise variable. Were one to introduce new questions, even logically similar ones, we cannot be certain that the change in score reflects a change in understanding, as we then need to add assumptions about the understandability and comparability of different questions. If there were variability between pupils in the understandability of any new questions, we would have introduced an unnecessary noise variable. While our approach might affect interpretation of absolute change in scores, we are, however, interested in increase in response compared to other increases in response or as they correlate with another factor. We are not first and foremost interested in the absolute change *per se*. Put differently, even if all scores go up – possibly because the students better understand the same questions – the issue is why some students' scores go up more than others.

The median gap between pre and post assessment was 63 days. We are confident that teachers did not "teach to the test" as the anonymity of students and schools in the study was explained to teachers prior to their agreeing to partake. Moreover, the teachers were instructed to teach their normal GCSE syllabus. We also control for within class-effects which would remove any better-teacher effects, should such confound exist.

Focus groups

Focus groups were designed to better understand the responses found in the student questionnaires, i.e., why students were or were not accepting of evolution; how these views related to knowledge of evolution; how these related to knowledge of genetics; and what other factors are important. Seventy-six students were involved in 16 focus groups. These students were from 10 different schools. The largest focus groups contained 7 students and the smallest, 2. All students were from groups identified as "higher-ability," with most students being from among the top sets in each school. The majority of students were in Years 9, 10, and 11 and studying towards their GCSE examinations. Six students were in Year 12 and studying for A-Level exams. Most focus groups comprised students of the same age and from the same class, however there were 3 groups that contained a mixture of ages and classes.

Teacher parameters

To estimate teacher acceptance we conducted teacher surveys via an online MATE resource that we developed. The survey is "highly reliable": evolution acceptance has a Cronbach's alpha of 0.94 and G6 of 0.96 (maximum value is 1).

Background information

A mixture of state, faith, and independent schools have been involved in this project. All schools are from the South of England and Mid and South Wales. All schools within the accessible area were invited but not all accepted. All are English language schools. Schools included students from socially and economically diverse communities, including rural, suburban, and inner city. A number of schools are single-sex. Although data were not collected specifically on student demographics, a wide range of ethnic backgrounds and faiths were represented. Background data on schools have been collected from inspection (OFSTED/ESTYN) reports, school websites, and from meetings with teachers. For further detail see [19]. We do not release information on demographics on a school-by-school basis as this might impinge on anonymity of schools, teachers and pupils, anonymity that was guaranteed.

Statistics. All statistics were conducted in R with data processing via Tcl scripts. Loess was performed using R using the loess function. We note that the loess method has the advantage over binning methods of not enforcing arbitrary bin sizes that can in turn distort proportionality between bins. Where rho is specified it may be assumed that the method was Spearman's rank correlation.

To test exclude between-class effects we consider each class in isolation and consider for each class the correlation of interest (e.g. between the normalised improvement in evolution understanding and the preteaching acceptance scores). We then take the values of this correlation (via Spearman's rank correlation) for all classes and test this set of intraclass rho values against a median correlation of zero using Wilcox signed rank test. As the strength of this test is dependent on both the number of classes being considered and the number of students in any given class, we consider the test for a variety of minimum class sizes (from a minimum of 5 students in a class to a minimum of 15).

Hartigan and Hartigan's dip test[29] was implemented in the R package diptest. The dip test metric is sensitive to sample size such that two otherwise identical distributions can report different dip scores depending on the sample size. To test for a difference between higher and foundation classes, we thus control sample size by randomly subsampling 172 from the 1055 higher ability students (without replacement) and calculated the dip score for this subsample, being identical in size to the foundation class.

Repeating this 10000 times we derive the distribution of dip scores of the population of higher ability students that is directly comparable to the dip score for the foundation class. We calculate 95% confidence intervals using the quantile function in R. We then compare the dip score of the foundation class against those confidence intervals and present the median of the dip scores of the subsamples.

To calculate the significance of the difference in the frequency of low acceptors (score ≤ 32) a nonparametric randomization was employed. We randomly reassigned the data to two paritons (1055 and 172 in size) without replacement, these being the sizes of the higher and foundation samples respectively. For each randomised pair of partitions, we calculate the frequency of low acceptors in both and consider the modulus of the difference between these two as the reporting statistic. We ask of 10,000 simulations how many have an absolute difference between frequencies that is greater than or as great as that seen in the real data. If this number is n, with m simulations, $P=n+1/m+1$.

To determine whether a correlation between x and y is significantly stronger or weaker than the correlation between z and y, we performed a nonparametric Monte Carlo simulation. We calculated the 2 Spearman correlations and asked about the difference in the Spearman rank coefficient. We then randomised the vector y, and considered for each randomised version the correlation between x and randomised y and z and randomized y. We again considered the modular difference in Spearman rho value for the correlation of these 2 individually against variable y (the mean difference in the simulants is zero). Repeating the simulation 10,000 times, we asked how often the modular difference was as great or greater than that observed in the real data. As we employed modular data, the test is 2-tailed. The type 1 error rate is then given by $P = (n + 1)/(m + 1)$, where n is the number of randomizations in which the diference is as extreme or more extreme than that observed in the real data and m the number of simulations. Randomization was done in all cases uses the sample function in R. Other tests are explained in text.

Significance is taken at alpha <0.05.

Item nonresponse levels were low. We considered alternative means to handle nonresponse, but as the numbers are so low, they make no difference to results (see prior analysis[19] for further details including raw data files).

Acknowledgements

We thank Dr Jonathan Milner and the Evolution Education Trust for funding.

Author contributions

LDH, MH and RM devised the program of work. MH and LDH supervised the project. RM collected the data. LDH analysed the data. RM, MH and LDH wrote or edited the paper.

Data availability

All data are freely available from the supplementary information of our prior paper: https://doi.org/10.1371/journal.pbio.2002255

Competing Financial Interests

The authors declare no competing financial interests.

REFERENCES

1. Sherkat, D.E. Religion and Scientific Literacy in the United States. *Social Science Quarterly* **92**, 1134-1150 (2011).
2. McKeachie, W.J., Lin, Y.-G. & Strayer, J. Creationist vs. Evolutionary Beliefs: Effects on Learning Biology. *The American Biology Teacher* **64**, 189-192 (2002).
3. Festinger, L. *A Theory of cognitive dissonance.* , (Stanford University Press., Stanford, CA, 1957).
4. Kahan, D.M. *et al.* The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change* **2**, 732-735 (2012).
5. Nyhan, B. & Reifler, J. The roles of information deficits and identity threat in the prevalence of misperceptions. (2017).
6. Steele, C. The psychology of self-affirmation: Sustaining the integrity of the self. in *Advances in experimental social psychology*, Vol. 21 (ed. Berkowitz, L.) 261-302 (Academic Press, New York, 1988).
7. Nickerson, R.S. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* **2**, 175-220 (1998).
8. Dodick, J. & Orion, N. Introducing Evolution to Non-Biology Majors via the Fossil Record: A Case Study from the Israeli High School System. *The American Biology Teacher* **65**, 185-190 (2003).

9. Allum, N., Sturgis, P., Tabourazi, D. & Brunton-Smith, I. Science knowledge and attitudes across cultures: a meta-analysis. *Public Understanding of Science* **17**, 35-54 (2008).
10. Bauer, M.W. The Evolution of Public Understanding of Science-Discourse and Comparative Evidence. *Science Technolgy and Society* **14**, 221-240 (2009).
11. Fischhoff, B. Risk perception and communication unplugged - 20 years of process. *Risk Analysis* **15**, 137-145 (1995).
12. Smith, P.J. *et al.* Parental Delay or Refusal of Vaccine Doses, Childhood Vaccination Coverage at 24 Months of Age, and the Health Belief Model. *Public Health Reports* **126**, 135-146 (2011).
13. Eveland, W.P. & Cooper, K.E. An integrated model of communication influence on beliefs. *Proceedings of the National Academy of Sciences* **110**, 14088-14095 (2013).
14. Joslyn, M.R. & Haider-Markel, D.P. Who Knows Best? Education, Partisanship, and Contested Facts. *Politics & Policy* **42**, 919-947 (2014).
15. Burke, K.L. 8 Myths About Public Understanding of Science. Vol. 2017 (American Scientist, 8 Myths About Public Understanding of Science, 2015).
16. Raine, L., Funk, C. & Page, D. An elaboration of AAAS Scientists' Views. Vol. 2017 (Pew Research, http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/07/Report-AAAS-Members-Elaboration_FINAL.pdf, 2015).
17. Lawson, A.E. & Worsnop, W.A. Learning about Evolution and Rejecting a Belief in Special Creation: Effects of Reflective Reasoning Skill, Prior Knowledge, Prior Belief and Religious Commitment. *Journal of Research in Science Teaching* **29**, 143-166 (1992).
18. Williams, J.D. Evolution Versus Creationism: A matter of acceptance versus belief. *Journal of Biological Education* **49**, 322-333 (2015).
19. Mead, R., Hejmadi, M. & Hurst, L.D. Teaching genetics prior to teaching evolution improves evolution understanding but not acceptance. *PLoS Biol* **15**, e2002255 (2017).
20. Donnelly, L.A., Kazempour, M. & Amirshokoohi, A. High School Students' Perceptions of Evolution Instruction: Acceptance and Evolution Learning Experiences. *Research in Science Education* **39**, 643-660 (2008).
21. Rutledge, M.L. & Warden, M.A. Evolutionary Theory, the Nature of Science & High School Biology Teachers: Critical Relationships. *The American Biology Teacher* **62**, 23-31 (2000).
22. Ehrle, E.B. Notes on the Teaching of Evolution. *The American Biology Teacher* **22**, 418-419 (1960).
23. Banet, E. & Ayuso, G.E. Teaching of biological inheritance and evolution of living beings in secondary school. *International Journal of Science Education* **25**, 373-407 (2003).
24. Lawson, A.E. & Thompson, L.D. Formal Reasoning Ability and Misconceptions Concerning Genetics and Natural Selection. *Journal of Research in Science Teaching* **25**, 733-746 (1988).
25. Lawson, A.E. & Weser, J. The Rejection of Nonscientific Beliefs about Life: Effects of Instruction and Reasoning Skills. *Journal of Research in Science Teaching* **27**, 589-606 (1990).

26. Deniz, H., Donnelly, L.A. & Yilmaz, I. Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers: Toward a more informative conceptual ecology for biological evolution. *Journal of Research in Science Teaching* **45**, 420-443 (2008).
27. Woods, C.S. & Scharmann, L.C. High School Students' Perceptions of Evolutionary Theory. *Electronic Journal of Sciene Education* **6**(2001).
28. Drummond, C. & Fischhoff, B. Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proc Natl Acad Sci U S A* **114**, 9587-9592 (2017).
29. Hartigan, J.A. & Hartigan, P.M. The dip test of unimodality. *Annals of Statistics* **13**, 70-84 (1985).
30. Nadelson, L.S. & Sinatra, G.M. Shifting Acceptance of Evolution: Promising Evidence of the Influence of the Understanding Evolution Website. *The Researcher* **23**, 13-29 (2010).
31. Berkman, M.B. & Plutzer, E. Defeating Creationism in the Courtroom, But Not in the Classroom. *Science* **331**, 404-405 (2011).
32. Nyhan, B., Reifler, J., Richey, S. & Freed, G.L. Effective Messages in Vaccine Promotion: A Randomized Trial. *Pediatrics* **133**, E835-E842 (2014).
33. Lewandowsky, S., Gignac, G.E. & Oberauer, K. The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. *Plos One* **8**(2013).
34. Lewandowsky, S., Gignac, G.E. & Oberauer, K. The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science (vol 8, e75637, 2013). *Plos One* **10**(2015).
35. Dixon, R.M. & Jones, J.A. Conspiracist Ideation as a Predictor of Climate-Science Rejection. *Psychological Science* **26**, 664-666 (2015).
36. Rutledge, M.L. & Sadler, K.C. Reliability of the Measure of Acceptance of the Theory of Evolution (MATE) Instrument with University Students. *The American Biology Teacher* **69**, 332-335 (2007).
37. BERA. Ethical Guidelines for Education Research. (2011).
38. Rutledge, M.L. & Warden, M.A. The Development and Validation of the Measure of Acceptance of the Theory of Evolution Instrument. *School Science and Mathematics* **99**, 13-18 (1999).
39. Rutledge, M.L. & Sadler, K.C. University Students' Acceptance of Biological Theories - Is Evolution Really Different? *Journal of College Science Teaching* **41**, 38-43 (2011).
40. Lovely, E.C. & Kondrick, L.C. Teaching Evolution: Challenging Religious Preconceptions. *Integrative and Comparative Biology* **48**, 164-74 (2008).
41. Bowling, B.V. *et al.* Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics* **178**, 15-22 (2008).
42. Lewis, J. & Wood-Robinson, C. Genes, chromosomes, cell division and inheritance - do students see any relationship? *International Journal of Science Education* **22**, 177-195 (2000).

Figure legends

Fig. 1 The relationship between the acceptance of evolution and the understanding of evolution prior to teaching. Here we present that scores for the preteaching acceptance of each pupil and the preteaching evolution understanding scores. Spearman's rho = 0.22, P=9 x 10^{-15}. The regression line is the best fit line of *y* predicted by *x*. However, as assumptions of linear regression are not fully met it is provided for illustrative purposes alone to indicate the trend. N=1227.

Fig. 2 The stratification of evolution acceptance scores prior to teaching and teacher-derived classification of student ability (foundation or higher ability). Here we present that scores for the preteaching acceptance of each pupil stratified by teacher-derived classification of student ability visualised as violin plot. Higher ability N= 1055, foundation, N=172. Median higher = 51, 95% CI 37.35 - 61.65; median foundation = 47, 95% CI 35 – 60. Mann Whitney U test, P=9 x 10^{-11}.

Fig 3. Relationship between change in understanding of evolution score and preteaching evolution understanding score. Here we plot for each student the change in understanding of evolution score (post teaching score – preteaching score) against the preteaching understanding score. The blue line is the loess regression line around which residuals are generated. Loess was run under default settings. Equivalent number of parameters=4.88, residual standard error = 1.289. N=1227.

Fig 4. Normalised gain in evolution understanding owing to teaching is positively correlated with preteaching acceptance in evolution score. Here we plot the residuals of the loess regression (shown in Fig 3) as normalized gain in evolution understanding, as a function of preteaching acceptance of evolution. Spearman's rho=0.17, P=4.6 x 10^{-9}. The regression line is the best fit line of *y* predicted by *x*. However, as assumptions of linear regression are not fully met it is provided for illustrative purposes alone to indicate the trend. N=1227.

Fig 5. Normalised gain in genetics understanding owing to teaching is positively correlated with preteaching acceptance in evolution score. Here we plot the residuals of the loess regression of change in genetics understanding predicted by preteaching genetics understanding (normalized gain in genetics understanding), as a function of preteaching acceptance of evolution. Spearman's rho=0.15, P=6.4 x 10^{-8}. The regression line is the best fit line of *y* predicted by *x*. However, as assumptions of linear regression are not fully met it is provided for illustrative purposes alone to indicate the trend. N=1227.

Table 1. The correlation between preteaching acceptance scores and normalised increase in evolution understanding in stratified analysis. For each stratification we consider the Spearman rank correlation between preteaching acceptance score and the residuals from the loess of change in evolution understanding predicted by preteaching evolution understanding. Change is defined as post score – pre-teaching score. Rho is the Spearman rank correlation coefiicent, P the significance and N the sample size.

Stratification	Level	Rho	*P*	*N*
Teaching order	Genetics first	0.18	3.5×10^{-7}	776
	Evolution first	0.10	0.032	451
Ability	High	0.13	1.2×10^{-5}	1055
	foundation	0.23	0.0029	172