PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Sauce B, Bendrath S, Herzfeld M, Siegel D, Style C, Rab S, Korabelnikov J, Matzel LD. 2018 The impact of environmental interventions among mouse siblings on the heritability and malleability of general cognitive ability. *Phil. Trans. R. Soc. B* **373**: 20170289.
http://dx.doi.org/10.1098/rstb.2017.0289

Accepted: 5 June 2018

One contribution of 15 to a theme issue 'Causes and consequences of individual differences in cognitive abilities'.

Subject Areas:
behaviour, genetics, cognition

Keywords:
individual differences, heritability, malleability, intelligence, environment intervention, mouse

Author for correspondence:
Louis D. Matzel
e-mail: matzel@psych.rutgers.edu

The impact of environmental interventions among mouse siblings on the heritability and malleability of general cognitive ability

Bruno Sauce[1], Sophie Bendrath[2], Margalit Herzfeld[2], Dan Siegel[2], Conner Style[2], Sayeeda Rab[2], Jonathan Korabelnikov[2] and Louis D. Matzel[2]

[1]Department of Neuroscience, Karolinska Institutet, Solnavägen 9, Solna 171 65, Sweden
[2]Department of Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA

BS, 0000-0002-9544-0150; LDM, 0000-0003-0462-7188

General cognitive ability can be highly heritable in some species, but at the same time, is very malleable. This apparent paradox could potentially be explained by gene–environment interactions and correlations that remain hidden due to experimental limitations on human research and blind spots in animal research. Here, we shed light on this issue by combining the design of a sibling study with an environmental intervention administered to laboratory mice. The analysis included 58 litters of four full-sibling genetically heterogeneous CD-1 male mice, for a total of 232 mice. We separated the mice into two subsets of siblings: a control group (maintained in standard laboratory conditions) and an environmental-enrichment group (which had access to continuous physical exercise and daily exposure to novel environments). We found that general cognitive ability in mice has substantial heritability (24% for all mice) and is also malleable. The mice that experienced the enriched environment had a mean intelligence score that was 0.44 standard deviations higher than their siblings in the control group (equivalent to gains of 6.6 IQ points in humans). We also found that the estimate of heritability changed between groups (55% in the control group compared with non-significant 15% in the enrichment group), analogous to findings in humans across socio-economic status. Unexpectedly, no evidence of gene–environment interaction was detected, and so the change in heritability might be best explained by higher environmental variance in the enrichment group. Our findings, as well as the 'sibling intervention procedure' for mice, may be valuable to future research on the heritability, mechanisms and evolution of cognition.

This article is part of the theme issue 'Causes and consequences of individual differences in cognitive abilities'.

1. Introduction

Cognitive abilities can be separated into multiple factors (derived from domain-specific cognitive processes) that, at least in some species, are influenced by a common, general factor (derived from domain-general cognitive processes) [1]. This general cognitive ability (GCA), sometimes interpreted as 'intelligence', is defined as the general capacity to learn, reason, plan and solve problems [2]. GCA can vary greatly across individuals, and studying these variations provide precious information on the genetic and environmental factors that shape this trait [3]. In addition, heritable individual differences are the necessary fuel for evolution via natural selection, and so studying the heritability of intelligence can shed light on how cognition evolved [1].

Heritability is a statistic that captures how much of the variation in a trait is due to genetic differences, and the metric ranges from 0.0 to 1.0. Because the

THE ROYAL SOCIETY PUBLISHING

genetic and environmental effects often covary (e.g. related individuals often share some environments in addition to some genes), studies of heritability exploit special cases where the two effects can be separated, such as cases of families with twins [4]. These studies usually find quite high heritabilities for GCA: around 0.60 (or sometimes higher) in adult humans [5], 0.50 in chimpanzees [6] and 0.40 in mice [7].

The influence of genes on GCA, however, might not be as powerful as estimates of heritability might lead many to believe. There is evidence in humans suggesting that despite its high heritability, intelligence is also quite malleable, with a great deal of intelligence's variation being attributable to environmental factors across families and socioeconomic status [8]. Adoption studies, for example, often reveal an average increase in intelligence of one standard deviation (15 IQ points) within several years of adoption [9]. This represents a remarkable cognitive gain of the adopted children over their biological, non-adopted siblings.

An interesting question arises from the above observations: how can intelligence be at the same time highly heritable and highly malleable? One possible solution to this odd paradox is the role of gene–environment interactions between genetic and environmental factors [8]. In one type of gene–environment interaction, genetically different individuals will have a different subjective experience (i.e. pay attention to, absorb or respond differently) to the same objective experience, and this can lead to further increases or reductions in intelligence. Gene–environment interactions in intelligence can be especially elusive to detection, and are often disregarded or ignored by researchers [10].

The difficulties associated with tests of gene–environment interactions are due in part to limitations on work with humans, including being largely confined to the assessment of heritability and malleability using only correlational methods (i.e. methods without experimental manipulations that focus on variation between individuals). With laboratory animals, we can easily control the environment, and can combine correlational and experimental designs to more precisely understand the effects of genes, environment and their interaction. However, there are relatively few studies on individual difference in cognitive abilities in non-human animals [11]. Most animal studies have primarily used experimental approaches (i.e. studies with manipulation that look at group-level effects and ignore inter-individual variation), and focus on single cognitive domains [12]. While these studies have proven fruitful in delineating certain neurobiological substrates of task performance [13,14], they do not capture how genetic and environmental factors contribute to create the differences in general cognitive skills.

In previous research by our group, genetically diverse mice were tested on batteries of learning tasks, each of which with unique sensory, motor and motivational demands [15,16]. Mice that do well in one task of the battery tend to perform well in other tasks within the battery too, revealing a positive correlation of each animal's learning across all tasks. The 'general learning' scores derived from a factor analysis also covaries with other cognitive domains, such as inductive and deductive reasoning [17], spatial ability [18] and working memory and attention [19]. That means that the common factor behind performance in the learning batteries is capturing something cognitively more general (across domains) than simply learning. In fact, others have described our results as qualitatively analogous to what is described in humans as intelligence [20].

Previous studies by our group also found that differences in mouse intelligence correlate with difference in expression of genes known to play a role in learning and synaptic plasticity [21], and also correlate with dopamine-induced activity in the prefrontal cortex [22,23]. Because these studies were performed in laboratory mice living in a fairly homogeneous environment, the results suggest that individual differences in a mouse's IQ have strong genetic influences. Similar to the human literature, there is also evidence for the malleability of mouse intelligence. For example, a study found that a combination between exercise and novel environments in mice increases neurogenesis in the hippocampus and retention of these new neurons [24]. A review by van Praag *et al*. [25] concluded that environmental enrichment in rodents (defined as 'a combination of complex inanimate and social stimulation') can have lasting effects on learning and brain growth.

Here, we attempted to test the prediction (based on the evidence above) that mouse intelligence can have both high heritability and malleability. For this, we used groups of full-sibling mice and exposed subsets of each sibling cohort to different environments. In other words, our study combined the design of a sibling study with a controlled environmental intervention. This allowed us to estimate how many of the differences in mouse intelligence are influenced by genetic and the environmental factors, as well as test for expected gene–environment interactions.

2. Material and methods

(a) Subjects

We used 232 CD-1 outbred male mice from Harlan Laboratories (Indianapolis, IN, USA). Estimates of genetic variation in this line have indicated that despite over 50 years of breeding, they are very similar to wild mouse populations [26]. The mice arrived in our laboratory between at four and five weeks of age, and they were singly housed in clear shoe box cages inside a temperature-controlled colony room under a 12 L : 12 D cycle. In order to minimize any differential stress responses due to experimenter handling, we handled the animals for 90 s a day for a period of 7 days prior to the start of the experiment. Handling consisted of removing the mice from their home cage and holding them while walking throughout the laboratory space.

The 232 mice comprised 58 sets of four siblings (fraternal quadruplets), totalling 58 families whose parents were unrelated to each other (as guaranteed by the supplier Envigo). Two siblings of a set, randomly chosen, stayed in the home environment (control group) and the two other siblings received an environmental 'enrichment' treatment consisting of physical exercise and exposure to novel and engaging environments (enrichment group).

All mice had continuous access to both food and water. The only exception was during the tests requiring food deprivation, when mice were provided with food in their home cages for only 90 min a day, beginning on the day prior to testing. Although mild, this level of deprivation is sufficient to maintain stable performance on learning tasks [15]. All experiments were conducted in accordance with protocols approved by the Rutgers University Institutional Animal Care and Use Committee (IACUC).

(b) Environmental enrichment

The enrichment manipulation lasted for 16 days, and the two groups of mice were maintained in separate, though nominally identical, colony rooms. (During enrichment, it was necessary

to separate the groups into two colony rooms owing to the auditory stimulation associated with aspects of the enrichment procedure.) The enrichment group lived in home cages containing a running wheel for exercise throughout the treatment. These mice were exposed to one novel environment each day for 30 min outside their home cage. Animals were individually exposed to a new novel environment at approximately the same time each day for each of 16 days.

The environments encountered by the enrichment group were: (i) a large, black, plastic box with two concave towers on each side and a platform in the centre reachable by jumping; (ii) a narrow Plexiglas tube where the ends have two small boxes where the mice could traverse between the boxes by going through the tube; (iii) an eight-arm radial arm maze with all doors left open; (iv) an acoustic chamber with foam on the wall with a fan inside as the only sound mice would experience; (v) a black box with a white stripe on the walls and the floor covered with soft, plastic spikes; (vi) a white box with six different plastic toys; (vii) a large social box with a second mouse inside a cylindrical cage to interact with; (viii) an open rat shoe-box cage, with one-quarter of its depth filled with bedding and with 15 marbles on the top of the bedding that mice are prone to manipulate and then hide; (ix) a closed rat shoe-box cage with standard level of bedding containing four pieces of paper towel to be shredded; (x) a white box with a fixed 'merry-go-round' like structure inside; (xi) a metal pot with holes on the sides for nose poking, and a cover closing the pot; (xii) a large, white, plastic box with the two angled cylindrical beams originating on the floor that the mice could climb; (xiii) a closed mouse shoe-box cage put upside down with 10 strings of rope crossing the top of it creating a net where mice could walk; (xiv) a white box containing a white PVC tube with a mirror at one of its ends; (xv) an acoustic chamber with foam on the wall with a metal plate inside containing jars filled with small metal bells to produce sound whenever the mice roll the jars; and (xvi) a large white plastic box with an angled ramp which ended at a large metal grid that the animals could climb onto.

Upon completion of the 16 days, the enrichment group was moved back to standard cages in the colony room with the control siblings. All mice were then handled again for 90 s a day for 7 days. This ensured that both groups were receiving similar contact with humans and with the surrounding laboratory before the start of behavioural testing. Also, these 7 days of break would function as 'rest' for mice in the enrichment group to minimize any differences in metabolic levels between them and mice in the control group (metabolic differences might arise as a consequence of the environmental experience, and confound results by affecting the state of attention and of blood glucose levels during the learning tasks). After this, all mice were tested in a battery of learning tasks to provide an index of their GCA.

(c) Learning battery to measure general cognitive ability

All mice were tested on a battery of five learning tasks, presented in the following order: Lashley maze, passive avoidance, T-maze alternation, odour discrimination and spatial water maze. Our group had used these tests in the past to estimate GCA, and described them in detail elsewhere [15,27]. In each task, we obtained an outcome variable of learning performance for later analyses. These variables were meant to capture the differences in rate of learning among mice, and so only consider a mouse's early performance. (As opposed to, for example, considering performance in all trials, because mice are typically at high levels of performance during later trials. In this study, we are interested in learning rates in tasks, not maximum task performance. In most instances, mice reach comparable levels of asymptotic performance.) A brief description of each task is provided below.

In the Lashley maze, mice must navigate four interconnected alleys to reach a goal box that contains a food reward. This task is designed to measure the learning of a stable route, involves egocentric navigation, requires ambulation and has food as motivator. During each of five total trials, we tracked the two types of errors that could be committed: backtracking, which we define as a mouse going from one alley opening to the prior alley opening, and dead end, which we define as a mouse walking past an alley opening towards a dead end. Between each trial, the mice were placed back in their home cage for 20 min. We defined the outcome variable in the Lashley maze as the mean errors (backtracking and dead end combined) across the first three trials after acclimation.

In the passive avoidance task, a mouse was confined to a 'safe' platform for 5 min, after which the exit door was opened. When a mouse stepped from the safe platform onto a grid floor (i.e. baseline latency), it would encounter a 5 s compound aversive stimulus composed of a bright white light and noise (a loud oscillating tone, or 'siren'). During the aversive stimulus presentation, the mice retreat onto the safe platform, where they were then confined for a 5 min interval. At the end of this interval, the door from the platform was again opened, so that the mouse was again free to exit the platform (i.e. avoidance latency). This task is designed to measure the learning of operant avoidance, involves fear, requires passivity and has aversive light and sound as motivator. We defined the outcome variable in the passive avoidance as the ratio of avoidance latency divided by baseline latency. Mice with better learning should have relatively longer latencies to step from the platform during the avoidance period.

In the T-maze alternation task, mice must alternate their foraging for a food reward between two arms. This task is designed to measure the learning of choice alternation, involves attentional capacity to ignore place preference (and the tendency to return to the last location of reinforcement), requires ambulation and has food as motivator. The apparatus was a start arm that intersected at its extremity with two choice arms, forming a 'T' shape. To help the mice distinguish between arms, one of the arms' walls had vertical white stripes, and the other had horizontal white stripes. If an incorrect choice was made, the animal could correct its mistake and find the food in the other arm. After the correct choice was made, we placed the animal back in the start area where it waited 20 s for the following trial. We administered 2 days of testing with 12 trials per day. We defined the outcome variable in the T-maze alternation as the trial when a mouse first made four correct choices in a row. This variable is meant to capture early learning performance by looking at the beginning of minimum competency in the task. (It is notable that mice initially tend to return to a location previously reinforced, and so a streak of four correct alternations is unlikely to be only due to chance.)

In the odour discrimination task, mice had to use a specific odour cue (mint) to find food. The task was administered in a square box, where three of the box's four corners always contained cups, and the fourth corner served as a start location. Immediately before each trial, fresh swabs were loaded with lemon, almond or mint (the target) odorants. This task is designed to measure the learning of odour discrimination, involves olfactory stimuli, requires ambulation and has food as motivator. Each mouse received a total of four trials. After each trial, we rearranged the location of the food cups, and waited 6 min before another trial. An error was recorded any time a mouse sampled an incorrect cup, or when it sampled the target cup without retrieving the available food. We defined the outcome variable in the odour discrimination as the mean errors across the first three trials after acclimation.

In the spatial water maze task (or Morris water maze), mice are placed in a circular pool of opaque water and can find an underwater, hidden escape platform using spatial cues for navigation. Once the mice had all four of their paws on the platform, it was allowed to stay on the platform for 5 s, and then was removed from the pool for a 20 min inter-trial interval. If a mouse could not find the platform after 90 s, we placed it on the platform for 5 s. Each mouse received a total of six trials, on which the starting location was changed on each trial (thereby mitigating strategies based on egocentric navigation). This task is designed to measure the learning of triangulation, involves spatial navigation, requires swimming and has water immersion as motivator. We recorded path lengths from the start position to the platform during each trial as the measure of learning. We defined the outcome variable in the spatial water maze as the mean path lengths across the first two trials after acclimation.

(d) Statistical analyses

For all the tasks in the learning battery, we defined univariate outliers as any values above or below two interquartile ranges. We then applied the technique known as 'bring it to the fence' to modify the outliers to values at either the lower fence (first quartile minus twice the interquartile range) for low outliers or the upper fence (third quartile plus twice the interquartile range) for high outliers [28]. We also tested all data for the presence of kurtosis and skewness to check if the variables conformed to a normal distribution. These pre-analyses were all done in SPSS 24. We treated missing data by estimating values for each case with multiple imputation, a technique that estimates missing data points based on the observed data. Multiple imputation provides less biased information than simpler procedures for dealing with missing data such as listwise deletion, pairwise deletion or imputation of means [29].

Each mouse's value of learning performance was determined for each of the learning tasks. Using an exploratory factor analysis, a statistical method that is used to explore underlying factors capturing the common covariation among variables, we assessed individual differences in learning on all the tasks. An exploratory factor analysis captures only the variance shared in common between the variables, and therefore is ideal to reveal a common construct influencing learning in all tasks (as opposed to techniques such as principal component analyses that capture both shared and non-shared variance, and that are better suited for purposes of dimension reduction). From this analysis, each animal was then assigned a factor score, which represents their GCA, or intelligence score. In principal, our primary factor could have captured a common influence other than 'general cognitive ability', such as exploratory tendencies, anxiety or stress reactivity. While this is always a possibility, extensive prior analyses measuring these traits suggest that 'non-cognitive' influences load onto secondary factors independent of the primary factor [30,31].

We performed a parallel analysis in SPSS 24 to verify if the GCA factor we obtained has meaningful exploratory value, by contrasting its eigenvalue with a 'meaningless' eigenvalue based on random data (1000 datasets) that recreate the same parameters (five variables, $n = 231$) [32]. We also performed a confirmatory factor analysis to test our assumption that there is a single factor (GCA) explaining the common variance between learning tasks of the battery. We used the maximum-likelihood estimation in AMOS 24 to acquire the solution for the model. We assessed model fit by using two absolute indices—model χ^2 (χ^2_M) and root mean square error of approximation (RMSEA). For χ^2_M, the null hypothesis is the model itself, so failing to reject it indicates a good fit [33]. Similarly, RMSEA values of 0.06 and below are considered good [34]. In addition to these two absolute indices, we also assessed model fit with an incremental index, the comparative fit index (CFI), which indicates an adequate model fit at values of 0.95 or above [34]. We chose these tests due to their statistical relevance and frequent use [33].

We used the framework of linear mixed models in SPSS 24 for all further analyses. We estimated the heritabilities of GCA scores from both groups (enrichment and control) combined as well as from each group separately. We followed the classic full-sibling formulas by Falconer to obtain full-sib heritability (h_{FS}), its standard deviations (σh_{FS}) and significance [35]. We obtained the terms σ^2_F and σ^2_w (and consequent full-sibling intraclass correlation) from a mixed model with only a random effect of sibling families. All the variance explained by this random effect is σ^2_F, and thus represents genetic factors due to different parental origin (which also includes any shared early environment effect, such as the womb environment, as we discuss later). Meanwhile, all the residual variance is σ^2_w.

$$h_{FS} = \frac{2\sigma^2_F}{(\sigma^2_F + \sigma^2_w)} \quad \text{and}$$

$$\sigma h_{FS} = 2\left\{\frac{2[1 + (n-1)t]^2}{n(n-1)(N-1)}\right\}^{1/2},$$

where h_{FS} is the full-sibling heritability, σh_{FS} the standard deviation of the full-sibling heritability, σ^2_F the difference between the siblings of different families, σ^2_w the difference between siblings within a family, n the number of individuals per family, N the number of families, t the full-sibling intraclass correlation: $\frac{1}{2} h_{FS}$.

We also used the framework of linear mixed models in SPSS 24 to test for environmental effects in GCA scores. The model included the group treatments as a fixed effect (i.e. independent environmental effect), and sibling families as a random effect (i.e. independent genetic effect), and was estimated with maximum likelihood using an unstructured covariance structure. Lastly, we used the linear mixed models framework to test for gene–environment interactions in GCA. To accomplish that, we compared a baseline mixed model with sibling families as a random intercept (i.e. allowing for different family values of GCA between control and environment group treatments; in other words, a model with independent effects) against a mixed model with sibling families as a random intercept and group treatment as a random slope (i.e. allowing for different rates of change between control and enrichment treatments in GCA in each different genetic family; in other words, a model with gene–environment interactions). We compared these two models using a likelihood ratio (LR) χ^2 difference test [36]. An LR test compares nested models, and here it will test if the addition of the random slopes (gene–environment interaction model) to the random-intercepts model (independence model) results in a significantly improved fit.

3. Results

(a) Descriptive statistics

By examining all learning variables for the presence of univariate outliers, we found up to eight cases of outliers in each of the variables for Lashley maze, passive avoidance, T-maze and odour discrimination. We did not find any outliers for spatial water maze. We treated the outliers using the technique 'bring it to the fence' as described in Material and methods. We also tested all variables for skewness and kurtosis, and all variables had values of skewness and kurtosis well within the recommend range for normality (−2 and +2).

Less than 8% of the whole sample was missing (due to mice's natural death/illness and to experimenter's error during tests), and that data were missing at random, Little's

Table 1. Means, standard deviations, heritabilities (h_{FS}) and standard deviations of heritability (σh_{FS}) for all outcome variables of the learning battery, as well as the extracted variable of GCA (factor scores, where '0' is the anticipated median; values above 0 reflect performance better than the median) in all mice, and in each group separately.

variable	mean	s.d.	heritability	s.d. of heritability
Lashley maze (mean errors across trials 1–3)	11.73	4.78	0.27	0.15
passive avoidance (ratio of avoidance latency by baseline latency)	1.92	0.96	0.16	0.13
T-maze alternation (trial at first four correct choices in a row)	9.61	8.04	0.08	0.12
odour discrimination (mean errors across trials 1–3)	5.73	4.57	0.20	0.14
spatial water maze (mean path length in cm across trials 1–2)	1214.91	606.15	0.18	0.14
general cognitive ability—all mice (scores extracted from EFA of all learning tasks)	0	0.89	0.24	0.15
general cognitive ability—enrichment group	0.20	0.89	0.15	0.29
general cognitive ability—control group	−0.20	0.85	0.55	0.34

MCAR test: $\chi^2 = 20.51$, d.f. = 19, $p = 0.364$. As described in Material and methods, we estimated values for each missing case by using multiple imputation.

The means and standard deviations for all learning variables used in the subsequent analyses are shown in table 1.

(b) Factor analyses of general cognitive ability

The unrotated exploratory factory analysis of the performance data for the five learning tasks of the learning battery isolated a factor that accounted for a total of 19.5% of the variance in performance (table 2). That is equivalent to accounting for 28% of the total variance from a principal component analysis, a value similar to what we have found in the past. Performance from all of the learning tasks loaded consistently on this factor, and in the same direction. We used this first factor from the exploratory factor analysis, then, to extract factor scores to represent the mice's GCA. The parallel analysis showed that the eigenvalue of our factor (eigenvalue = 0.98, $n = 231$) is greater than the eigenvalue of a randomly created factor (eigenvalue = 0.21, $n = 231$), which suggests that GCA as a factor has meaningful exploratory value. We also performed a confirmatory factor analysis to ensure that the measured variables of learning would form a coherent latent variable. The model from the confirmatory factory analysis had a good fit to the data ($\chi^2_M = 7.04$, d.f. = 6, $p = 0.317$; RMSEA = 0.03; CFI = 0.96), with all measured variables having significant factor loadings ($p < 0.05$).

(c) Heritabilities of general cognitive ability

We estimated the heritabilities for each individual learning task, as well as for GCA derived from the exploratory factor analysis described above (table 1). The heritability for all mice combined was moderate–low, with a value of 0.24, $n = 231$, $p = 0.017$. By contrast, the mice from the enrichment group expressed a heritability of 0.15, and was not significantly different from zero, $n = 115$, $p = 0.284$, while the mice in the control group had a moderate–high heritability of 0.55, $n = 116$, $p = 0.017$. Thus, environmental enrichment was associated with a *decrease* in the estimate of the heritability of animals' GCA.

Table 2. Factor loadings and variance explained by the first factor (general cognitive ability, or intelligence) extracted from the five learning tasks using an exploratory factor analysis. $n = 231$.

learning task	general cognitive ability
Lashley maze	0.89
passive avoidance	0.33
T-maze	0.22
odour discrimination	0.15
spatial water maze	0.11
eigenvalue	0.98
proportion of common variance	19.5%

(d) Environmental effects on general cognitive ability

The means and standard deviations of GCA scores in all mice, in enrichment group, and the control group can be seen in table 1. The linear mixed model (with group treatments as a fixed effect, and sibling families as a random effect) revealed a significant effect of group treatment on mice's GCA ($t = 3.69$, $p < 0.001$). The estimate of the model showed an effect size of 0.39 (s.e. = 0.11). That represent a gain in 0.44 standard deviations in GCA for the mice in the enrichment group. This result suggests that experience with the enriched environment had a positive influence on animals' overall cognitive performance.

We also checked for the existence of gene–environment interactions by comparing a model with only a random

intercept of sibling families (independent effects) with a model with a random intercept of litter family and a random slope of group treatment (gene–environment interaction effects). The difference in LR test statistic between the models was 0.22, with degrees of freedom = 2. The *p*-value of this statistic was 0.896, which is not significant, and where the null hypothesis is the simpler model. The LR test indicates, therefore, that there was no gene–environment interaction because the independent effects model is more parsimonious and explains the data equally well.

4. Discussion

Here, we found that GCA in mice has substantial heritability and malleability. Overall, mice siblings had much more similar intelligence scores, i.e. siblings were more similar to one another. Mice that were exposed to enriched environments and physical exercise exhibited better performance than their siblings in a control group (that were maintained in the standard laboratory environment). To our knowledge, this is the first study to show in any non-human animal that a trait analogous to human's general intelligence is both substantially heritable and malleable. We also found that heritability itself changed between groups, a result that, as some researchers argue, sometimes can reflect gene–environment interactions. Unexpectedly, however, our tests showed no gene–environment interactions in our study. A closer examination of these results might help to clarify all of these conclusions.

Similar to our previous work, there was a positive correlation of each mouse's rate of acquisition across all learning tasks. The GCA factor accounted for 19.5% of the common variance in mice's performance. This is equivalent to accounting for 28% of the total variance from a principal component analysis, which is similar to what we have reported previously [16]. Relatedly, in prior work, we have determined that this general factor is unrelated to differences in stress reactivity, fear or anxiety [30,31]. Here, a parallel analysis showed that the first factor from the exploratory factor analysis had exploratory value much above one from a random dataset. Furthermore, a confirmatory factor analysis also revealed a good fit of the model with a single latent variable influencing all learning tasks of our battery, and the loadings were all significant. At first glance, 19.5% might seem a low value for a general cognitive factor in comparison to typical values of 50% in humans. Note, however, that the cognitive tasks composing modern human IQ tests are the result of a slow and gradual intentional selection for tasks that load well with others [37]. Tasks that had poor correlations with other tasks were changed or removed over the decades. The mouse learning battery we used here is not the culmination of a similar process, and thus reflect less of that 'bias'. In fact, the learning tasks in our battery were designed to be distinct in many parameters (described in Material and methods), and so the existence of a single factor that explains one-fifth of that variance is rather striking.

The present study combines a full-sibling design with a procedure loosely analogous to a human adoption study, or a randomized clinical trial, or school intervention. Two of the siblings in a litter of four were removed from their usual environment and experienced a new, more complex environment. In contrast with human adoption studies, here we had direct control of the environment into which some siblings were immersed. We found that mice in the environmental enrichment condition had GCA scores 0.44 standard deviations higher than their peers in the control group. In humans, this difference would represent 6.6 IQ points, which is a substantial and functionally important gain. At first glance, the gains we found might seem large in response to an 'intervention' which lasted for only 16 days. However, mice's typical life-span is less than 2 years, and a substantial part of their development occurs during the first 10 weeks of life, at which point they have reached sexual maturity. At birth, mice are hairless, blind, deaf, have minimal motor skills and are fully dependent on their mother, while by the sixth week, mice are already fully functional, with fine motor skills, a broad and complex repertoire of social behaviour and a remarkable capacity for learning [38]. In that context, 16 days of environmental enrichment during this maturation phase of development are probably quite meaningful. And, our environmental enrichment included substantial physical exercise. In total, this enrichment protocol was a dramatic intervention relative to the standard treatment of isolated laboratory mice.

Numerous theories have been proposed to account for the beneficial effect of environmental enrichment on cognition. Among them, the 'learning and memory' hypothesis seems to be favoured. This theory holds that when animals are confronted with novelty and environmental complexity, there are physiological and morphological changes that impact the mechanisms that underlie learning [25]. Physical exercise alone, however, stimulates synaptogenesis and neurogenesis, but does not seem to promote improvements in general intelligence [39]. This can be explained by the 'use it or lose it' paradigm in neuroscience: new neurons need to be recruited for a specific cognitive function if they are to last beyond a few days [40]. Therefore, if physical exercise had an influence on the intelligence gains that we found here, it is likely to have been a synergistic influence combined with the exposure to novel and complex stimuli. In past research, for example, we found that physical exercise alone did not improve mice's GCA, but when physical exercise was combined with cognitive training, the treatment had a greater effect than cognitive training alone [41].

Instead of (or in addition to) the conclusion that GCA was *helped* by environmental enrichment, it is possible that GCA was *harmed* by adverse environments (such as the sterile home conditions encountered by our control group). In typical laboratory conditions (as experienced by the control mice), mice are inhabiting an environment not expected by natural selection, while in our enrichment condition, mice encounter an environment that is a little closer to what their genotypes might be adapted to. Note, however, that not everything that is 'natural' is helpful, and not everything that is 'artificial' leads to harm. Mice in the laboratory environment have free and guaranteed access to food, water and shelter. In the wild, they do not. These 'artificial' experiences might reasonably be expected to *help* in promoting cognitive performance. However, laboratory mice are also deprived socially, are deprived sexually, are deprived from exploration, and from physical exercise (among other things). These 'artificial' experiences might reasonably be expected to *harm* cognitive performance. This leads to the question: which set of experiences matter more, the experiences that help, or the experiences that harm? We cannot answer this question with our current analysis. Nonetheless,

our results here show that GCA is substantially malleable—being helped by enrichment and/or harmed by adverse environments. We hope that future studies can more directly address the help/harmed distinction.

When considering all mice as part of one population, our estimate of heritability was 0.24 (or 24%), a moderate–low range comparable to the value (between 0.34 and 0.42) obtained by the only other study that estimated heritability of general intelligence in mice [7]. Note that both ours and this other study used sibling designs, and so the family effect we interpret as 'genetic' also include any shared early environment effect, such as in the womb. Because of that, our heritability estimate is an upper-limit heritability, and will reflect maximum genetic influences in the case of minimum maternal/litter effects [35]. In other animals (primarily primates), estimates of GCA tend to be moderate with values ranging from 0.3 to 0.6 [1]. Like ours, some of these primate studies also using sibling designs, and hence the estimates that they generate reflect upper-limit heritabilities.

The separate learning tasks in our study had similar heritabilities compared with GCA and among each other (with the exception of the T-maze alternation). This result, combined with the good model fit of a single factor explaining the variation in the learning tasks, supports the view of a general influence on cognitive abilities. By contrast, Sorato *et al*. (this edition) [42] found no covariation between discriminative and reversal learning tasks, which, as the authors argue, support the view of independent modularity. Of course, these two views are not mutually exclusive, as domain-specific cognitive abilities (modularity) might share domain-general processes (general intelligence) [1]. More empirical evidence will be critical to determine where in the modularity spectrum particular species and environments fall.

When considering mice in the study as part of two different populations, the enrichment group had an estimated heritability not significantly different from zero, while mice in the control group had a moderate heritability of 0.55. Interestingly, estimates of the heritability of intelligence in humans also seem to change across environments. A recent meta-analysis by Tucker-Drob *et al*. [43] showed that among affluent families, most of IQ's variation was associated with genetic variation (heritability of 0.70). However, among the poorest families, the reverse was true: most of variation in IQ was associated with the shared familial environment, and little of the variation was attributable to genetic variation (heritability of 0.10). Those authors and others argue that changes in heritability are likely to be cases of gene–environment interactions; what is sometimes described as the bioecological model of intelligence [44].

Note, however, that the direction of the changes in heritability that we observed here (in response to environmental enrichment) were opposite those that would be expected based on studies of humans. Mice exposed to the more complex environment had a *lower* estimated heritability than mice maintained in the more sterile environment. In human populations, gene–environment interactions in wealthy groups are believed to inflate the estimates of heritability [8]. Typical methods in quantitative genetics usually give priority to genetics, and so gene–environment interactions are counted as independent genetic effects [4]. In our study, however, there was no gene–environment interaction (discussed below). A possible explanation for our results is that mice in the enrichment group showed lower heritability because of more independent environmental variance, while independent genetic variance remained the same. Because the enrichment group had a more complex environment than the control group, this has the potential of decreasing the estimate of heritability of the enrichment group. Regardless of its source, the present results highlight the sensitivity of estimates of heritability to the environment in which the estimate is obtained.

To our surprise, there was no interaction between family (genetic) effects and group (environment) effects. Even though we did not find direct evidence for gene–environment interactions, it is important to note that gene–environment correlations might still have exerted some influence, given our finding that heritability changed across the two environments. To directly test for these correlations and the 'snowballing' influence that they can foster, however, would require us to specify what environmental factors influence intelligence, and also restrict individuals with particular genes to get more/less of specific environments without correlating it with confounding factors. Of course, this is much more feasible to be tested in non-human, laboratory animals and future studies might well follow such a strategy.

The results here could help laying the groundwork for future studies identifying specific genes, neural and developmental mechanisms associated with GCA, as well as the development of interventions to improve cognition. A clear understanding of the causes of variation of intelligence is also critical for us to know how different species adopted different cognitive capacities, what the related selective pressures were and how intelligence differs across populations or species.

Ethics. All experiments were conducted in accordance with protocols approved by the Rutgers University IACUC.

Data accessibility. All data are available at: Research Gate repository http://dx.doi.org/10.13140/RG.2.2.25565.10723. All raw data are available at https://www.researchgate.net/publication/323128667_Data_-_Heritability_and_Malleability_of_mouse_intelligence_2018.

Authors' contributions. B.S.: conception and design, acquisition of data, and analysis and interpretation of data; drafting the article and revising it critically for important intellectual content; and final approval of the version to be published. S.B.: (i) conception and design, and acquisition of data; (ii) revising the article critically for important intellectual content; and (iii) final approval of the version to be published. M.H.: (i) conception and design, and acquisition of data; (ii) revising the article critically for important intellectual content; and (iii) final approval of the version to be published. D.S.: (i) acquisition of data; (ii) revising the article critically for important intellectual content; and (iii) final approval of the version to be published. C.S.: (i) acquisition of data; (ii) revising the article critically for important intellectual content; and (iii) final approval of the version to be published. S.R.: (i) acquisition of data; (ii) revising the article critically for important intellectual content; and (iii) final approval of the version to be published. J.K.: (i) acquisition of data; (ii) revising the article critically for important intellectual content; and (iii) final approval of the version to be published. L.D.M.: (i) conception and design, acquisition of data, interpretation of data; (ii) revising the article critically for important intellectual content; and (iii) final approval of the version to be published.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by grants from the National Institute of Mental Health (R03MH108706), the Busch Foundation and the Office of Naval Research (N000141210873).

Acknowledgements. Our thanks to Tracey Shors, Edward Selby and Christopher Chabris for the helpful suggestions and comments on this work.

References

1. Burkart JM, Schubiger MN, van Schaik CP. 2017 The evolution of general intelligence. *Behav. Brain Sci.* **40**, e195. (doi:10.1017/S0140525X16000959)
2. Neisser U *et al.* 1996 Intelligence: knowns and unknowns. *Am. Psychol.* **51**, 77–101. (doi:10.1037/0003-066X.51.2.77)
3. Matzel LD, Sauce B. 2017 Individual differences: case studies of rodent and primate intelligence. *J. Exp. Psychol. Anim. Learn. Cogn.* **43**, 325–340. (doi:10.1037/xan0000152)
4. Tenesa A, Haley CS. 2013 The heritability of human disease: estimation, uses and abuses. *Nat. Rev. Genet. Nat. Publ. Group* **14**, 139–149. (doi:10.1038/nrg3377)
5. Deary IJ, Spinath FM, Bates TC. 2006 Genetics of intelligence. *Eur. J. Hum. Genet.* **14**, 690–700. (doi:10.1038/sj.ejhg.5201588)
6. Hopkins WD, Russell JL, Schaeffer J. 2014 Chimpanzee intelligence is heritable. *Curr. Biol.* **24**, 1649–1652. (doi:10.1016/j.cub.2014.05.076)
7. Galsworthy MJ, Paya-Cano JL, Liu L, Monleón S, Gregoryan G, Fernandes C, Schalkwyk LC, Plomin R. 2005 Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behav. Genet.* **35**, 675–692. (doi:10.1007/s10519-005-3423-9)
8. Sauce B, Matzel LD. 2018 The paradox of intelligence: heritability and malleability coexist in hidden gene-environment interplay. *Psychol. Bull.* **144**, 26–47. (doi:10.1037/bul0000131)
9. van Ijzendoorn MH, Juffer F, Poelhuis CWK. 2005 Adoption and cognitive development: a meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychol. Bull.* **131**, 301–316. (doi:10.1037/0033-2909.131.2.301)
10. Finkel D, Pedersen NL. 2016 Gene–environment interplay in adulthood. *Behav. Genet.* **46**, 1–3. (doi:10.1007/s10519-015-9775-x)
11. Galsworthy M, Arden R, Chabris C. 2014 Animal models of general cognitive ability for genetic research into cognitive functioning. In *Behavior genetics of cognition across the lifespan advances in behavior genetics* (eds D Finkel, C Reynolds), pp. 257–278. New York, NY: Springer.
12. Sauce B, Matzel LD. 2013 The causes of variation in learning and behavior: why individual differences matter. *Front. Psychol.* **4**, 395. (doi:10.3389/fpsyg.2013.00395)
13. Christian KM, Thompson RF. 2003 Neural substrates of eyeblink conditioning: acquisition and retention. *Learn. Mem.* **10**, 427–455. (doi:10.1101/lm.59603)
14. Eichenbaum H. 2000 A cortical–hippocampal system for declarative memory. *Nat. Rev. Neurosci.* **1**, 41–50. (doi:10.1038/35036213)
15. Matzel LD, Han YR, Grossman H, Karnik MS, Patel D, Scott N, Specht SM, Gandhi CC. 2003. Individual differences in the expression of a 'general' learning ability in mice. *J. Neurosci.* **23**, 6423–6433 (doi:10.1523/JNEUROSCI.23-16-06423.2003)
16. Matzel LD, Sauce B, Wass C. 2013 The architecture of intelligence: converging evidence from studies of humans and animals. *Curr. Dir. Psychol. Sci.* **22**, 342–348. (doi:10.1177/0963721413491764)
17. Wass C, Denman-Brice A, Rios C, Light KR, Kolata S, Smith AM, Matzel LD. 2012 Covariation of learning and 'reasoning' abilities in mice: evolutionary conservation of the operations of intelligence. *J. Exp. Psychol. Anim. Behav. Process* **38**, 109–124. (doi:10.1037/a0027355)
18. Kolata S, Light K, Matzel LD. 2008 Domain-specific and domain-general learning factors are expressed in genetically heterogeneous CD-1 mice. *Intelligence* **36**, 619–629. (doi:10.1016/j.intell.2007.12.001)
19. Kolata S, Light K, Townsend DA, Hale G, Grossman HC, Matzel LD. 2005 Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiol. Learn. Mem.* **84**, 241–246. (doi:10.1016/j.nlm.2005.07.006)
20. Blinkhorn S. 2003 Neuroscience: of mice and mentality. *Nature* **424**, 1004–1005. (doi:10.1038/4241004a)
21. Kolata S, Light K, Wass CD, Colas-Zelin D, Roy D, Matzel LD. 2010 A dopaminergic gene cluster in the prefrontal cortex predicts performance indicative of general intelligence in genetically heterogeneous mice. *PLoS ONE* **5**, e14036. (doi:10.1371/journal.pone.0014036)
22. Wass C, Pizzo A, Sauce B, Kawasumi Y, Sturzoiu T, Ree F, Otto T, Matzel LD. 2013 Dopamine D1 sensitivity in the prefrontal cortex predicts general cognitive abilities and is modulated by working memory training. *Learn. Mem.* **20**, 617–627. (doi:10.1101/lm.031971.113)
23. Wass C, Sauce B, Pizzo A, Matzel LD. 2018 Dopamine D1 receptor density in the mPFC responds to cognitive demands and receptor turnover contributes to general cognitive ability in mice. *Sci. Rep.* **8**, 4533. (doi:10.1038/s41598-018-22668-0)
24. Fabel K, Wolf SA, Ehninger D, Babu H, Leal-Galicia P, Kempermann G. 2009 Additive effects of physical exercise and environmental enrichment on adult hippocampal neurogenesis in mice. *Front. Neurosci.* **3**, 50. (doi:10.3389/neuro.22.002.2009)
25. van Praag H, Kempermann G, Gage FH. 2000 Neural consequences of environmental enrichment. *Nat. Rev. Neurosci.* **1**, 191–198. (doi:10.1038/35044558)
26. Aldinger KA, Sokoloff G, Rosenberg DM, Palmer AA, Millen KJ. 2009 Genetic variation and population substructure in outbred CD-1 mice: implications for genome-wide association studies. *PLoS ONE* **4**, e4729. (doi:10.1371/journal.pone.0004729)
27. Sauce B, Wass C, Smith A, Kwan S, Matzel LD. 2014 The external–internal loop of interference: two types of attention and their influence on the learning abilities of mice. *Neurobiol. Learn. Mem.* **116**, 181–192. (doi:10.1016/j.nlm.2014.10.005)
28. Hoaglin DC, Iglewicz B. 1986 Performance of some resistant rules for outlier labeling. *J. Am. Stat. Assoc.* **81**, 991–999. (doi:10.1080/01621459.1986.10478363)
29. Schafer JL. 1999 Multiple imputation: a primer. *Stat. Methods Med. Res.* **8**, 3–15. (doi:10.1191/096228099671525676)
30. Matzel LD, Townsend DA, Grossman H, Han YR, Hale G, Zappulla M, Light K, Kolata S. 2006 Exploration in outbred mice covaries with general learning abilities irrespective of stress reactivity, emotionality, and physical attributes. *Neurobiol. Learn. Mem.* **86**, 228–240. (doi:10.1016/j.nlm.2006.03.004)
31. Grossman HC, Hale G, Light K, Kolata S, Townsend DA, Goldfarb Y, Kusnecov A, Matzel LD. 2007 Pharmacological modulation of stress reactivity dissociates general learning ability from the propensity for exploration. *Behav. Neurosci.* **121**, 949–964. (doi:10.1037/0735-7044.121.5.949)
32. O'connor BP. 2000 SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instrum. Comput.* **32**, 396–402. (doi:10.3758/BF03200807)
33. Kline RB. 2011 *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
34. Hu L, Bentler PM. 1999 Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. A Multidiscip. J.* **6**, 1–55. (doi:10.1080/10705519909540118)
35. Falconer DS. 1989 *Introduction to quantitative genetics*, 3rd edn. New York, NY: Longman.
36. Tom AB, Bosker TASRJ, Bosker RJ. 1999 *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Beverley Hills, CA: Sage.
37. Mackintosh N. 2011 *IQ and human intelligence*. Oxford, UK: Oxford University Press.
38. Weber EM, Olsson IAS. 2008 Maternal behaviour in *Mus musculus* sp.: an ethological review. *Appl. Anim. Behav. Sci.* **114**, 1–22. (doi:10.1016/j.applanim.2008.06.006)
39. Daugherty AM, Zwilling C, Paul EJ, Sherepa N, Allen C, Kramer AF, Hillman CH, Cohen NJ, Barbey AK. 2017 Multi-modal fitness and cognitive training to enhance fluid intelligence. *Intelligence* **66**, 32–43. (doi:10.1016/j.intell.2017.11.001)
40. Shors TJ, Anderson ML, Curlik DM, Nokia MS. 2012 Use it or lose it: how neurogenesis keeps the brain fit for learning. *Behav. Brain Res.* **227**, 450–458. (doi:10.1016/j.bbr.2011.04.023)
41. Smith AM, Spiegler KM, Sauce B, Wass CD, Sturzoiu T, Matzel LD. 2013 Facilitation of the cognitive

enhancing effects of working memory training through conjoint voluntary aerobic exercise. *Behav. Brain Res.* **256**, 626–635. (doi:10.1016/j.bbr.2013.09.012)

42. Sorato E, Zidar J, Garnham L, Wilson A, Lovlie H. 2018 Heritabilities and co-variation among cognitive traits in red junglefowl. *Phil. Trans. R. Soc. B* **373**, 20170285. (doi:10.1098/rstb.2017.0285)
43. Tucker-Drob EM, Briley DA, Harden KP. 2013 Genetic and environmental influences on cognition across development and context. *Curr. Dir. Psychol. Sci.* **22**, 349–355. (doi:10.1177/0963721413485087)
44. Bronfenbrenner U, Ceci SJ. 1994 Nature-nuture reconceptualized in developmental perspective: a bioecological model. *Psychol. Rev.* **101**, 568–586. (doi:10.1037/0033-295X.101.4.568)