THE GENERAL COGNITIVE ABILITY OF MICE HAS MODERATE HERITABILITY AND IS INFLUENCED BY ENVIRONMENTAL FACTORS

By

BRUNO SAUCE SILVA

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Louis D. Matzel

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2018

ABSTRACT OF THE DISSERTATION

THE GENERAL COGNITIVE ABILITY OF MICE HAS MODERATE HERITABILITY AND IS INFLUENCED BY ENVIRONMENTAL FACTORS

by BRUNO SAUCE SILVA

Dissertation Director:

Louis D. Matzel

Among humans, estimates of the heritability of intelligence are high, although there is also evidence that this trait is quite malleable. However, little is known about the genotypes and environmental states that are relevant to intelligence, much less how they react together. This is in part due to limitations on work with humans, which are largely confined to assessing heritability and malleability using only correlational methods. In the past, our group has developed a battery of cognitive tasks in mice where the performance of individuals is correlated across all tests. Aggregate performance on this battery of tests predicts individuals' performance on other cognitive abilities implicated in human general intelligence such as reasoning and working memory. Thus it has been asserted that the "general cognitive ability" (GCA) scores obtained in this battery are analogous to human IQ scores. Here we attempted to answer a critical question: How much the individual differences in mouse intelligence can be influenced by genetics and the environment? To make these determinations, we used a strategy analogous to classic human twin/adoption studies. Unlike human studies, in the present case, different

environments were under experimental control and thus could be administered differently to different members of sibling cohorts. We used 232 outbred male mice consisting of 58 families of 4 full siblings (fraternal twins) and provided different environments to the siblings from each family (analogous to "twins raised apart"). For the environmental manipulation, half of the sibling groups stayed in a standard colony room with little stimulation, while the other half of the siblings were exposed to environments that are prone to influence cognitive development: physical exercise and daily exposure to novel experiences. Similar to our previous work, here GCA (i.e., intelligence) accounted for 19.5% of the common variance in mice's performance from an exploratory factor analysis. We found that GCA was moderately malleable, with environmental enrichment increasing GCA scores by 0.44 standard deviations. In humans, this increase would represent an increase of 6.6 IQ points. Although here we used a combination of environmental factors that were likely to be effective, these were admittedly limited in scope and strength, and so GCA can potentially be more malleable than these results suggest. The population of all mice combined showed a GCA heritability of 0.24, quite comparable to values estimated by others in rodents and non-human primates. To our surprise, our Enrichment group had a heritability not significantly different from zero, while our Control group had a moderate heritability of 0.55. This is the opposite of what is typically found in humans, where heritability is higher in populations drawn from higher SES environments. Also, unexpectedly, we did not find evidence for G×E in our study. The results from the current study could help laying the groundwork for future studies on the independent effects and interactions between environment and genes in shaping general intelligence in humans.

Acknowledgement

This work was supported by grants from the National Institute of Aging (R01AG029289), the Busch Foundation, and the Office of Naval Research (N000141210873).

I would like to thank the members of my committee, Tracey Shors, Edward Selby, and Christopher Chabris for their help and insightful comments. I would also like to express my enormous gratitude to my advisor, Louis D. Matzel, for all the guidance, motivation, knowledge, and precious inspiration over the years of my PhD studies.

Table of Contents

List of tables

List of illustrations

Introduction

Individuals vary in their intelligence, and these differences are immediately apparent to even casual observers. More technically, the American Psychological Association (Neisser et al., 1996) described differences in intelligence as the differences in "the ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, and to engage in various forms of reasoning" (p. 81). Although this definition is somewhat vague, the impact of intelligence on academic, career, and social success is empirically well established (Gottfredson, 1998).

People who perform well in one cognitive test (e.g., mental arithmetic) tend to perform well in other tests (e.g., spatial rotation). This consistent pattern of correlations is the basis for the concept of general intelligence, or "g", proposed originally by Charles Spearman in 1904 (Spearman, 1904). Standard methods for assessing intelligence in humans (e.g., the Wechsler Adult Intelligence Scale; The Stanford-Binet) reflect this concept, using an individual's aggregate performance across diverse and varied cognitive tests to estimate his/her IQ score. These standardized methods have determined that 25-50% of the variance among individuals' performance in diverse cognitive tasks can be accounted for by the single factor of general intelligence (Plomin, 2001; Sternberg & Kaufman, 1998).

There is a substantial degree of genetic influence on differences in intelligence. In fact, intelligence is one of the most heritable psychological phenotypes (Bouchard, 2004). Heritability is a statistic that captures how much of the variation on a trait is due to genetic differences. The most common and cost-effective approach to estimate heritability is by using genetic relatedness as a proxy for genes *per se*. (Genetic

relatedness is easier to measure because we typically cannot specify the genes behind a trait, but we *can* identify daughters, uncles, siblings, or other relatives.) Studies on heritability look for special cases where environmental effects can be somehow minimized or at least balanced, such as cases of adoption, families with identical twins, or families with siblings born at the same time. (For more on the methods for estimating heritability, see Tenesa & Haley, 2013).

Heritability can be estimated for any trait, and it ranges from 0.0 (meaning that the trait has no genetic component) to 1.0 (meaning that the trait is completely heritable). For example, the heritability of breast cancer is reported to be 0.27; the heritability of body mass index is 0.59; and the heritability of Type 1 diabetes is 0.88 (Hyttinen, Kaprio, Kinnunen, Koskenvuo, & Tuomilehto, 2003; Lichtenstein et al., 2000; Silventoinen, Magnusson, Tynelius, Kaprio, & Rasmussen, 2008). Using similar methods, the heritability of general intelligence is estimated to range from 0.60 to 0.80 in adults (Davies et al., 2011; Deary, Penke, & Johnson, 2010). To put this value in perspective, note that the heritability of other psychological traits described as "highly heritable" rarely approach the degree of heritability of IQ (see Bouchard, 2004, for extensive examples). The most comparable is schizophrenia (heritability of 0.64; Lichtenstein et al., 2009), while alcoholism (heritability of 0.50), neuroticism (heritability of 0.48; Riemann et al., 1997), and major depression (heritability of 0.40; Sullivan et al., 2000) are markedly lower.

The influence of genes on IQ, though undeniable, might not be as powerful and constrictive as many believe. There is evidence in humans suggesting that despite its high heritability, intelligence is also quite malleable, and that changes in the environment explain a great deal of IQ's variation across families and socioeconomic status (Sauce & Matzel, 2017).

In adoption studies, for example, researchers can compare an adopted child with his/her biological siblings who were not adopted, or with other peers who were left behind. Being raised by different parents changes not only a child's family environment, but also all factors contributed by the neighborhood, peer, and school environments. These often-drastic environmental changes make adoption studies particularly useful to assess the malleability of intelligence in humans. A meta-analysis of 62 adoption studies from a multitude of countries (totaling 18,000 adopted children) found an average increase in IQ of 17.6 points within several years of adoption (van Ijzendoorn et al., 2005). That is a remarkable cognitive gain over their biological, nonadopted siblings and their peers who stayed behind. To put that value in perspective, keep in mind that standard IQ tests set 100 points as the age/population average test score, and 15 points as the age/population standard deviation (so, an IQ of 130 is at the 97th percentile). For example: a successful college graduate has, on average, 15 units of IQ above the average of 100; and a child is considered "gifted" in the US (and goes to special schools) if he/she has 30 units of IQ above the average value of 100.

Interestingly, estimates of the heritability of intelligence also seem to change across socio-economic status. In a seminal study, Turkheimer et al. estimated genetic and environmental effects on IQ in 7-year-old twins in high and low SES families

(Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). The authors reported that among affluent families, most of IQ's variation was associated with genetic variation (heritability of 0.72). 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). These results suggest that environmental differences in families across SES influence the genetic potential for intelligence. Differences in genes are more accentuated in favorable environments, while on the other extreme, differences in familial environment are strongest for IQ's variation among poor families. And, regardless of the direction, changes in heritability resulting from changes in the environment are likely to be cases of G×E interactions and correlations. It's hard to imagine how independent genetic effects alone could differently affect wealthy and poor communities so to create the observed heritability changes across SES.

A decade prior to the report by Turkheimer et al. (2003), Bronfenbrenner and Ceci (1994) predicted changes in heritability in relation to SES and the importance of GE interplay in what is described as their bioecological model of intelligence. This same pattern was confirmed empirically in a recent meta-analysis by Tucker-Drob, Briley, & Harden (2013), where they claim GE interplay to be behind the process described as the "transactional model" of heritability.

Although general intelligence is a single, abstract construct, it needn't be a single, concrete property of the brain. Instead, the higher-order construct of g may emerge from the interplay between a multitude of systems, including genetic networks, influences from psychological traits like working memory capacity, attentional capacity, motivation

and personality, and reciprocal relations between cognitive and environmental processes (as held, for example, by on the "mutualism" model of intelligence originally proposed in van der Maas et al., 2006). Intelligence is a hugely complex trait, and we should expect a high number of moving parts behind differences in its expression across individuals.

Independent environmental effects as well as G×E interactions (as well as correlations, which we don't address here) are attributes of a trait's "malleability". The more the environment causes (directly or indirectly) the variation in a trait, the higher the trait's malleability. But what exactly are these effects? Does, for example, an independent genetic effect imply that a single isolated gene produces the trait? In any literal sense, this cannot be true. If you put a strand of DNA double helix in a tube and wait, no distinguishable trait will emerge from it. And what about the gene-environment interactions (G×E)? Does this merely refer to the evident truth that an egg needs both genes and environment to create an adult chicken? Again, the answer is "no". The effects that we are referring to here are related to *causes of variation*, not causation.

To illustrate the difference between causation and causes of variation, let us imagine that a building fire has resulted in dozens of casualties. An ensuing investigation determined that the fire was started by a dropped cigarette that ignited paper that was stored in the building. The cigarette and papers, therefore, are the proximal "causes" of the fire. Remove either the source of ignition (the cigarette) or the source of fuel (stored papers) from this particular fire, and the fire would not have ignited or spread. However, the chief firefighter of the city knows through experience that some causes are more important than others. The fire chief is particularly concerned about stored papers in her city's extremely dry climate, and less concerned about the behavior of smokers. In most

instances, in that city the fires are caused by big piles of papers that get so dry that any trivial ignition of any type will suffice to start a fire. Reducing the piles of papers, the firefighter knows, is *the way* to save hundreds of future lives from fire, while prohibiting smoking in buildings would be of only minor consequence. Thus in one sense, the stored paper can be described as an "equal contributor", but in another sense, could be described as the "principal cause" of deadly fire.

Although a fire is necessarily a product of both ignition and fuel, a particular cause/factor can be more critical to the "causes of variation" of fire (differences in occurrence and intensity) among a particular set of buildings (i.e., the *environment* is a critical cause of variation). In Silicon Valley, for instance, overloaded electrical systems may be a more relevant factor than stored paper in the variation of fires, and defective heaters will not be relevant at all in the tropical city of Mumbai. In other words, the causes of *variation* for some event are not the same as its *causation* (we used this and other analogous examples in Sauce & Matzel, 2013, where you can also find a more in depth treatment on "causes"). Heritability is a measure of causes of variation, and not a measure of what causes the trait. As with the fires, the analysis of causes of variation of intelligence can give valuable insights into how much and in which way factors interact in the real world to make some people brilliant while some others are cognitively impaired.

In light of what causes of variation mean, we can now return to defining independent and interaction effects. Independence (or additivity) is conceptually analogous to what a statistician would describe as the "main effect", i.e., it is how much a predictor/input variable (a gene) relates to a predicted/output variable (a trait) when

averaging across the other predictor variables (environment and other genes). A G×E effect is conceptually close to a statistical interaction — it is when the effect of a variable (gene) with another variable (an environmental factor) has multiplicative consequences, with the result (variation in a trait) being more than the sum of its parts. More concretely (and of particular importance), in G×E interactions, 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. It is this role of the G×E interaction that is difficult to capture in estimates of heritability, and might explain why a highly heritable trait can also be malleable.

Although the role of gene-environment interaction is not a new idea, even within the field of behavior genetics the implications of these effects for the variation of traits is just beginning to be appreciated. Gene-environment interplay in intelligence can be especially elusive, and is often disregarded or ignored, even by many in the immediate field. In a recent editorial (Finkel & Pedersen, 2016), the journal *Behavior Genetics* pointed to the novelty and immediate topical importance of gene-environment studies: "The Behavior Genetics Association has seen an increase in presentations addressing various forms of GE interplay, not the least at the association's annual meetings. For example, in 1995 there were only two papers at the annual meeting that addressed GE interplay, and at that time most of the work was conducted using mouse models of behavior. By the 2015 conference, there were 18 papers and posters addressing GE interplay, using a variety of methodologies and applied to several phenotypes across the lifespan." Despite its elusive nature,, the study of gene-environment interplay is starting

to gain momentum in our current scientific age of big data, complex networks, and systems biology (Chandler, Chari, & Dworkin, 2013; Geschwind & Konopka, 2009; Lazer et al., 2009; Philip Chen & Zhang, 2014; Rockman, 2008). Researchers have recently found traits in animals, plants, and microorganisms that are under strong influences of gene-environment interplay (Bhatia et al., 2014; El-Soda, Kruijer, Malosetti, Koornneef, & Aarts, 2015; E. N. Smith & Kruglyak, 2008; Volkers et al., 2013; Wolf, Vaughn, Pletscher, & Cheverud, 2002).

Although studies have established that a high proportion of individual variations in intelligence emerges from genetic influences, specific genetic determinants are unknown (Deary, Johnson, & Houlihan, 2009) and false positive identifications have been exceedingly common (Chabris et al., 2012). If gene-environment interactions play a substantial role in IQ differences, maybe the failure to identify specific genes associated with this trait (i.e., "missing heritability") reflects the failure to account for such interplays. We do not know much about the genotypes and environmental states that are relevant to intelligence, much less how they react together (Bailey, 1997). Because of this ignorance (as well as limitations on direct manipulations of the environment), the effects of G×E are mostly beyond the reach of current (and ethically permissible) methods in human studies. And as pointed by Tenesa and Haley (Tenesa & Haley, 2013), these flaws lead to an overestimation of genetic effects. In other words, when estimating the heritability of IQ, those gene-environment interplays that we don't recognize or can't manipulate are usually attributed to the genetic component. Consequently, that could be inflating the estimates of IQ's heritability. (For a more sophisticated treatment on the

genetics and mathematics of this issue regarding hypothetical traits, see Templeton, 2006.)

The difficulties associated with tests of GE interplay 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. In contrast, with laboratory animals, we can easily control the environment, and can combine correlational and experimental designs. Hence, studies with animals might reveal gene-environment effects to account for problems such as the "missing heritability" of intelligence. Animal models could be of great value in understanding the malleability and G×E in cognitive abilities. However, most animal studies have primarily focused on the processes and mechanisms which underlie single domains of learning (e.g., spatial learning). While this approach has proven fruitful in delineating certain neurobiological substrates of these processes (Christian & Thompson, 2003; Eichenbaum, 2000; Phelps & LeDoux, 2005), they fail to capture how these systems work in conjunction with one another to create individual differences in general cognitive skills.

In previous research by our group, genetically diverse mice were tested on batteries of five to nine cognitive tasks, each of which made unique sensory, motor, motivational, and information processing demands on the animals (Matzel et al., 2003, 2006). These test batteries were analogous to the design of "classic" human intelligence tests, where various tasks impinged on different information processing skills. The tasks were designed to be rudimentary in nature (e.g., fear conditioning, passive avoidance, odor discrimination, and spatial navigation). As a consequence, all tested animals could attain comparable levels of performance, but could do so with different efficiencies. Mice

that perform well in one task of the battery tend to do so in the other tasks of the battery – revealing a positive correlation of each animal's rate of acquisition across all tasks. Factor analyses have indicated that 32% to 48% of the variance across tasks was attributable to a single factor, which in our first studies, was described as "general learning ability."

Subsequent studies performed in our laboratory showed that general learning abilities in mice also co-varies with other cognitive abilities implicated in human general intelligence. Performance on inductive and deductive reasoning tasks was strongly correlated with animals' aggregate performance in the learning battery (Wass et al., 2012), as well as performance in working memory and attentional tasks (Kolata et al., 2005; Kolata, Light, Grossman, Hale, & Matzel, 2007). And in a comprehensive test of 264 mice, our group reported a hierarchical structure of this general cognitive abilities in mice, i.e., a general factor influenced domain-specific factors, such as spatial abilities (Kolata, Light, & Matzel, 2008). As described above, this hierarchy of abilities is a hallmark of general intelligence in humans. In fact, others described our results as qualitatively analogous to what is described in humans as "intelligence" (Blinkhorn, 2003).

Having established the existence of a general cognitive ability in genetically heterogeneous mice, we later asked questions about the neural and genetic factors regulating those individual differences. Kolata et al. (2010) characterized the general intelligence of 60 outbred mice, and quantified the expression of approximately 25,000 genes in specific brain areas. In that study, our group reported a difference in expression of ten genes among mice classified of high intelligence (compared to mice of low

intelligence). Of those ten genes, three were of particular interest (Drd1a, Darp-32, and Rgs9) due to their ability to modulate dopaminergic signaling cascades (Kolata et al., 2010), as well as learning and synaptic plasticity (Calabresi et al., 2000; Genoux et al., 2002). Dopamine is well known to be involved in executive attention and working memory, two processes highly linked to general intelligence. Our group later assessed dopamine signaling in prefrontal networks of mice classified for their general intelligence (Wass et al., 2013). In that study, we observed a significant correlation between dopamine-induced activity and general intelligence in the prefrontal cortex. In other words, D1 receptive neurons of high intelligence mice have an enhanced sensitivity level compared to low intelligence mice. The genetic and neurological studies combined suggest that the individual differences in a mouse's IQ have a strong biological basis, and, therefore, might have a moderate to high heritability.

In a study by another group of investigators (Galsworthy et al., 2005), a sibling analysis was used to estimate the heritability of intelligence in mice. Although the test battery administered by Galsworthy et al. included a few tasks with minimal cognitive requirements (e.g. marble burying), their battery also included interesting reasoning tasks, and they isolated a consistent single factor regulating the animals' performance across the battery of tests. By looking at the correlation between sibling pairs, Galsworthy et al. estimated a heritability of intelligence in those mice as 0.24. Although this value is nominally lower than we would expect given previous studies of humans, it is still quite meaningful.

In addition to evidence pointing towards genetic contributions to individual difference in mice's IQ, there is also evidence for environmental effects. Recently, we

found that mice's general intelligence (as well as the sensitivity of dopamine-receptor neurons in the prefrontal cortex) is improved through cognitive training (Wass et al., 2013). This indicates that intelligence in mice is not a "fixed" ability. In another study, we found that the effects of cognitive training are greatly enhanced if combined with physical exercise. In other words, mice under a regimen of cognitive stimulation and aerobic exercise undergo greater improvements in their general intelligence than was observed in response to either intervention in isolation (A. M. Smith et al., 2013). These findings are supported by a neurobiological study that found a combination effect between exercise in the home cage and novel environments on neurogenesis in the hippocampus and retention of these new neurons (Fabel et al., 2009). There is also evidence that the brain of rodents get bigger from environmental experience (Walsh, Cummins, & Budtz-Olsen, 1973), and that cognitive performance relates to brain size (Anderson, 1993; Jensen & Fuller, 1978). In addition, a review by van Praag, Kempermann, & Gage (2000) 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. Thus, these studies seem to suggest that physical exercise paired with novel experiences (complex stimulation) might be the key to better cognitive ability.

Building on current knowledge of intelligence in both humans and mice, here we attempted to answer a critical question: How much the individual differences in mouse intelligence can be influenced by genetics and the environment? For this, we used groups of full siblings (fraternal twins) mice and provided different environments to subsets of each sibling cohort. For the environmental manipulation, half of the sibling groups stayed

in a standard colony room with little stimulation, while the other half of the siblings were exposed to environments that are prone to affect their intelligence: physical exercise and daily exposure to novel experiences. At the end, we assessed all mice on a learning battery designed to characterize their general cognitive performance. In short, our study combined the design of a sibling study with an "adoption" study (where "adoption" was under experimental control), and that allowed us to better assess heritability, environmental effects, as well as G×E effects on the expression of intelligence. This type of study combines correlational and experimental approaches, and is extremely hard to perform in humans due to practical and ethical limitations.

A specific prediction of the current study was that general cognitive ability would have a moderate to high heritability. And we estimated heritability not only in a standard population of lab mice, but also of lab mice under special environmental enrichment. In humans, estimates of heritability tend to be higher in wealthier relative to impoverished communities, and thus we expected mice raised in an enriched environment to support a higher estimate of heritability as well. Another specific prediction was that the enriched environment given to one of the groups of siblings will have a beneficial impact in mice's general cognitive ability. In other words, we expected mouse intelligence to be malleable. We also predicted that brain weight would be higher in the environment enrichment group. Finally, we predicted that there would be G×E interactions for general cognitive ability. We believe that manipulating the environment among sibling sets was a unique opportunity to directly test GE interplay – especially of the G×E effects, as mice within families will have similar genetic background and experience different environments,

while mice between families will have different genetic backgrounds and experience similar environments.

Materials and Methods

Subjects

We used 232 CD-1 outbred male mice from Harlan Laboratories (Indianapolis, IN). Since this strain exhibits a robust amount of behavioral as well as genetic variability, they are well suited for experiments examining individual differences. Estimates of genetic variation in this line have indicated that despite over 50 years of breeding they are very similar to wild mouse populations (Aldinger, Sokoloff, Rosenberg, Palmer, & Millen, 2009). The mice arrived in our laboratory between 4-5 weeks of age at which time their weight ranged from 25-30 grams. They were singly housed in clear standard shoe box cages in a temperature and humidity controlled colony room under a 12-hour light/dark cycle. In order to minimize any differential stress responses due to experimenter handling, we handled the animals for 90 seconds a day for a period of seven days.

The 232 mice were comprised of 58 sets of four siblings (fraternal quadruplets), totaling 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 control home environment (Control Group) and the two other siblings received the environmental "enrichment" treatment consisting of physical exercise and exposure to novel and engaging environments (Enrichment Group).

Except for the food-dependent learning tasks, all mice had continuous access to both food and water. For the two tests requiring food deprivation, ad libitum food were removed from the animals' home cages at the end of the light cycle on the day before the rest day that precedes the start of training. During the deprivation period, animals were

provided with food in their home cages for 90 min per day during the last two hours of the light cycle, and thus they were food deprived for 16 hours at the time of training or testing. Although mild, this level of deprivation was sufficient to maintain stable performance on these tasks. All experiments were conducted in accordance with protocols approved by the Rutgers University IACUC (Institutional Animal Care and Use Committee).

Exploration

All mice received two tests of exploration: The Open Field at the start of the study (prior to differential treatments) and the Elevated plus Maze at the end of the environment treatment (the day before the learning battery starts). According to past results, we anticipate that exposure to complex environments would promote an increase in exploratory behaviors (Light, Kolata, Hale, Grossman, & Matzel, 2008).

Open field

The Open Field is a square field (46 cm × 46 cm) with 13 cm high walls, constructed of white Plexiglas and located in a brightly lit room (300 lx). The field is divided into a 6 × 6 grid comprised of 7.65 cm2 quadrants, where 20 of the quadrants are next to the outer walls of the field (i.e., "wall" quadrants), and 16 quadrants are located in the center (i.e., "open" quadrants). Mice are placed in the center of the open field, and their behavior monitored for five minutes. Throughout this time, the animals' entries into walled and open quadrants are recorded. An entry is recorded whenever both front paws crossed the border of a quadrant. We recorded the time spent in unwalled (open)

quadrants of the field as well as the time spent in walled quadrants. This measure has been previously interpreted to reflect exploratory tendencies, as opposed to non-specific motor activity.

Elevated plus maze

The elevated plus maze was used to assess the exploratory behaviors of mice after their group treatments. The maze is constructed of grey Plexiglas in the shape of a "plus." Each arm of the maze is 28 cm long and 6 cm wide, and the maze is elevated 30 cm above a white floor. Two opposing arms of the maze are enclosed in 8 cm high, grey Plexiglas walls and two of the arms were open. The maze is located in a brightly lit room (300 lx). Animals were placed in the center of the maze facing a closed arm, and their behavior in the maze recorded for 3 minutes. We recorded the percent time in closed and open arms. Generally, open arms are considered to be stressful to animals, thus measures in the open arms provide indices of exploratory tendencies similar in nature to that of exploration of the open quadrants of the open field.

Environmental enrichment

For this study, we separated the mice into two groups: a control group and an environmental-enrichment group. Our enriched group was in a different colony room and with home cages containing a running wheel for exercise throughout the 16 days of treatment. These mice were exposed to one novel environment each day for 30 minutes outside their home cage in a large room under dim lights. There were a total of 16 novel environments administered at the same time each day.

The environments were: 1) A big, black, plastic box with two concave towers on each side and a platform in the center reachable by jumping. 2) A narrow Plexiglas tube where the ends have two small boxes where the mice would reach going through the tube. 3) An 8-arm Radial Arm Maze with all doors left open. 4) An acoustic chamber with foam on the wall with a fan inside as the only sound mice would experience. 5) A black box with white stripe on the walls and the floor covered with soft, plastic spikes. 6) A white box with six different plastic toys. 7) A social box where there is always a second mouse inside a cylindrical cage to interact with. 8) An open rat cage with ¼ filled with bedding and 15 marbles on the top, which mice are prone to manipulate and then hide. 9) A closed rat cage with standard level of bedding containing four pieces of paper towel to be shredded. 10) A white box with a fixed "merry-go-round" like structure inside. 11) A metal pot with holes on the sides for nose poking, and a cover closing the pot. 12) A big, white, plastic box with the two cylindrical beams crossed in X and reaching high for the mice to climb. 13) A closed mouse cage put upside down with 10 strings of rope crossing the top of it creating a net where mice could slowly walk on. 14) A white box with a big white PVC tube crossing it and a mirror at one of its ends. 15) An acoustic chamber with foam on the wall with a metal plate inside containing jars filled with small metal jingle bells to produce sound whenever the mice roll the jars. 16) A big, white, plastic box with a large platform going up and ending with a large metal grid, and chicken wire around the bottom.

As soon as the mice had completed their 16 days of environmental treatment, they were moved back to standard cages in the colony room with the control siblings. All mice

were then handled again for 90 seconds a day for seven days. This ensured that both groups were receiving similar contact with humans and with the laboratory outside the colony room before testing. Also, the seven days of break would function as "rest" for mice in the environmental group to get at metabolic levels closer to mice in the control group (so potential confounds such as withdrawal from intense physical activity wouldn't be as salient during testing for intelligence). After this, all mice were tested in the Elevated plus Maze, followed by testing in the learning battery to assess their intelligence.

Learning battery

Lashley Maze

In the Lashley maze, mice were placed in a start box that leads to four interconnected alleys and a "goal box," that contained a food reward. In order to familiarize the mice with the novel reward they were provided the specific food reward the day prior to acclimation. The Lashley III maze is scaled for mice, and parameters were developed that supported rapid acquisition. The maze is constructed of black Plexiglas. A 2-cm wide, 0.1 cm deep white cup was located in the rear portion of the goal box, and a piece of chocolate-flavored puffed rice serves as the reinforcer. Illumination was 80 Lux at the floor of the maze. The maze is isolated behind a shield of white Plexiglas to mitigate against extra-maze landmark cues. Food-deprived animals were acclimated and trained on 2 successive days. On the day before acclimation, all animals were provided with eight pieces of chocolate-flavored puffed rice in their home cages to familiarize them with the novel reinforcer. On the acclimation day, the mice were placed

in the four alleys of the maze, but the openings between the alleys were blocked so that the animals could not navigate the maze. Four minutes were spent in each alley, and the goal box contained 3 pieces of chocolate-flavored puffed rice. Once testing starts, the mice were placed in the start box and would have to find their way to the goal box during five consecutive trials. During each trial, the time it takes the mice to reach the goal box, and the errors made along the way were tracked. The errors are of two types: backtracking, which we define as the mice going from one alley opening to the prior alley opening, and dead end, which we define as the mice walking past an alley opening towards a dead end. If the mice already make a backtrack error in an alley, a dead end in that same alley is not considered an error. In between each trial, the mice were placed back in their home cage for 20 minutes, and the maze was cleaned for the next trial.

Passive Avoidance

A chamber illuminated by dim (5 fc) red light was used for training and testing. At the rear of a 16 x 12 cm (length x width) white grid floor is an enclosed platform (70 x 45 x 45 cm; length x width x height) constructed of black Plexiglas and closed on all sides except the side facing the grid floor. The platform floor is 5 cm above the grid floor, and a black Plexiglas sloping ramp extend 5 cm from the floor of the platform to the grid floor. The exit from the platform can be blocked by a remotely operated, clear Plexiglas sliding door. When an animal step from the platform and contact the grid floor, the compound aversive stimulus composed of a bright (550 Lux) white light, noise, and vibration is initiated. Noise and vibration are produced by a flexible nylon rod attached to a motor outside of an exterior wall of the chamber such that the rod strikes the wall of the chamber twice during each revolution (1400 rpm) of the motor, producing a noise 65 dBa

above a 45 dBa background and a 46 Hz vibration of the chamber surfaces. We placed mice on the platform and spent five minutes in confinement, where the door to the grid flooring was lowered. After this interval of time, the door was lifted so that the mice can step down. Once the mice step off the platform, the door is lowered (not allowing them back onto the platform at that time) and the combination of light, noise, and vibration was presented for four seconds. After the four second interval, the door was lifted to allow the mice to return to the platform, where they were confined again for five minutes. After the five-minute interval, the door was again lifted for the mice to step onto the grid floor and we recorded their latency. Those with better cognitive abilities should have a longer latency to step down than those with lower cognitive abilities.

T-Maze Alternation

The apparatus consists of one longer start arm (75 cm) and two shorter arms (31.5 cm), with two sliding opaque doors located 0.2 cm into the short arms. The arms are all 12-cm wide and 20 cm high. The floor of the maze is black plastic and the walls are clear plastic.

During the task, the mice were tested on 12 consecutive trials to choose the alternate arm to find the food reward. The day before acclimation, the mice were food deprived and introduced to 8 pieces of the food reward to be used in this task. On the day of acclimation, mice are given four forced choices. A mouse was held in the start area for 20s, and then allowed to pass through, with only the door leading to the right arm opened. After the food is eaten, we returned the mouse to the start area for a 20-s intertrial interval. We then repeated this procedure with the right door closed and the left door

opened instead. After the food is eaten, this sequence restarts for a total of four forced-choice trials.

On the day following acclimation, we administered the actual task. Mice chose between either open arm. On the first trial, food was available in both arms, and the animal was able to make one free choice. On the second trial, food was only available in the arm not chosen on the first trial. If an incorrect choice is made, we allowed the animal to correct its mistake and find the food in the other arm. After the correct choice is made, we placed the animal back in the start area and wait 20 s for the following trial, now with food only in the opposite choice arm. Accordingly, the mice have to alternate their choices to most efficiently obtain food. We administered two days of training with 12 trials per day. From our previous experience, this level of training is sufficient to support a high level of efficacy.

Odor Discrimination

A black Plexiglas 60-cm-square field with 30-cm-high walls is located in a dimly lit (10 fc) testing room. Three 4 x 4 x 2.0 cm (length, width, height) aluminum food cups are placed in three corners of the field. A food reinforcer (30 mg portions of chocolate-flavored puffed rice) is placed in a 1.6-cm-deep, 1-cm-diameter depression in the center of each cup. The food in two of the cups is covered (1.0 cm below the surface of the cup) with a wire mesh so that it is not accessible to the mice, whereas in the third cup (the "target" cup), the food can be retrieved and consumed. A cotton-tipped laboratory swab, located between the center and rear corner of each cup, extends vertically 3 cm from the surface of the cups. Immediately before each trial, fresh swabs were loaded with 25 µl of lemon, almond, or mint odorants (McCormick flavor extracts). The mint odor was always

associated with the target food cup. (It should be noted that in pilot studies, the odor associated with food was counterbalanced across animals, and no discernible differences in performance could be detected in response to the different odors.)

On the acclimation day, each food-deprived mouse was placed in the field for 10 min with no food cups present. Later, three pieces of chocolate-flavored puffed rice that subsequently served as the reinforcer was placed in the animals' home cages to acquaint them with the reinforcer.

On the subsequent test day, the mice received four training trials in the field with three food cups present. On each trial, the mouse was placed in the empty corner of the field. On trial 1, the reinforcing food (chocolate-flavored puffed rice) is available to the mice in the cup marked by mint odor. The trial continues until the mouse retrieved and consumed the food from the target cup, after which the mouse is then returned to its home cage to begin a 6 minute ITI. On trials 2–4, the location of the food cups was rearranged, but the baited cup remained consistently marked by the mint odor. Both the corner location of the mint odor and its position relative to the remaining odors changed on each trial. On each trial, we recorded the latency to retrieve the food and errors. An error is recorded any time that a mouse made contact with an incorrect cup or its nose crosses a plane parallel to the perimeter of an incorrect cup. Similarly, an error is recorded when a mouse sample (as above) the target cup but does not retrieve the available food.

Spatial Water Maze

In the spatial water maze, the mice were placed in a circular tub of opaque water and forced to find a hidden platform using extra maze cues, or "landmarks." A round white pool (140 cm diameter, 56 cm deep) is filled to within 20 cm of the top with water made opaque by the addition of nontoxic, water soluble, black paint. A hidden 12-cm-diameter perforated white platform is in a fixed location 1.5 cm below the surface of the water midway between the center and perimeter of the pool. The pool is enclosed within a ceiling-high black curtain on which lights and geometric shapes (landmarks) are positioned at heights (relative to water surface) ranging from 90 to 150 cm. A video camera lens extends through a 30-cm-diameter black circle 180 cm above the center of the water surface. The tub is divided into four quadrants, one of which includes the platform, making the starting point one of the other three quadrants. On the day before testing, the mice were confined to the platform for 300 seconds for acclimation. On the testing day, the mice were placed in unique locations during each of the six trials within the three quadrants (i.e., that do not include the platform). Once the mice have all four of their paws on the platform, they were left there for 5 seconds. If a mouse cannot find the platform after 90 seconds, we placed them on the platform and leave them there for 5 seconds, similar to when the mice finds the platform. We recorded path lengths from the start position to the platform during each trial as the measure of learning.

Brain weight

Two days after the end of behavioral tests, all mice were euthanized and decapitated. We then removed their whole brains following standard procedures and weighed each brain.

Statistical Analyses

For all exploration tasks as well as the tasks in the learning battery, we defined univariate outliers as any values above or below two interquartile ranges. We then applied the technique popularly 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. (The validity of this formula is shown in Hoaglin & Iglewicz, 1986. The advantages of this technique over the deletion of data are reviewd in DiLalla & Dollinger, 2006.) We also tested all data for the presence of kurtosis and skewness, and when necessary, did relevant transformations to fit the variables to a normal distribution. These pre-analyses were all done in SPSS 24. Regarding any potential missing data, we estimated values for each case by using Multiple Imputation (MI). MI is a technique for calculating maximum-likelihood estimates based on the observed data, and provides less biased information than simpler procedures for dealing with missing data such as listwise deletion, pairwise deletion or imputation of means (Schafer, 1999, 2010).

The mice's rate of acquisition was determined for each of the cognitive tests. Using an exploratory factor analysis, a statistical method that is used to describe variability among multiple correlated variables, we assessed the individual differences

between acquisitions on all the tasks. From this matrix, we estimated the extent to which the mice's performance in one test indicates their performance on the other tasks (Matzel et al., 2003). Each animal was then assigned a factor score, which represents their general cognitive performance, or intelligence score. (A factor score is analogous to an average z score of an animal's performance on each task, where the performance on each task is weighted according to the degree of loading on the first factor.)

To test our assumption that there is a single factor (general cognitive ability) explaining the common variance between learning tasks of the battery, we performed a confirmatory factor analysis. We used the maximum likelihood estimation in AMOS 24 to acquire the solution for the model. This particular estimation is considered robust in comparison to other procedures like Generalized Least Squares and Asymptotically Distribution-Free, and allows reliable fit indices with relatively small samples (Chou & Bentler, 1995). We assessed model fit by using two absolute indices – Model Chi-Square and Root Mean Square Error of Approximation (RMSEA) – that describe how the model represents the observed data, and for which lower values indicate better fit. For the Model Chi-Square (χ^2_M), the null hypothesis is the model itself, so failing to reject it (i.e., a small Model Chi-Square) indicates a good fit (with alpha here set at 0.05) (Kline, 2011). Similarly, RMSEA values of 0.06 and below are considered good (Hu & Bentler, 1999). In addition to these two absolute indices, we also assessed model fit with the incremental index called Comparative Fit Index (CFI). This describes how well the model fits in comparison to a baseline model in which all variables are uncorrelated and without latent variables, and for which higher values indicate better fit (Kline, 2011). CFI indicate an adequate model fit at values of 0.95 or above (Hu & Bentler, 1999). We chose these tests

due to their statistical relevance and frequent use (Hooper, Coughlan, & Mullen, 2008; Kline, 2011; Schreiber, Nora, Stage, Barlow, & King, 2006). For assessing the significance of individual parameters such as regression paths and correlations, we chose an alpha value of 0.05.

We estimated the heritabilities of the exploration tests (Open Field and Elevated plus Maze), brain weight, and each of the learning tasks of the battery for all mice. We also estimated the heritabilities of general cognitive ability scores from both groups combined as well as from each group separately. We followed the full-sibling formulas by Falconer to obtain full-sib heritability (h_{FS}) and its standard deviation (σh_{FS}):

$$h_{FS} = 2\sigma_F^2 / (\sigma_F^2 + \sigma_w^2)$$

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

Where

h_{FS} is the full-sibling heritability

σh_{FS} is the standard deviation of the full-sibling heritability

σ_F^2 is the difference between the siblings of different families

σ_w^2 is the difference between siblings within a family

n is the number of individuals per family

N is the number of families

t is the full-sibling intraclass correlation: ½ h_{FS}

We used one-way analyses of variance (ANOVA) with family as the main effect to test for significance of the heritabilities (Falconer 1989).

We used t-tests in SPSS 24 to compare mice in the Enrichment group to the Control group in all exploratory and learning tests, as well as brain weight. We also used ANOVAs and mixed model linear regressions to test for the effect of the environment enrichment in general cognitive ability scores (i.e., intelligence), as well as the random effect from sibling families. Finally, we tested for gene-environment interactions using a GLM procedure in SPSS 24.

Results

We first describe the results from each test in the chronological order of our study design –including the results for all mice and the differences between Control and Enrichment groups. We start by describing the results from the Open Field, then the results from each task of the learning battery, then amount of physical exercise, then results from the Elevated plus Maze, and finally brain weights. (Brain weights were obtained for two purposes. First, the brain weight of mammals is known to be heritable. Thus the heritability of brain weight can be expected to be heritable and can serve as a validation of our statistical methods. Second, it is conceivable that environmental enrichment will promote changes in brain weight and might interact with heritability, a possibility that can be easily assessed.) After that, we describe the descriptive statistics of all relevant measures used in further analyses, and show their correlations. With this, we then explore the results of our factor analyses, which revealed the latent variable of general cognitive ability. Finally, we show estimates of heritability, environmental effects, and GE interactions in the model of confirmatory factor analyses.

Exploration before group treatments: Open Field

Results for the Open Field test can be seen in Figure 1. We used the measure of relative time spent in the center (the unwalled area), and found that mice in both groups (Enrichment and Control) had similar levels of exploration before the environmental enrichment was begun, $t(195) = 0.66$, $p = 0.510$. This suggests that mice in both groups had similar levels of anxiety/stress reactivity before the start of the start of any manipulation.

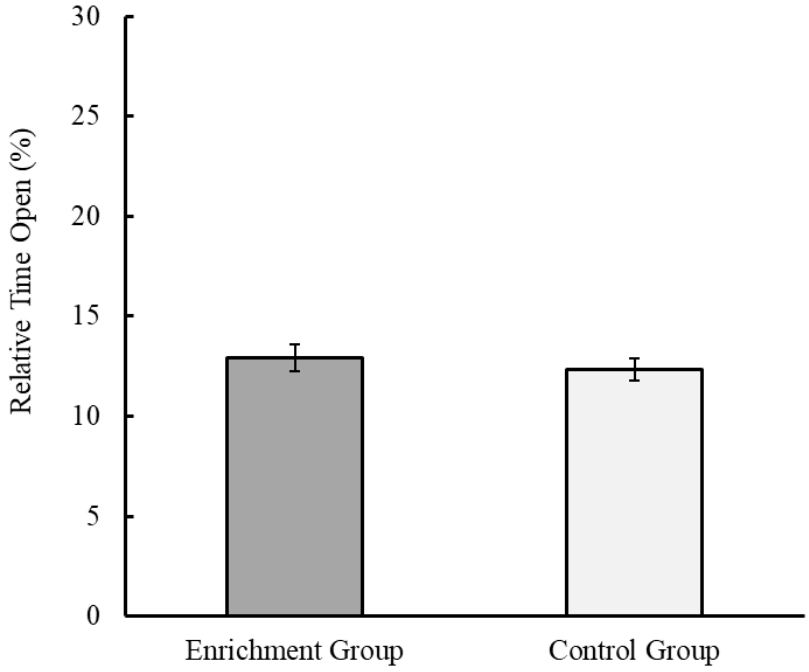


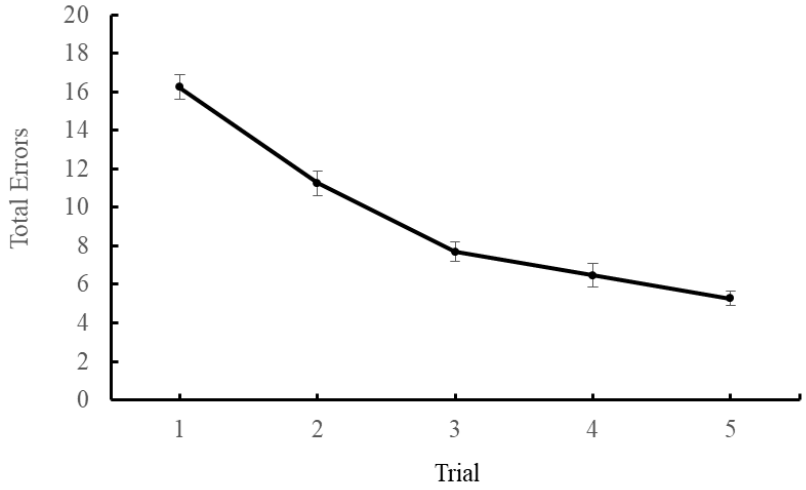
Figure 1. Mean relative time spent in the open quadrants (%) of the Open Field for mice in the Enrichment group and the Control group. Note that this data was obtained before differential treatment between groups. There was no significant difference in exploration between the two groups. Brackets indicate standard error of the mean.

Individual tests of the learning battery

Lashley Maze

In the Lashley Maze, mice started the test making 16.25 (±0.64) errors on average, and ended with 5.27 (±0.36) errors (Figure 2A). This reduction of errors across trials was significant, $F(4,824) = 78.79$, $p < 0.001$, which indicates that the mice learned the task. There was no difference between groups across the five trials, $F(1,206) = 2.19$, $p = 0.141$. However, when we consider the trials where the learning phase was most pronounced (Trials 2 and 3), there was a difference between groups, $F(1,206) = 4.93$, $p = 0.027$, i.e., environmental enrichment promoted an increase in the rate of learning (see Figure 2B).

A



B

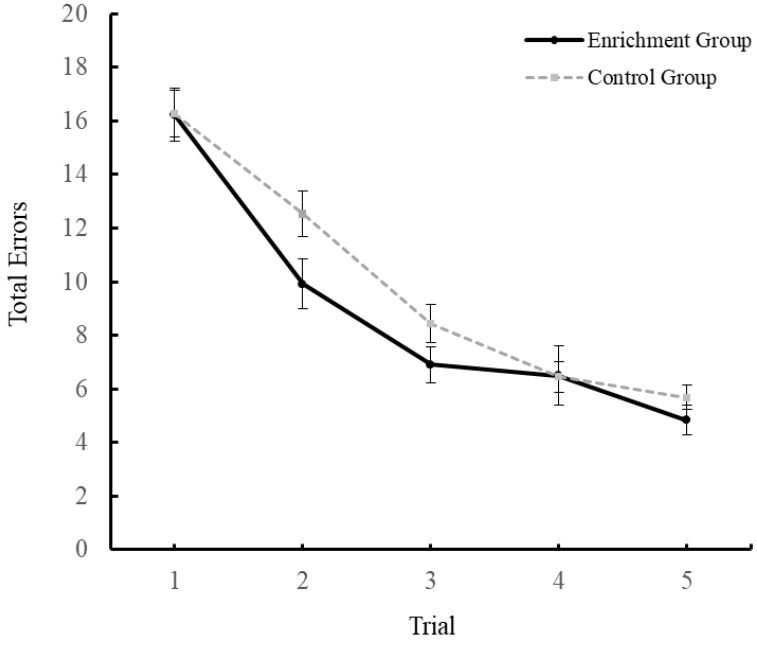
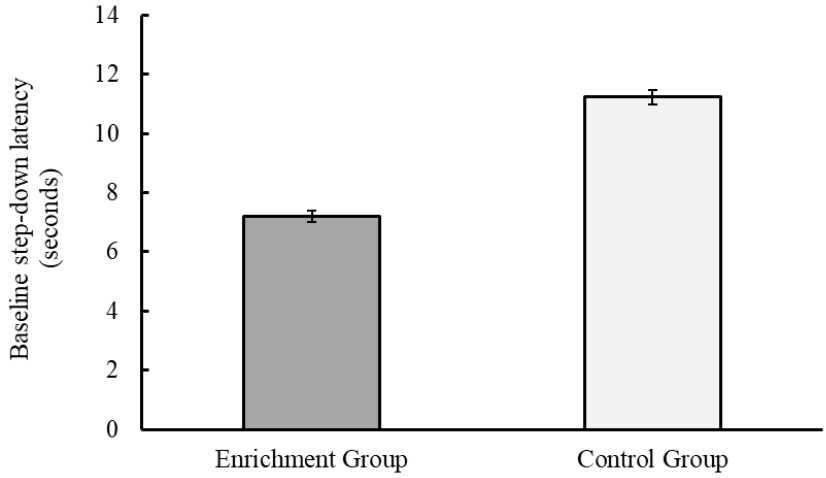


Figure 2 A. Mean number of errors in the Lashley Maze for all mice. B. Mean number of errors in the Lashley Maze for the mice in the Enrichment group and the Control group. There was a significant difference in rate of learning between the two groups. Brackets indicate standard error of the mean.

Passive Avoidance

For the passive avoidance task, there was a significant difference between the groups for baseline step-down latencies (Figure 3A), which is the latency to step into a novel environment before pairing that step with aversive light and noise, $t(225) = 12.93$, $p < 0.001$. This latency is obtained *before* exposure to the aversive stimuli, and thus is indicative of exploration, where lower latencies represent more exploration. Thus it appears that environmental enrichment promoted increases in exploration independent of learning. After the aversive light and noise were encountered (pursuant to stepping off the platform), the mean latency for all mice went from 9.25 (±0.21) seconds to 16.68 (±0.51), a significant increase, $F(1,225) = 209.59$, $p < 0.001$, that indicates the mice learned the avoidance response. Given the difference in baseline performance, we assessed learning in the passive avoidance by computing a ratio of the post-training step-down latency divided by the baseline step-down latency, where higher ratios would reflect better learning. Between groups, a comparison of their learning ratio showed significant differences, $t(225) = 6.36$, $p < 0.001$ (Fig. 3B), indicating that environmental enrichment improved the rate of avoidance learning.

A



B

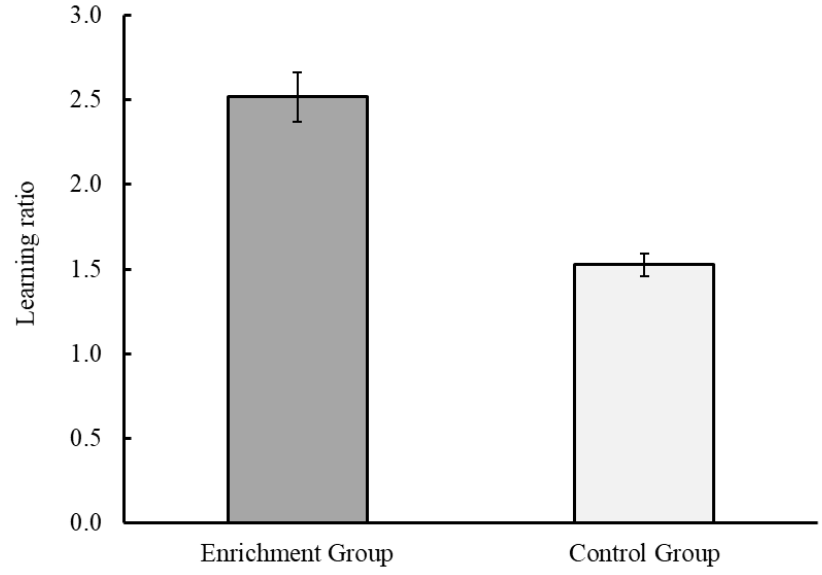


Figure 3 A. Mean latency (seconds) to step down from the platform in the Passive Avoidance task before pairing that step with aversive light and noise for mice in the Enrichment group and the Control group. There was a significant difference in exploration between the two groups. B. Mean latency (seconds) to step down from the platform in the Passive Avoidance after pairing that step with aversive light and noise for mice in the Enrichment group and the Control group. There was a significant difference in learning between the two groups. Brackets indicate standard error of the mean.

T-Maze

There are only two possible choices every trial, so random choices would yield a 50% correct response rate. A binomial test indicated that, at the end of the first training day (12th trial), the proportion of correct choices (0.64) was already better than expected by chance (0.50), $p < 0.001$. For the measure of learning we used mice's performance based on the previously defined learning criterion of 4 correct choices in a row. As seen in Figure 4, mice in the Enrichment group did not differ significantly from the Control group, $t(222) = 0.20$, $p = 0.840$, indicating that on this task, environmental enrichment did not promote improvements in learning.

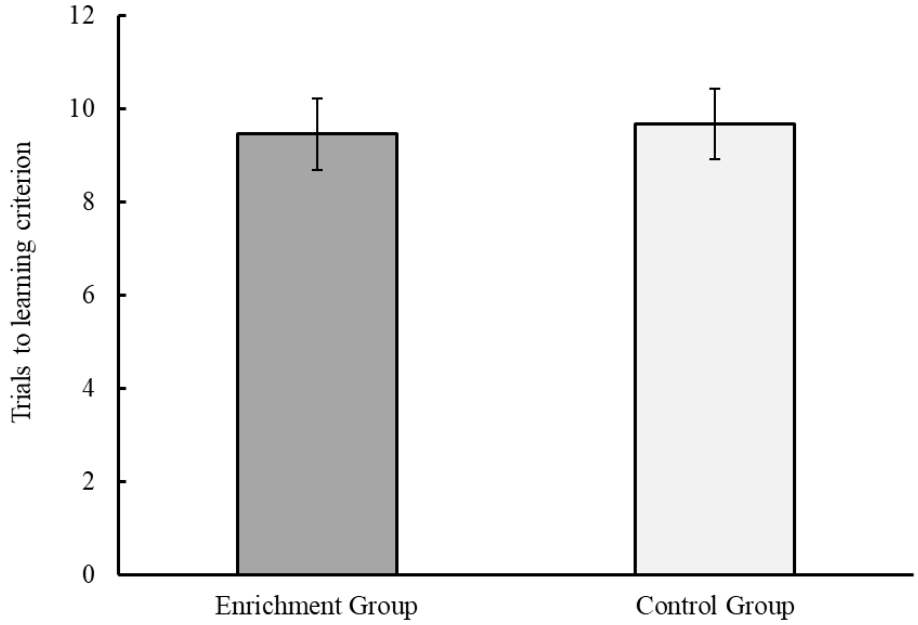
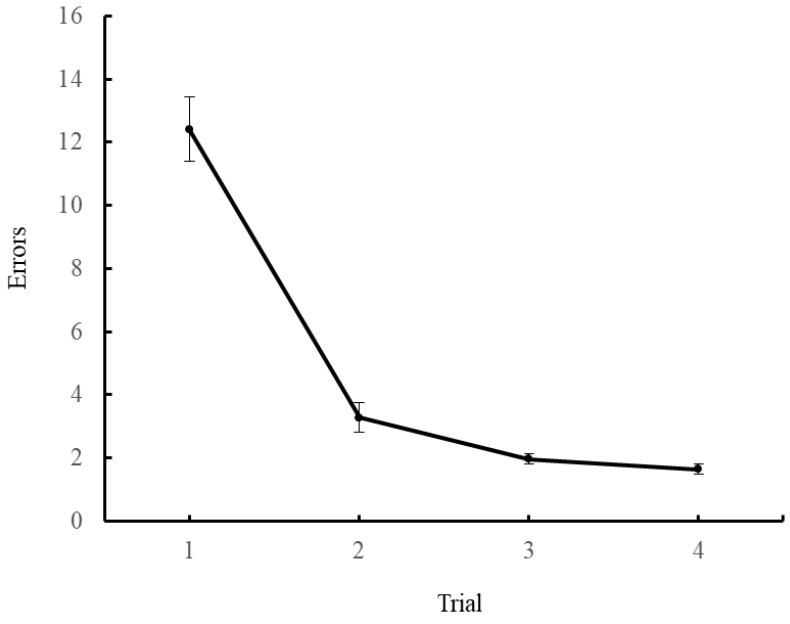


Figure 4. Mean number of trails for the learning criterion of 4 consecutive correct choices in the T-Maze for mice in the Enrichment group and the Control group. There were no significant differences in learning between the two groups. Brackets indicate standard error of the mean.

Odor Discrimination

In the Odor Discrimination task, mice started the test (Trial 1) making 12.40 (±0.80) errors on average, and ended (by Trial 4) with 1.64 (±0.14) errors (Figure 5A). This reduction of errors across trials was significant, $F(3,663) = 148.34$, $p < 0.001$, indicating that the mice learned the task. There was also a difference between groups across the four trials, $F(1,221) = 5.34$, $p = 0.022$ (Figure 5B), indicating that environmental enrichment improved learning.

A



B

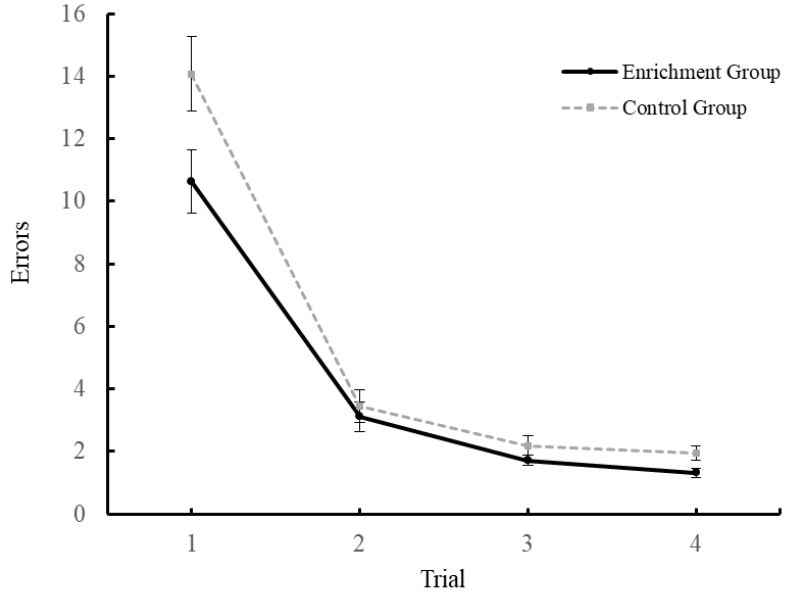
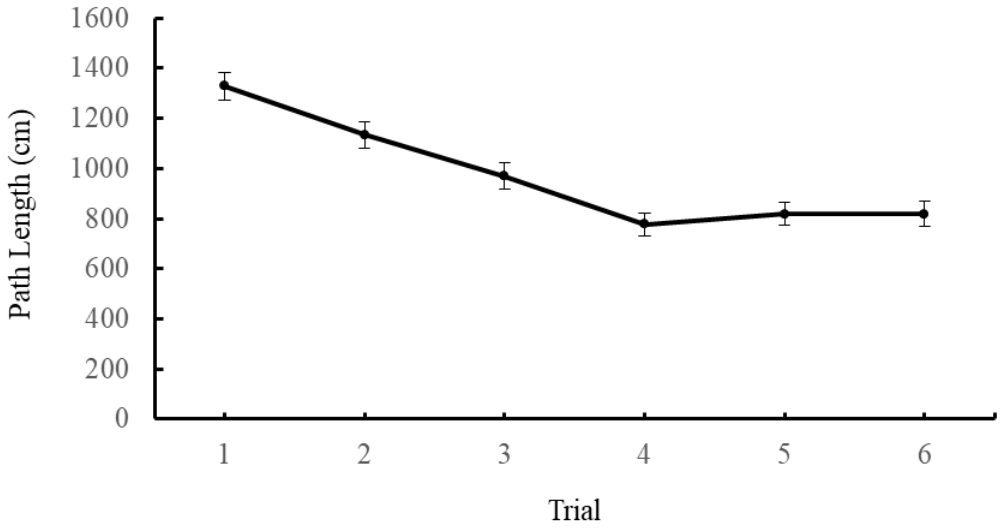


Figure 5 A. Mean number of errors in the Odor Discrimination task for all mice. B. Average number of errors in the Odor Discrimination task for the mice in the Enrichment group and the Control group. There was a significant difference in rate of learning across the four trials between the two groups. Brackets indicate standard error of the mean.

Spatial Water Maze

In the Spatial Water Maze, the mean path length to reach the hidden platform decreased from 1328.35 (±55.29) cm at the first trial to 818.35 (±51.06) cm on the last trial (Figure 6A). This reduction in path length across trials was significant, $F(5,1020) = 20.22$, $p < 0.001$, indicating that the mice learned the task. There was no difference between groups across the six trials, $F(1,204) = 1.39$, $p = 0.239$ (Figure 6B). And even when we considered only the trials where the learning phase was most pronounced (Trials 1 and 2), there was still no significant difference between groups, $F(1,216) = 2.92$, $p = 0.089$, suggesting that on this task, environmental enrichment did not promote an increase in the rate of learning.

A



B

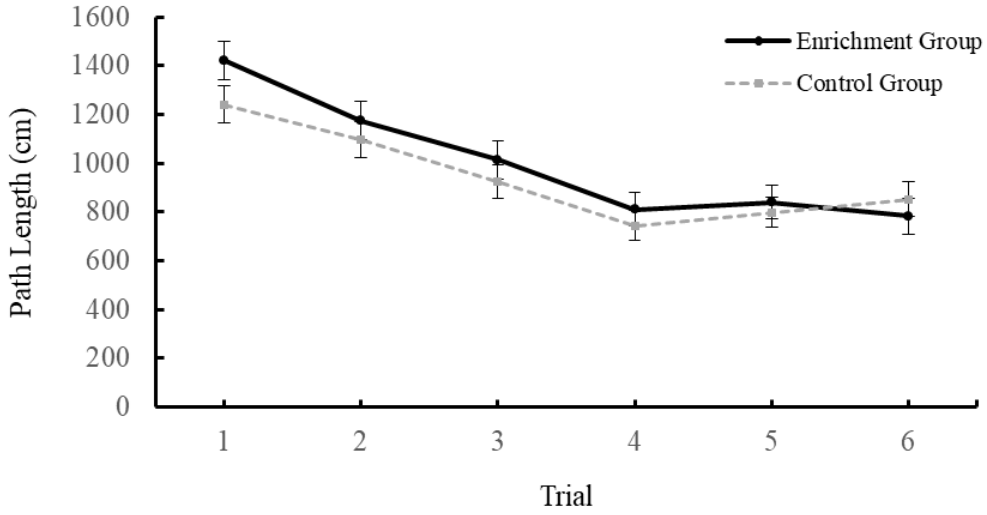


Figure 6 A. Mean distance traveled (path length, in cm) in the Water Maze for all mice. B. Mean distance traveled (path length, in cm) in the Water Maze for the mice in the Enrichment group and the Control group. There were no significant differences in rate of learning or asymptotic performance between the two groups. Brackets indicate standard error of the mean.

Exercise in the Enrichment Group

The environmental enrichment manipulation included providing access to running wheels in their home cages. We counted the revolutions of the running wheel for all mice in the Enrichment group for a total of 16 days. These measurements were taken at roughly the same time every day in order to give a measure of the number of revolutions per 24-h period. Throughout this period, animals exhibited a steady increase in the amount that they ran each day, with a significant effect of changes in revolutions across trials, $F(15, 1425) = 47.25$, $p < 001$. The mean revolution per day over the duration of the treatment was 15,774.04 (±686.46), which translates to each mouse running 6.3 Km every day.

Exploration after group treatments: Elevated plus Maze

Results for the Elevated plus Maze (obtained after differential treatment of the two groups) can be seen in Figure 7. The pattern for the Elevated plus Maze results contrast with results in the Open Field (obtained before differential treatment of the two groups). We used the measure of relative time spent in the open arms, and found that mice in the Enrichment group explored more than mice in the Control group after receiving the environment enrichment, $t(222) = 3.06$, $p = 0.002$. This result suggests that mice in the Enrichment group had lower levels of anxiety/stress reactivity compared to Control animals.

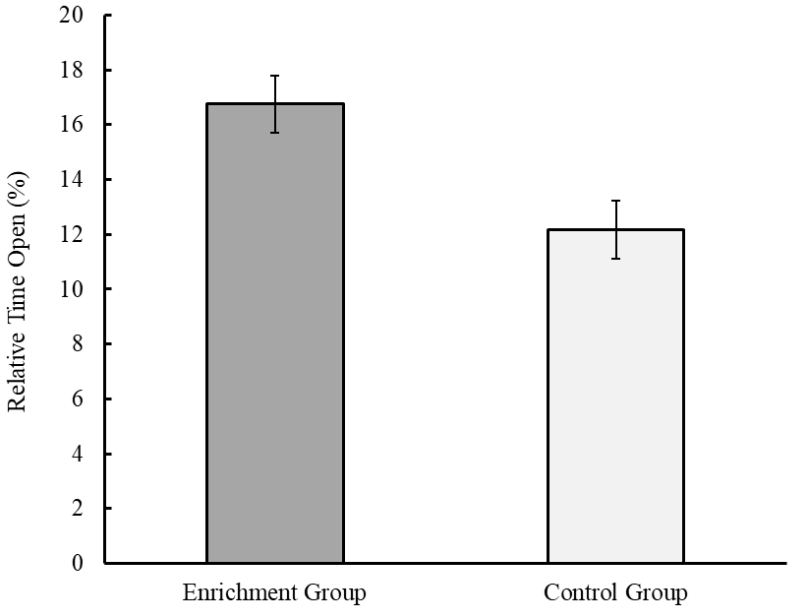


Figure 7. Mean relative time spent on the open arms (%) of the Elevated plus Maze for mice in the Enrichment group and the Control group. Note that this data was obtained after differential treatment between groups. There was a significant difference in exploration between the two groups. Brackets indicate standard error of the mean.

Brain weight

At the end of the study, we euthanized all mice and weighed their brains. Results can be seen in Figure 8. There was no difference between Enrichment group and the Control group, $t(223) = 0.50$, $p = 0.615$.

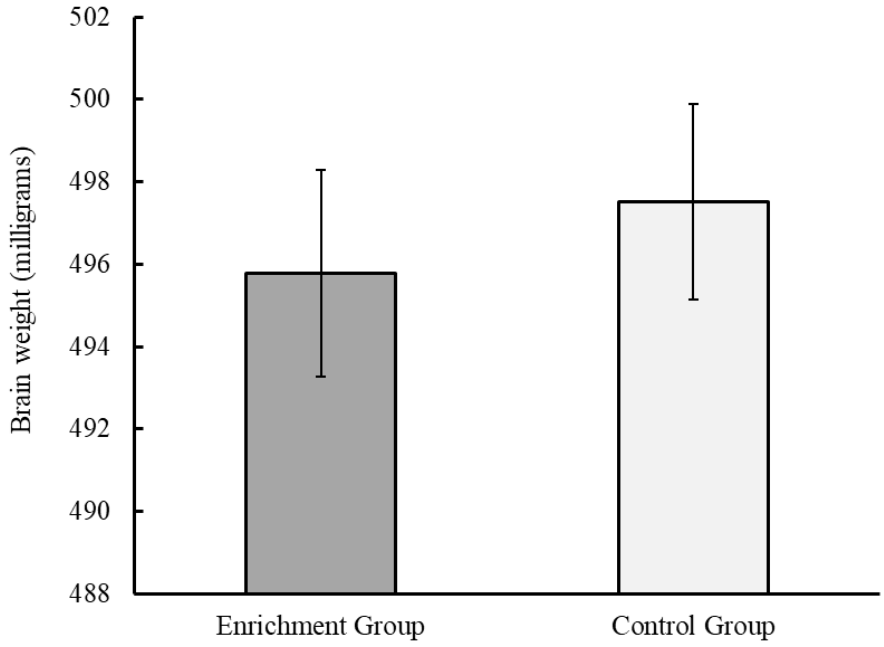


Figure 8. Mean brain weight (milligrams) for mice in the Enrichment group and the Control group. There were no significant differences in brain weight between the two groups. Brackets indicate standard error of the mean.

Descriptive Statistics of all relevant measures

As noted above, we chose only specific trials from the learning test to represent the learning in that test (thus providing a sensitive measure of differences in *rate* of learning). For the Lashley Maze, we used the mean performance in trials 1, 2, and 3 as the greatest gains in performance occurred across those trials. For Passive Avoidance, we used the ratio of the post-training step-down latency divided by baseline step-down latency. For the T-Maze, we used the number of trials required to reach a criterion of four correct choices in a row. For the Odor Discrimination, we used trials 1, 2, and 3 (again, where the greatest gains in performance occurred). Finally, for the Spatial Water Maze we used trials 1 and 2.

The means and standard deviations for all variables used in further analyses are shown in Table 1, while inter-correlations between these variables are shown in Table 2. By examining all variables for the presence of univariate outliers (defined in the Methods sections), we found up to eight cases of outliers in each of the variables for Lashley Maze, Passive Avoidance, T-Maze, and Odor Discrimination. We did not find any outliers for Spatial Water Maze. Instead of deleting outliers from the analyses, however, we "brought them to the fence" by setting their values to two interquartile ranges either above the 3rd quartile or below the 1st quartile – depending on the value.

Because factor analyses, ANOVAs, and multiple regressions can be sensitive to non-normal distributions (just like any parametric analysis), we tested all variables for skewness and kurtosis. Fortunately, all variables had values of skewness and kurtosis well within the recommend range for normality (-2 and +2). Therefore, there was no need for transformations.

Less than 8% of the whole sample was missing, and that data were missing at random, Little's MCAR test: Chi-Square = 20.512, DF = 19, Sig. = .364. As described in Methods, we estimated values for each missing case by using Multiple Imputation (MI).

Table 1. Means and standard deviations of all variables used in the study.

Category	Measured variable	Mean	Standard Deviation
Exploration	Open Field (% of relative time in the center)	12.42	6.21
	Elevated plus Maze (% of relative time in open arms)	14.32	11.30
Brain Weight	Brain weight (milligrams)	496.67	25.90
Learning Tasks	Lashley Maze (Trials 1, 2, and 3. Errors)	11.73	4.78
	Passive Avoidance (learning ratio)	1.92	0.96
	T-Maze (Trials to learning criterion of 4 correct choices)	9.61	8.04
	Odor Discrimination (Trials 1, 2, and 3. Errors)	5.73	4.57
	Spatial Water Maze (Trials 1 and 2. Path length in cm)	1214.91	606.15

Table 2. Pearson correlations between all variables used in the study.

	1	2	3	4	5	6	7	8
1. Open Field	—							
2. Elevated plus Maze	.14*	—						
3. Brain weight	.06	.03	—					
4. Lashley Maze	-.08	-.18**	-.04	—				
5. Passive Avoidance	-.01	-.11	.10	.27**	—			
6. T-Maze	.03	-.06	.03	.22**	.02	—		
7. Odor Discrimination	-.02	.04	-.06	.14*	.08	-.02	—	
8. Spatial Water Maze	-.01	.01	-.03	.13*	-.10	.04	-.04	—

* $p < 0.05$

** $p < 0.01$

Factor Analyses of General Cognitive Ability

An unrotated principal components factor analysis of the performance data for the five learning tasks that comprise the learning battery isolated a factor that accounted for a total of 19.5% of the variance in performance (eigenvalue = 1.00, n = 232; Table 3). This is equivalent to accounting for 28% of the total variance from a Principal Component Analysis, which is 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 then to extract factor scores to represent animals' aggregate performance across all learning tasks (i.e., their general cognitive ability).

Table 3. Factor loadings and variance explained by the first factor (General Cognitive Ability) extracted from the five learning tasks using an exploratory factor analysis. n = 232.

Learning Task	General Cognitive Ability
Lashley Maze	.89
Passive Avoidance	.33
T-Maze	.22
Odor Discrimination	.15
Spatial Water Maze	.11
Eigenvalue	1.00
Proportion of common variance	19.5%

We performed a confirmatory factor analysis to ensure that the measured variables of learning would form a coherent latent variable. In other words, it tests how the variables are related in the absence of any hypothesized relationship between them, while still capturing only the variance shared in common between the variables (as opposed to techniques such as principal component analyses that capture both shared and non-shared variance, and are better suited for purposes of dimension reduction). The model from the confirmatory factory analysis (Figure 9) had a good fit to the data ($\chi 2$ = 7.04, df = 6, p = 0.317; RMSEA = 0.027; CFI = 0.964), with all measured variables having significant factor loadings ($p < 0.05$).

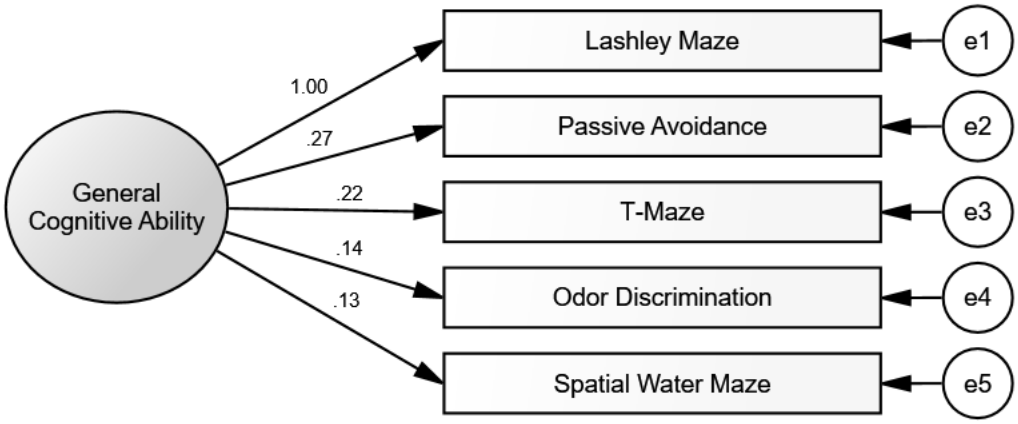


Figure 9. Confirmatory factor analysis of the model with a single latent factor (general cognitive ability) influencing all observed learning tasks in our study. The model overall had a good fit to the data. Small circles with an "e" represent error (i.e., leftover variance due to flaws in measurements and estimations). Arrows going out of the latent factors (big circle) represent factor loadings. All parameters shown are standardized.

Heritabilities of tests and general cognitive ability

Table 4 gives the calculated heritabilities for learning and exploration tasks, brain weight, as well as for general learning ability from the factors scores derived from the exploratory factor analysis described above. The estimates for general cognitive ability are given for all mice combined and for the Enrichment and Control groups separated. ANOVA indicated significant family effects for all variables ($p < 0.05$). Full-sib heritabilities were moderate for exploration tasks, as well as for brain weight. For the learning tasks, only the Lashley Maze and Odor Discrimination had significant heritabilities, and they were both moderate-low. Regarding GCA scores, the population of all mice combined showed a moderate-low heritability. In contrast, the population of only Enrichment group had a heritability not significantly different from zero, while the population of only Control group had a moderate heritability.

Table 4. Estimates of full-sib heritabilities with their respective standard deviations (SD) and p values. The heritabilities for General Cognitive Ability (GCA) are shown for all mice combined and for each group (Enrichment and Control) separately. Significant heritabilities are marked with *.

Variable	Heritability	SD	p value
Open Field	**0.40***	0.17	0.0003
Elevated plus Maze	**0.29***	0.16	0.0057
Brain weight	**0.42***	0.18	0.0002
Lashley Maze	**0.27***	0.15	0.0086
Passive Avoidance	**0.15**	0.13	0.0813
T-Maze	**0.08**	0.12	0.2300
Odor Discrimination	**0.20***	0.14	0.0400
Spatial Water Maze	**0.18**	0.14	0.0523
GCA score – All mice	**0.24***	0.15	0.0170
GCA score – Enrichment group	**0.15**	0.29	0.2844
GCA score – Control group	**0.55***	0.34	0.0168

Environmental and G×E effects on General Cognitive Ability

The mean general cognitive ability score in the Control group was -0.195 while the Enrichment group had a mean score of 0.195, an average increase in 0.44 standard deviations (Figure 10). This result suggests that experience with the enriched environment had a positive influence on animals' overall cognitive performance.

The mixed model revealed a main of effect of group on general cognitive ability scores, $F(1,57) = 12.67$, $p = 0.001$. The partial eta squared was 0.18, which is considered a medium effect size. There was also a random effect from family, $F(57,57) = 1.55$, $p < 0.05$, which means that genetic differences between each family of siblings mattered to GCA scores. The partial eta squared was 0.61, which is considered an extremely large effect size. However, there was no interaction between the genetic differences and the environmental manipulation, $F(57,116) = 1.12$, $p < 0.319$. This lack of effect suggests no gene-environment interaction behind the variation in GCA in our mice.

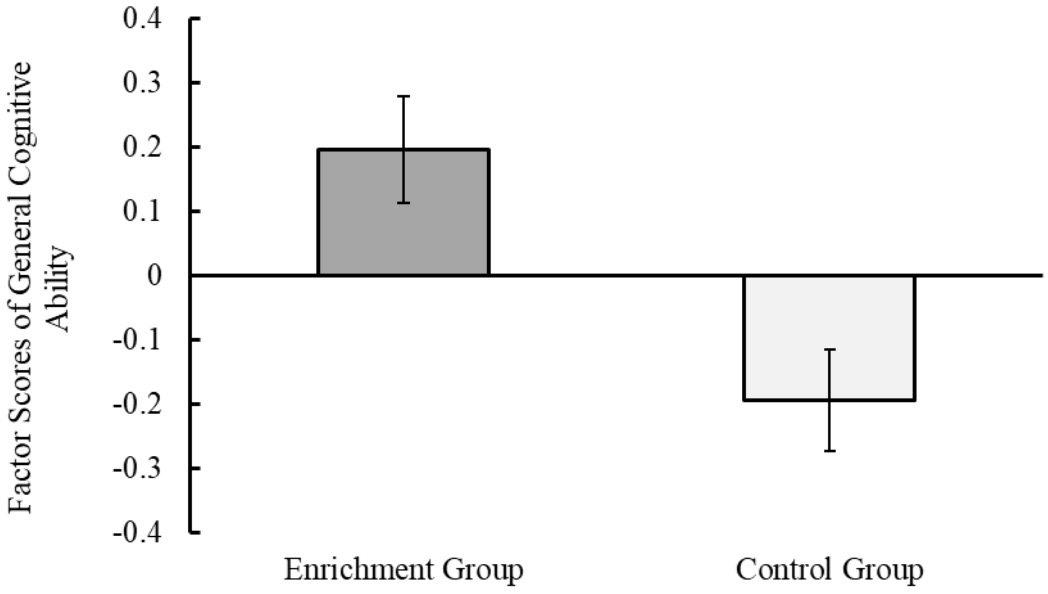


Figure 10. Mean factor scores of General Cognitive Ability for mice in the Enrichment group and the Control group. Factor scores were derived from the first factor of an exploratory factor analysis of the learning battery. There was a significant difference in General Cognitive Ability between the two groups. Brackets indicate standard error of the mean.

Discussion

Here we found that general cognitive ability (GCA) in mice has low-moderate heritability and is moderately malleable, where a group receiving enriched environment and physical exercise expressed substantially better performance than a control group. 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 heritable and malleable. We also found that heritability itself changed between groups, a result similar to what is seen in humans across SES (though in our case the change was in the opposite direction from what these studies of humans might lead us to expect). Unexpectedly, however, we did not find evidence for G×E in our study. A closer look at the results can provide a richer picture of our findings, and we discuss this below.

Similar to our previous work, there was a positive correlation of each mouse's rate of acquisition across all learning tasks. A GCA (i.e., intelligence) 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. That result is similar to what we have found in the past, where GCA explained 32% to 48% of the total variance across tests in a learning battery (for review, see Matzel, Sauce, & Wass, 2013). In addition, the 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. Those results suggest that the empirical measurement of the latent factor (GCA) is consistent with the exploratory factor analysis, and is comparable to results from previous studies.

Although expected based on previous work, it is notable that performance in tasks as distinct as the Lashley Maze and the Passive Avoidance share common variance. (In other words, a portion of the individual differences observed in those measures have the same cause.) One is a task that requires egocentric navigation and short-term memory, and uses food as motivation, while the other task requires associative learning between CS and US, uses fear as motivation, and passivity is the response requirement. The other tasks in the learning battery are also, by design, quite distinct. Even more notable is the fact that a single factor could explain about 20% of all the common variance in the battery here. In human studies, a single factor (the "g" factor) can explain around 50% of the common variance in performance in modern IQ batteries. At first glance, this value seems high relative to the present results. However, 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 (Mackintosh, 2011). Tasks that had poor correlations with other tasks were changed or removed over time. The mouse learning battery we used here has very little of that bias, and so it is remarkable that it captures a general cognitive factor with relatively "raw" tasks.

It is worth noting that while we assessed intelligence in mice by using a battery of learning tasks, modern IQ batteries in humans barely test learning capacity at all. That absence, however, might be more due to pragmatism than theory. In practice, testing for learning takes more time and resources in humans. In theory, learning ability is at the core of the definition of intelligence given by experts in the field (Neisser et al., 1996), and seen above in the Introduction section. General intelligence and learning seem to be so highly related that some claim the distinction between them may be more semantic

than real. Based on an extensive analysis, Arthur Jensen concluded (Jensen, 1998, p. 226):

> *"A general factor common to all learning tasks [the RATE at which they are acquired]…is highly correlated with the g factor extracted from psychometric tests. The general factor of both domain-learning and psychometric abilities-is essentially one and the same g."*

In addition, we also have direct evidence that at least in mice, general cognitive ability and general intelligence are probably synonymous, as presented in the Introduction section. In contrast to the mouse heritability study by Galsworthy et al. (2005), the batteries of tests that we used to estimate mouse intelligence have been designed to more closely emulate the rationale that underlies tests of human intelligence. Our tasks in this test battery isolate basic learning skills that are presumed to underlie a broad range of more complex forms of learning, as well as reasoning, working memory, and attention. Therefore, we can expect less contamination of estimates of cognitive ability from non-cognitive influences.

Even showing moderate heritability values (as discussed below), we found that mouse exploratory tendencies as well as GCA are fairly malleable via independent environmental effect (statistical main effects). Again, part of our study can be loosely described as an "adoption study", where 2 of the siblings in a litter of 4 were removed from their usual environment and experienced a new, more complex environment, a treatment that was administered for 16 days. At first glance, this might seem a short period of time, but keep in mind that mice's typical life span is less than two years, and a substantial part of their development occurs during the first 6 weeks of life. At birth, mice are hairless, blind, deaf, have minimal motor skills and are fully dependent on their

mother for nutrition and thermoregulation, while by the 6th week mice are already fully functional and approaching adolescence, with fine motor skills, showing complex social behavior, and a remarkable capacity for learning (Weber & Olsson, 2008). In that context, 16 days of environmental enrichment during pre-maturity phase is probably quite meaningful.

Indicative of the effectiveness of environmental enrichment, there was a substantial effect of that treatment on exploration. Prior to the treatment, both groups showed similar levels of exploration in the Open Field. After the treatment, however, they differed by a substantial margin, with mice who got the environmental enrichment showing much higher level of exploration in the Elevated plus Maze. In addition, we were also able to assess mice's exploration using the first part of the Passive Avoidance procedure. A prolonged baseline step-down latency has previously been interpreted as indicative of a reduction in exploratory behavior (Matzel et al., 2006). Given that Passive Avoidance was only the second task in the learning battery, these results also suggests that the increase in exploration is due to the enrichment itself, and not from a potential interaction between enrichment and the learning battery.

Similar to our results, a study with environment enrichment in rats (Franks, Champagne, & Higgins, 2013) found that the enrichment group showed a significantly greater exploratory tendency despite having equivalent levels of motor activity. Interestingly, they also showed that rats with (vs. without) enriched housing conditions are more likely to abandon known rewards and incur risk in order to explore, indicating that they take exploration to be highly valuable in its own right (Franks et al., 2013).

It is worth taking a detour to mention the relationship between performance in exploration tasks and in the learning battery. It is possible that increased exploratory tendencies might have played a role into the mice's capacity to learn. However, "general cognitive ability" is not simply "general exploratory behavior", and increased exploratory behavior by itself does not causally impact an animal's general cognitive ability. Anxiolytic administration, for example, promotes increases in exploration but generally causes a mild impairment in cognitive performance. In addition, a past study by our group increasing exploration in mice did not have a commensurate effect on general cognitive performance (Light et al., 2008). Those suggest that even if perhaps necessary, increases in exploration are not sufficient to promote increases in intelligence.

Regarding individual learning tasks, the Enrichment group had improved performance in the Lashley Maze, Passive Avoidance, and Odor Discrimination. So, although the increase was not observed consistently across all tasks, it was observed in tasks different enough from one another to assume the change was in a latent psychological factor; i.e., general cognitive ability. Indeed, mice from the Enrichment group were substantially higher in GCA scores than mice from the Control group. Assuming that the groups had similar intelligence before the start of the study (an assumption supported by the group similarities in the Open Field, which can be predictive of general cognitive performance), then the environment enrichment increased the intelligence by 0.44 standard deviations. In humans, this increase would represent an increase in 6.6 IQ points. Although here we used a combination of environmental factors (novelty and physical exercise) that were likely to be effective, these were admittedly only a limited in scope and strength. A complete change in the environment, such as seen

in adoption from poor countries to developed countries, probably has many more favorable factors in play. Also, given that brain weight differences in our mice did not correlate with GCA, it's not surprising that there was no effect from the environmental enrichment on brain weight.

Numerous cognitive theories about how environmental enrichment affects the brain have been proposed. Among them, the "learning and memory" hypothesis seems to be favored – it claims that when animals are confronted with novelty and environmental complexity, there are physiological and morphological changes in the mechanisms underlying learning (Henriette van Praag et al., 2000). The exposure to novelty might have an even bigger effect if paired with physical exercise, as we believe it did here in our study. However, caution is need with this conclusion, as the cognitive benefits are still widely debated.

There have been positive reports on the effects of exercise in a variety of tasks of learning and executive function (for a review, see Daugherty et al., 2017). These cognitive effects are plausibly conferred by changes throughout the brain, including synapatogenesis and neurogenesis (Erickson, Hillman, & Kramer, 2015; Gomez-Pinilla & Hillman, 2013). In rodents, for example, voluntary running-wheel access resulted in shorter path lengths in the Spatial Water Maze and increased adult hippocampal neurogenesis in mice (H. van Praag, 2005) as well as in rats (Nokia et al., 2016). Despite the positive evidence, broad conclusions on the effects of physical exercise have been complicated by variations in the duration of the treatment (chronic versus acute), the nature of the treatment (anaerobic versus aerobic), and in particular, the transient nature of the beneficial effects of exercise. Besides, and more relevant to our current study,

physical exercise alone does not seem to bolster general intelligence (Daugherty et al., 2017). Therefore, if physical exercise had a substantial influence on the malleability in GCA that we found, it is likely to have played a support role in conjunction with the exposure to novel and complex environments.

A final note on physical exercise: Similar to results we obtained in a previous study, the voluntary physical exercise in the running wheels increased with time. According to our results, the exercise treatment mice received was "adaptive" (or motivation driven) in the sense that it increased as the animals were becoming more fit. Consequently, it is likely that the mice were maintaining peak physical effort. It is plausible that an adaptive schedule of physical exercise has a superior effect on cognition than the non-adaptive schedules (e.g., one hour of access per day) used in many studies.

Of course we cannot say with certainty that our effects are specifically attributable to environmental enrichment as opposed to detrimental effects of environmental deprivation in the Control group. The control conditions in a lab cage are quite sterile, and might have been detrimental to intelligence as compared to normal/wild conditions of development. In humans, for example, a meta-analysis by van Ijzendoorn et al. (2008) compared the intellectual development of children living in orphanages with that of children living with their adoptive families (covering studies 19 different countries). On average, children growing up in orphanages had an IQ that was 16.5 points lower than their peers who were adopted. As expected, orphanages in countries with a higher Human Development Index (a combined measure of life expectancy, literacy, education, standards of living, and quality of life) showed a smaller detrimental effect in children's intelligence (reduction of 11.9 IQ points) than countries with a lower Index (reduction of

21 IQ points). Also, children in orphanages with the most favorable caregiver-child ratio (maximally three children per caregiver) did not significantly differ from their adopted peers. These observations suggest that our Control group might have intelligence lower than a normally developing mouse, and so the gap between Enrichment group and Control group would perhaps be both due to environmental enrichment as well as environment impoverishment.

Full-sib heritabilities were moderate-low for exploration tasks, similar to estimates found elsewhere – heritability of exploration in the Open Field, for example, was found by DeFries & Hegmann (1970) to range between 0.24 and 0.58. For the learning tasks of our battery, only the Lashley Maze and Odor Discrimination had significant heritabilities, and they were both moderate-low. Using a full-sibling design, Oliverio, Castellano, & Messeri (1972) found a similar moderate heritability in performance in the Lashley Maze in mice (about 0.40), while estimates for avoidance learning were considerably higher (though their measures were of active avoidance, and not passive avoidance like in our battery).

Even though differences in brain weight were not related to differences in learning or GCA, the heritability of brain weight was moderate to high. In humans, whole brain volume is highly heritable, as are measures of volume for many cortical regions (Croston, Branch, Kozlovsky, Dukas, & Pravosudov, 2015). To date, there is relatively little information available on the heritability of brain volume in nonhuman animals, but the data available (mostly in primates) shows quite high heritabilities as well (see references in Croston et al., 2015).

Regarding GCA scores, the population of all mice combined showed moderate-low heritability. When considering all mice as part of one population, our estimates of heritability are quite comparable to values obtained by Galsworthy (Galsworthy et al., 2005). However, it is important to note that SES impacts the estimates of heritability in humans. Given that low SES populations studied so far (and in wealthy countries) have low heritabilities (around 0.2), it's reasonable to expect that most of the world's populations might have heritabilities of intelligences at low to moderate range. In fact, we did find differences in heritability depending on the environment, as discussed below. Estimates of general cognitive ability in other animals studied in the past, mostly in primates, tend to be moderate with values ranging from 0.3 to 0.6 (for a review of studies, see Burkart, Schubiger, & van Schaik, 2017.)

Given the evidence for individual differences in general cognitive ability and our finding that these variations are heritable, one might ask why so much variation in this ability exists across individuals of a species. One could reasonably expect that higher intelligence would be broadly selected for. A candidate hypothesis proposes that animals of lesser physical prowess may have evolved compensatory cognitive abilities to facilitate their survival within the group. This idea is consistent with a general theme that has been advanced in recent years among evolutionary biologists and psychologist (Holekamp, 2007; Kamil, 2004; McNally, Brown, & Jackson, 2012). Direct experimental support for this hypothesis has been limited, although such a trade-off between cognitive abilities and fitness (larval competition) has been previously observed in *Drosophila* (Mery & Kawecki, 2003) and wild birds (Cole & Quinn, 2012). In mice, we have recently reported in mice that the predisposition for social subordination was associated

with *superior* general cognitive ability mice, suggesting this can reflect a native predisposition that precedes exposure to social pressures (Matzel, Kolata, Light, & Sauce, 2016). (For a deeper exploration of this and other hypotheses on the fitness value and persistent variation of intelligence among humans and other animals, see Burkart et al., 2017; Thornton, Isden, & Madden, 2014.)

To our surprise, our Enrichment group had a heritability not significantly different from zero, while our Control group had a moderate heritability of 0.55. As seen above, this is the opposite of what is typically found in humans. A possible explanation is that mice in the Enrichment group showed lower estimated genetic variance because of more environmental variance. Typical methods in quantitative genetics usually give first dibs to genetics, and gene-environment interactions are counted as independent genetic effects. So in human populations, gene-environment interactions in high SES groups inflate the estimates of heritability. In our study, however, there was no gene-environment interaction, and so the environmental effects were independent for both groups. And because these effects were arguably stronger in the Enrichment group than the Control group (due their environment's higher complexity), this has the potential of decreasing the estimate of heritability of the Enrichment group.

It is also possible that our estimate of GCA's heritability would be different if we measured at a different age, since here the mice were already 5-6 weeks of age (thus entering adolescence) at the start of the differential treatment. The study done at an earlier age could have resulted in a different heritability in two different ways:

- A lower heritability in younger mice because the environmental enrichment would have had a greater independent effect improving GCA. In humans, an early

age is related to more overall malleability, including cognitive traits and skills (Cunha, Heckman, Lochner, & Masterov, 2006). As an example related directly with IQ measures, Rindermann & Thompson (2016) compared immigrants to natives worldwide and found that immigration was associated with a change in IQ toward that of the native inhabitants. Depending on the quality of the host country's educational system and economic resources, increases or decreases in 4 IQ were typical. And the size of the changes depended not only on the quality of the environment, but also the age at the time of immigration – an earlier age was always related to bigger changes in IQ (Rindermann & Thompson, 2016). Therefore, if our mice were younger during the environment enrichment, their GCA gains could have been higher, and that could have led to a smaller heritability for the same reason discussed above (i.e., lower estimated genetic variance because of more environmental variance).

- A lower heritability in younger mice because the heritability of IQ is known to increase with age (at least in humans) similar to what happens across SES (seen in the Introduction). In populations of age 4-5 years, the heritability of IQ is estimated at approximately 0.22. By 16 years of age, the heritability of IQ is estimated to be 0.62. Even more striking, by age, the heritability of IQ is greater than 0.80 (Bouchard, 1997; Haworth et al., 2009). Given that the genome is largely established at birth, a population is not gaining much genetic variation as it ages (except for age-dependent genes). This suggests that the increase in IQ's heritability (or the genetic variation related to IQ variation) observed across the lifespan is, in large part, an interplay of the genome with the environment. It is

important to note that G×E does not need to be completely cumulative. The increase in heritability across age could, in part, be because as people age, both the environments as well as the genetic effects relevant to IQ get more diversified (due, respectively, to cognitively complexity of adult life, and to genes with late onset). That, in turn, can lead to more opportunities for G×E to happen. Hence, if for a 6-year-old many of the environmental and genetic effects do not interact, by age 18 the number and complexity of active environment and genes is such that there are many more opportunities for "collisions" in G×E. This G×E would be interpreted as independent genetic effects and would inflate estimates of heritability in adulthood. Therefore, a study with younger mice and with a longer time of treatment would certainly have been favorable for detecting potential G×E interactions (if they are occurring at all).

Studies with animals have in the past used designs similar to ours to test for G×E effects on physical traits (examples in Christe, Møller, Saino, & De Lope, 2000; Dupont-Nivet et al., 2008; Merilä, 1996; Omasaki, Charo-Karisa, Kahi, & Komen, 2016; Woodgate et al., 2014). However, this was the first time this type of study has been performed to examine a complex psychological trait such as intelligence. This represented the first direct test from an experimental manipulation of a gene x environment interaction behind individual differences in intelligence. To our surprise, though, there was no G×E interaction in our study.

Even though we didn't find evidence for G×E, it is important to note that GE interplay is still likely to influence intelligence in mice. Evidence from evolutionary genetics strongly suggests that traits related to survival and reproduction (like

intelligence, as seen above) have a large reservoir of hidden variance in the form of interactions (Merilä & Sheldon, 1999). This type of diversity comes from both gene-gene and gene-environment interactions, and is generally referred to as "hidden" variation because it has the potential to affect a trait, but is not expressed under typical/current conditions (Le Rouzic & Carlborg, 2008). These interactions capture genetic variation and accumulate mutations that stay latent for long periods of time, since natural selection is "blind" to anything other than additive, independent effects (Hermisson & Wagner, 2004). And since additive genetic variation is what fuels evolution by natural selection, important traits like intelligence might only continue to evolve at a fast pace because of constant new G×G and G×E hidden effects. Perhaps future studies should test other environmental factors influencing intelligence that might interact with genes. In addition, the change in heritability across environments that we found in mice further suggests a role of gene-environment interplay. Maybe this change in heritability might have been due to GE correlations instead. However, it might be even harder to directly measure correlations relative to interactions. Not only one would need to know what environmental factors influences intelligence, but also restrict individuals with particular genes to get more/less of that specific environment without correlating it with other (confounding) factors. Of course, this is much more plausible to be done first with non-human animals, and perhaps future studies will be able to estimate in GE correlations.

Lastly, it is important to emphasize why the results in the current study matter. Intelligence is arguably one of the most widely influential psychological traits. Not only is IQ a recognizably consistent measure, it also independently predicts real (and sometimes quite surprising) outcomes such as academic grades, income, social mobility,

happiness, marital stability and satisfaction, general health, longevity, reduced risk of accidents, reduced risk of drug addiction, and reduced likelihood of committing violence and crimes (Gottfredson, 1998; Mackintosh, 2011). A clear understanding of the causes of variation in intelligence is critical for future research in psychology, and its potential applications in society are self-evident. The results from the current study could help laying the groundwork for future studies on the independent effects and interactions between environment and genes in shaping general intelligence in humans.

References

Aldinger, K. a, Sokoloff, G., Rosenberg, D. M., Palmer, A. a, & Millen, K. J. (2009). Genetic variation and population substructure in outbred CD-1 mice: implications for genome-wide association studies. *PloS One*, *4*(3), e4729. http://doi.org/10.1371/journal.pone.0004729

Anderson, B. (1993). Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: intelligence? *Neuroscience Letters*, *153*(1), 98–102. http://doi.org/10.1016/0304-3940(93)90086-Z

Bailey, R. C. (1997). Hereditarian scientific fallacies. *Genetica*, *99*(2–3), 125–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9463068

Bhatia, A., Yadav, A., Zhu, C., Gagneur, J., Radhakrishnan, A., Steinmetz, L. M., … Sinha, H. (2014). Yeast Growth Plasticity Is Regulated by Environment-Specific Multi-QTL Interactions. *G3: Genes|Genomes|Genetics*, *4*(5), 769–777. http://doi.org/10.1534/g3.113.009142

Blinkhorn, S. (2003). Neuroscience: of mice and mentality. *Nature*, *424*(6952), 1004–5. http://doi.org/10.1038/4241004a

Bouchard, T. J. (1997). IQ similarity in twins reared apart: findings and responses to critics. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Intelligence: Heredity and Environment*. Cambridge University Press.

Bouchard, T. J. (2004). Genetic Influence on Human Psychological Traits. A Survey. *Current Directions in Psychological Science*, *13*(4), 148–151. http://doi.org/10.1111/j.0963-7214.2004.00295.x

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, *40*(2), e195. http://doi.org/10.1017/S0140525X16000959

Calabresi, P., Gubellini, P., Centonze, D., Picconi, B., Bernardi, G., Chergui, K., … Greengard, P. (2000). Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(22), 8443–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11069952

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., … Laibson, D. (2012). Most Reported Genetic Associations With General Intelligence Are Probably False Positives. *Psychological Science*. http://doi.org/10.1177/0956797611435528

Chandler, C. H., Chari, S., & Dworkin, I. (2013). Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in Genetics*, *29*(6), 358–366. http://doi.org/10.1016/j.tig.2013.01.009

Chou, C.-P., & Bentler, P. M. (1995). Estimates and test in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications.* Sage.

Christe, P., Møller, A. P., Saino, N., & De Lope, F. (2000). Genetic and environmental components of phenotypic variation in immune response and body size of a colonial bird, Delichon urbica (the house martin). *Heredity*, *85*(1), 75–83.

http://doi.org/10.1046/j.1365-2540.2000.00732.x
Christian, K. M., & Thompson, R. F. (2003). Neural substrates of eyeblink conditioning: acquisition and retention. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *10*(6), 427–55. http://doi.org/10.1101/lm.59603
Croston, R., Branch, C. L., Kozlovsky, D. Y., Dukas, R., & Pravosudov, V. V. (2015). Heritability and the evolution of cognitive traits. *Behavioral Ecology*, *26*(6), 1447–1459. http://doi.org/10.1093/beheco/arv088
Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education* (Vol. 1, pp. 697–812). http://doi.org/10.1016/S1574-0692(06)01012-9
Daugherty, A. M., Zwilling, C., Paul, E. J., Sherepa, N., Allen, C., Kramer, A. F., … Barbey, A. K. (2017). Multi-modal fitness and cognitive training to enhance fluid intelligence. *Intelligence*, (November), 1–12. http://doi.org/10.1016/j.intell.2017.11.001
Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., … Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*(10), 996–1005. http://doi.org/10.1038/mp.2011.85
Deary, I. J., Johnson, W., & Houlihan, L. M. (2009). Genetic foundations of human intelligence. *Human Genetics*, *126*(1), 215–32. http://doi.org/10.1007/s00439-009-0655-4
Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*(3), 201–11. http://doi.org/10.1038/nrn2793
DeFries, J. C., & Hegmann, J. P. (1970). Genetic analysis of open-field behavior. In G. Lindzey & D. D. Thiessen (Eds.), *Contributions to behavior-genetic analysis: The mouse as a prototype* (pp. 23–56). New York: Appleton-Century-Crofts.
DiLalla, D. L., & Dollinger, S. J. (2006). Cleaning up data and running preliminary analyses. In F. T. L. Leong & J. T. Austin (Eds.), *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants* (pp. 241–253). London: Sage.
Dupont-Nivet, M., Vandeputte, M., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., & Chatain, B. (2008). Heritabilities and GxE interactions for growth in the European sea bass (Dicentrarchus labrax L.) using a marker-based pedigree. *Aquaculture*, *275*(1–4), 81–87. http://doi.org/10.1016/j.aquaculture.2007.12.032
Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience*, *1*(1), 41–50. http://doi.org/10.1038/35036213
El-Soda, M., Kruijer, W., Malosetti, M., Koornneef, M., & Aarts, M. G. M. (2015). Quantitative trait loci and candidate genes underlying genotype by environment interaction in the response of Arabidopsis thaliana to drought. *Plant, Cell & Environment*, *38*(3), 585–99. http://doi.org/10.1111/pce.12418
Erickson, K. I., Hillman, C. H., & Kramer, A. F. (2015). Physical activity, brain, and cognition. *Current Opinion in Behavioral Sciences*, *4*, 27–32. http://doi.org/10.1016/j.cobeha.2015.01.005
Fabel, K., Wolf, S. A., Ehninger, D., Babu, H., Leal-Galicia, P., & Kempermann, G. (2009). Additive effects of physical exercise and environmental enrichment on adult

hippocampal neurogenesis in mice. *Frontiers in Neuroscience*, *3*(NOV), 50. http://doi.org/10.3389/neuro.22.002.2009

Finkel, D., & Pedersen, N. L. (2016). Gene–Environment Interplay in Adulthood. *Behavior Genetics*, *46*(1), 1–3. http://doi.org/10.1007/s10519-015-9775-x

Franks, B., Champagne, F. A., & Higgins, E. T. (2013). How Enrichment Affects Exploration Trade-Offs in Rats: Implications for Welfare and Well-Being. *PLoS ONE*, *8*(12), e83578. http://doi.org/10.1371/journal.pone.0083578

Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., … Plomin, R. (2005). Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavior Genetics*, *35*(5), 675–92. http://doi.org/10.1007/s10519-005-3423-9

Genoux, D., Haditsch, U., Knobloch, M., Michalon, A., Storm, D., & Mansuy, I. M. (2002). Protein phosphatase 1 is a molecular constraint on learning and memory. *Nature*, *418*(6901), 970–5. http://doi.org/10.1038/nature00928

Geschwind, D. H., & Konopka, G. (2009). Neuroscience in the era of functional genomics and systems biology. *Nature*, *461*(7266), 908–15. http://doi.org/10.1038/nature08537

Gomez-Pinilla, F., & Hillman, C. (2013). The influence of exercise on cognitive abilities. *Comprehensive Physiology*, *3*(1), 403–428. http://doi.org/10.1002/cphy.c110063

Gottfredson, L. S. (1998). Intelligence and the American ambivalence toward talent. In *Talent development IV: Proceedings* (pp. 41–58).

Haworth, C. M. a, Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J. C., van Beijsterveldt, C. E. M., … Plomin, R. (2009). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 1–9. http://doi.org/10.1038/mp.2009.55

Hermisson, J., & Wagner, G. P. (2004). The population genetic theory of hidden variation and genetic robustness. *Genetics*, *168*(4), 2271–2284. http://doi.org/10.1534/genetics.104.029173

Hoaglin, D. C., & Iglewicz, B. (1986). Performance of Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association*, *81*(396), 991–999. http://doi.org/10.1080/01621459.1986.10478363

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling : Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. http://doi.org/10.1080/10705519909540118

Hyttinen, V., Kaprio, J., Kinnunen, L., Koskenvuo, M., & Tuomilehto, J. (2003). Genetic liability of type 1 diabetes and the onset age among 22, 650 young Finnish twin pairs: A nationwide follow-up study. *Diabetes*, *52*(4), 1052–1055. http://doi.org/10.2337/diabetes.52.4.1052

Jensen, C., & Fuller, J. L. (1978). Learning performance varies with brain weight in heterogeneous mouse lines. *Journal of Comparative and Physiological Psychology*, *92*(5), 830–836. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0018175284&partnerID=40&md5=af10201ff781ce4dd792ac80cb3b1c55

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford press.

Kolata, S., Light, K., Grossman, H. C., Hale, G., & Matzel, L. D. (2007). Selective attention is a primary determinant of the relationship between working memory and general learning ability in outbred mice. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *14*(1), 22–8. http://doi.org/10.1101/lm.408507

Kolata, S., Light, K., & Matzel, L. D. (2008). Domain-Specific and Domain-General Learning Factors are Expressed in Genetically Heterogeneous CD-1 mice. *Intelligence*, *36*(6), 619–629. http://doi.org/10.1016/j.intell.2007.12.001

Kolata, S., Light, K., Townsend, D. a, Hale, G., Grossman, H. C., & Matzel, L. D. (2005). Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiology of Learning and Memory*, *84*(3), 241–6. http://doi.org/10.1016/j.nlm.2005.07.006

Kolata, S., Light, K., Wass, C. D., Colas-Zelin, D., Roy, D., & Matzel, L. D. (2010). A dopaminergic gene cluster in the prefrontal cortex predicts performance indicative of general intelligence in genetically heterogeneous mice. *PloS One*, *5*(11), e14036. http://doi.org/10.1371/journal.pone.0014036

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., … Van Alstyne, M. (2009). Computational social science. *Science*, *323*(5915), 721–723. http://doi.org/10.1126/science.1167742

Le Rouzic, A., & Carlborg, Ö. (2008). Evolutionary potential of hidden genetic variation. *Trends in Ecology and Evolution*, *23*(1), 33–37. http://doi.org/10.1016/j.tree.2007.09.014

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., … Hemminki, K. (2000). Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, *343*(2), 78–85. http://doi.org/10.1056/NEJM200007133430201

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *The Lancet*, *373*(9659), 234–239. http://doi.org/10.1016/S0140-6736(09)60072-6

Light, K., Kolata, S., Hale, G., Grossman, H., & Matzel, L. D. (2008). Up-regulation of exploratory tendencies does not enhance general learning abilities in juvenile or young-adult outbred mice. *Neurobiology of Learning and Memory*, *90*(2), 317–29. http://doi.org/10.1016/j.nlm.2008.04.007

Mackintosh, N. (2011). *IQ and Human Intelligence*. Oxford: Oxford University Press.

Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., … Gandhi, C. C. (2003). Individual differences in the expression of a "general" learning ability in mice. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(16), 6423–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12878682

Matzel, L. D., Kolata, S., Light, K., & Sauce, B. (2016). The tendency for social submission predicts superior cognitive performance in previously isolated male mice. *Behavioural Processes*. http://doi.org/10.1016/j.beproc.2016.07.011

Matzel, L. D., Sauce, B., & Wass, C. (2013). The Architecture of Intelligence:

Converging Evidence From Studies of Humans and Animals. *Current Directions in Psychological Science*, *22*(5), 342–348. http://doi.org/10.1177/0963721413491764

Matzel, L. D., Townsend, D. a, Grossman, H., Han, Y. R., Hale, G., Zappulla, M., … Kolata, S. (2006). Exploration in outbred mice covaries with general learning abilities irrespective of stress reactivity, emotionality, and physical attributes. *Neurobiology of Learning and Memory*, *86*(2), 228–40. http://doi.org/10.1016/j.nlm.2006.03.004

Merilä, J. (1996). Genetic variation in offspring condition: an experiment. *Functional Ecology*, 465–474.

Merilä, J., & Sheldon, B. C. (1999). Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity*, *83 (Pt 2)*(June), 103–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10469197

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. http://doi.org/10.1037/0003-066X.51.2.77

Nokia, M. S., Lensu, S., Ahtiainen, J. P., Johansson, P. P., Koch, L. G., Britton, S. L., & Kainulainen, H. (2016). Physical exercise increases adult hippocampal neurogenesis in male rats provided it is aerobic and sustained. *The Journal of Physiology*, *594*(7), 1855–1873. http://doi.org/10.1113/JP271552

Oliverio, A., Castellano, C., & Messeri, P. (1972). Genetic analysis of avoidance, maze, and wheel-running behaviors in the mouse. *Journal of Comparative and Physiological Psychology*, *79*(3), 459–473. http://doi.org/10.1037/h0032835

Omasaki, S. K., Charo-Karisa, H., Kahi, A. K., & Komen, H. (2016). Genotype by environment interaction for harvest weight, growth rate and shape between monosex and mixed sex Nile tilapia (Oreochromis niloticus). *Aquaculture*, *458*, 75–81. http://doi.org/10.1016/j.aquaculture.2016.02.033

Phelps, E. a, & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *48*(2), 175–187. http://doi.org/10.1016/j.neuron.2005.09.025

Philip Chen, C. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347. http://doi.org/10.1016/j.ins.2014.01.015

Plomin, R. (2001). The genetics of G in human and mouse. *Nature Reviews Neuroscience*, *2*(2), 136–141. http://doi.org/10.1038/35053584

Riemann, R., Angleitner, A., & Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins teared together using the self- and peer report. *Journal of Personality*, *65*(3), 449–475. http://doi.org/10.1111/j.1467-6494.1997.tb00324.x

Rindermann, H., & Thompson, J. (2016). The cognitive competences of immigrant and native students across the world: An analysis of gaps, possible causes and impact. *Journal of Biosocial Science*, *48*(1), 66–93. http://doi.org/10.1017/S0021932014000480

Rockman, M. V. (2008). Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature*, *456*(7223), 738–44. http://doi.org/10.1038/nature07633

Sauce, B., & Matzel, L. D. (2013). The causes of variation in learning and behavior: why individual differences matter. *Frontiers in Psychology*, *4*(July), 395.

http://doi.org/10.3389/fpsyg.2013.00395
Sauce, B., & Matzel, L. D. (2017). The Paradox of Intelligence: Heritability and Malleability Coexist in Hidden Gene-Environment Interplay. *Psychological Bulletin*. http://doi.org/10.1037/bul0000131
Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*(1), 3–15. http://doi.org/10.1191/096228099671525676
Schafer, J. L. (2010). *Analysis of incomplete multivariate data*. CRC press.
Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, *99*(6), 323–338. http://doi.org/10.3200/JOER.99.6.323-338
Silventoinen, K., Magnusson, P. K. E., Tynelius, P., Kaprio, J., & Rasmussen, F. (2008). Heritability of body size and muscle strength in young adulthood: A study of one million Swedish men. *Genetic Epidemiology*, *32*(4), 341–349. http://doi.org/10.1002/gepi.20308
Smith, A. M., Spiegler, K. M., Sauce, B., Wass, C. D., Sturzoiu, T., & Matzel, L. D. (2013). Facilitation of the cognitive enhancing effects of working memory training through conjoint voluntary aerobic exercise. *Behavioural Brain Research*, *256*, 626–635. http://doi.org/10.1016/j.bbr.2013.09.012
Smith, E. N., & Kruglyak, L. (2008). Gene-environment interaction in yeast gene expression. *PLoS Biology*, *6*(4), 810–824. http://doi.org/10.1371/journal.pbio.0060083
Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292.
Sternberg, R. J., & Kaufman, J. C. (1998). Human Abilities. *Annual Review of Psychology*, *49*(1), 479–502. http://doi.org/10.1146/annurev.psych.49.1.479
Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry*, *157*(10), 1552–1562. http://doi.org/10.1176/appi.ajp.157.10.1552
Templeton, A. R. (2006). *Population Genetics and Microevolutionary Theory*. Hoboken, New Jersey: Wiley-Liss.
Tenesa, A., & Haley, C. S. (2013). The heritability of human disease: estimation, uses and abuses. *Nature Reviews Genetics*, *14*(2), 139–149. http://doi.org/10.1038/nrg3377
Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: linking individual cognitive differences to fitness. *Behavioral Ecology*, *25*(6), 1299–1301. http://doi.org/10.1093/beheco/aru095
Tucker-Drob, E. M., Briley, D. a., & Harden, K. P. (2013). Genetic and Environmental Influences on Cognition Across Development and Context. *Current Directions in Psychological Science*, *22*(5), 349–355. http://doi.org/10.1177/0963721413485087
Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of iq in young children. *Psychological Science*, *14*(6), 623–628. http://doi.org/10.1046/j.0956-7976.2003.psci_1475.x
van Ijzendoorn, M. H., Juffer, F., & Poelhuis, C. W. K. (2005). Adoption and cognitive development: a meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin*, *131*(2), 301–16.

http://doi.org/10.1037/0033-2909.131.2.301
van Ijzendoorn, M. H., Luijk, M. P. C. M., & Juffer, F. (2008). IQ of Children Growing Up in Children's Homes: A Meta-Analysis on IQ Delays in Orphanages. *Merrill-Palmer Quarterly*, *54*(3), 341–366. http://doi.org/10.1353/mpq.0.0002
van Praag, H. (2005). Exercise Enhances Learning and Hippocampal Neurogenesis in Aged Mice. *Journal of Neuroscience*, *25*(38), 8680–8685. http://doi.org/10.1523/JNEUROSCI.1731-05.2005
van Praag, H., Kempermann, G., & Gage, F. H. (2000). Neural consequences of environmental enrichment. *Nature Reviews. Neuroscience*, *1*(3), 191–8. http://doi.org/10.1038/35044558
Volkers, R. J., Snoek, L., Hubar, C. J. V. H., Coopman, R., Chen, W., Yang, W., … Kammenga, J. E. (2013). Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild Caenorhabditis elegans populations. *BMC Biology*, *11*(1), 93. http://doi.org/10.1186/1741-7007-11-93
Walsh, R. N., Cummins, R. A., & Budtz-Olsen, O. E. (1973). Environmentally induced changes in the dimensions of the rat cerebrum: A replication and extension. *Devel. Psychobiol.*, *6*(1), 3–7.
Wass, C., Denman-Brice, A., Rios, C., Light, K. R., Kolata, S., Smith, A. M., & Matzel, L. D. (2012). Covariation of learning and "reasoning" abilities in mice: evolutionary conservation of the operations of intelligence. *Journal of Experimental Psychology. Animal Behavior Processes*, *38*(2), 109–24. http://doi.org/10.1037/a0027355
Wass, C., Pizzo, A., Sauce, B., Kawasumi, Y., Sturzoiu, T., Ree, F., … Matzel, L. D. (2013). Dopamine D1 sensitivity in the prefrontal cortex predicts general cognitive abilities and is modulated by working memory training. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *20*(11), 617–627. http://doi.org/10.1101/lm.031971.113
Weber, E. M., & Olsson, I. A. S. (2008). Maternal behaviour in Mus musculus sp.: An ethological review. *Applied Animal Behaviour Science*, *114*(1–2), 1–22. http://doi.org/10.1016/j.applanim.2008.06.006
Wolf, J. B., Vaughn, T. T., Pletscher, L. S., & Cheverud, J. M. (2002). Contribution of maternal effect QTL to genetic architecture of early growth in mice. *Heredity*, *89*(4), 300–310. http://doi.org/10.1038/sj.hdy.6800140
Woodgate, J. L., Buchanan, K. L., Bennett, A. T. D., Catchpole, C. K., Brighton, R., & Leitner, S. (2014). Environmental and genetic control of brain and song structure in the zebra finch. *Evolution*, *68*(1), 230–240. http://doi.org/10.1111/evo.12261