

Research article

Effects of behavioural types on the problem-solving performance of wild house mice under controlled and semi-natural conditions

Alexandros Vezyrakis^{1,2}, Anja Guenther¹ and Valeria Mazza^{2,3}

¹RG Behavioural Ecology of Individual Differences, Max Planck Institute for Evolutionary Biology, Plön, Germany

²Animal Ecology, University of Potsdam, Germany

³Department of Ecological and Biological Sciences, University of Tuscia, Viterbo, Italy

Correspondence: Alexandros Vezyrakis (vezyrakis@evolbio.mpg.de)

Oikos

2025: e10787

doi: [10.1111/oik.10787](https://doi.org/10.1111/oik.10787)

Subject Editor: Denis Reale

Editor-in-Chief: Pedro Peres-Neto

Accepted 25 November 2024



Animals often face challenges that require them to come up with solutions to novel problems or to find new solutions to existing ones; i.e. they need to innovate. However, not all individuals in a population are equally likely to solve novel problems, and it is unclear which individual characteristics make a successful innovator. Theoretical frameworks suggest the importance of intrinsic (e.g. individual characteristics) and extrinsic (e.g. study condition) factors on problem-solving performance. Such frameworks have been empirically tested in model, highly neophobic species, leaving the generality of these processes unclear. We examined whether behavioural traits such as exploration and risk-taking are linked to problem-solving behaviour, using two replicated populations ($n = 121$) of wild house mice *Mus musculus domesticus* living under semi-natural conditions. There, we presented a battery of four problem-solving setups that individuals could access voluntarily, and we tested the mice for risk-taking and exploration. Furthermore, after acclimatising to cages, we tested a subset ($n = 50$) of the same individuals in controlled conditions, to validate the cross-context stability of cognitive performance and potential influences of behaviours. We placed single individuals overnight in arenas containing another four novel problem-solving setups. Contrasting most existing literature, we found no direct effects of behavioural type on the likelihood to problem-solve in either condition. However, there was an indirect effect, with shyer individuals visiting the problems more, which improved their likelihood of solving them. Additionally, mice were more likely to solve alone, and individuals were not consistent across conditions. Our findings suggest that exploration and risk-taking do not affect the ability to problem-solve across different conditions, but impact the non-cognitive steps that lead to the final performance. Also, individuals did not perform consistently across conditions, questioning the ecological validity of measures taken under controlled, artificial conditions when they do not reflect the animals' natural experience.

Keywords: Behavioural traits, ecological validity, innovation, personality, problem-solving, semi-natural conditions



www.oikosjournal.org

© 2025 The Author(s). Oikos published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

In the wild, animals often live under challenging and fast-changing conditions, and can greatly improve their survival and fitness by adjusting their behaviour to match the demands of their environment. The ability to selectively modify behaviours in response to changing conditions, also known as behavioural flexibility, can help them overcome challenges in foraging, attracting mates, or avoiding predators (Wright et al. 2010). These behavioural adjustments can often arise much more quickly than phenotypic changes in morphology, physiology or adjustments mediated by transgenerational effects. An important aspect of behavioural flexibility is innovation, which is the ability to solve novel problems or come up with novel solutions to old problems (Kummer and Goodall 1985, Laland and Reader 1999, Reader and Laland 2003, Ramsey et al. 2007, Amici et al. 2019). Innovative behaviours allow animals to exploit new resources, occupy new ecological niches, and even defend themselves against novel predators (Amici et al. 2019). Innovation has also been proposed to have an impact on macroevolution (Nicolakakis et al. 2003, Ramsey et al. 2007), either by accelerating speciation and diversification (Sol 2003) or by shielding species from environmental change e.g. through innovative niche construction or by rapid-response mechanisms ('plastic rescue', Day et al. 2003, Ramsey et al. 2007, Fox et al. 2019).

A large variety of innovative behaviours have been recorded in the wild, from the famous example of British tits opening milk bottles (Fisher 1949) to the Australian cockatoos *Cacatua galerita* opening household bins (Klump et al. 2021) and even engaging in a possible problem-solving arms race with humans (Klump et al. 2022). Other than observing innovations in the wild, researchers often use problems that the animals need to solve to obtain a reward, and solving these problems requires them to innovate (Morand-Ferron et al. 2011, Overington et al. 2011, Benson-Amram and Holekamp 2012).

A current focus in behavioural ecology and evolution is to understand why some individuals of the same species solve novel problems while others do not. Novel situations come with inherent risk–reward tradeoffs, which led to the formulation of hypotheses that link animal personality (consistent individual differences in behaviour across contexts and time; Réale et al. 2007, Stamps and Groothuis 2010, Kaiser and Müller 2021), to differences in problem-solving ability (Reader and Laland 2003, Carere and Locurto 2011, Sih and Del Giudice 2012, Griffin et al. 2015). Theory suggests that behavioural traits that enhance how individuals encounter, approach, explore, or sample their surroundings such as high levels of neophilia, exploration or boldness – are associated with increased problem-solving abilities (Reader and Laland 2003). For example, in many species, explorative or bold individuals have repeatedly been found to be better problem-solvers (Greenberg 2003, Reader and Laland 2003, Griffin and Guez 2014, Amici et al. 2019).

It is not always clear, however, whether these differences appear because bold and explorative animals have superior

problem-solving capacities or simply because they acquire and process information differently. Sih and Del Giudice (2012) proposed a theoretical framework under which individuals with certain behavioural types (e.g. bold, explorative) may appear better at a cognitive task because they encounter such tasks more often, and this could be irrespective of their cognitive abilities. Indeed, high boldness or exploration rates may have an indirect effect on problem-solving, such as improving participation and motivation (van Horik and Madden 2016, Van Horik et al. 2017). In contrast, neophobic (i.e. fearful of novelty) individuals are less likely to interact with novel problems in their environment, therefore limiting their probability to solve them (Tebbich et al. 2016). In fact, innovations themselves can arise from simple trial-and-error, and it is clear that individuals who get more opportunities to find, approach and interact with problem-solving tasks would have higher success in solving them (Seed and Byrne 2010, Thornton and Samson 2012, Amici et al. 2019).

Additionally, social context can both improve the problem-solving performance of individuals through social learning or hinder it due to competition for access to (novel) resources created from e.g. high densities or dominance hierarchies (Rowell et al. 2021). Social learning, on the one hand, can help animals solve problems by copying skilled demonstrators (Aplin et al. 2013). Social competition, on the other hand, can affect which individuals are more likely to have the possibility to problem-solve. Two opposing evolutionary hypotheses on innovation predict which individuals are likely to be problem-solvers: the 'excess of energy' (EE) hypothesis (Kummer and Goodall 1985) and the 'bad competitor' (BC) hypothesis (Laland and Reader 1999, Reader and Laland 2003, reviewed by Amici et al. 2019). According to the EE hypothesis, individuals with superior competitive abilities or social status (usually due to their better body condition or health status) should have more energy and time to devote to innovation, and thus perform better. Indeed, dominant individuals perform better than subordinates in some species (European starlings *Sturnus vulgaris*, Boogert et al. 2006, coyotes *Canis latrans*, Young et al. 2019), or the presence of dominant individuals reduces the performance of subordinates (spotted hyaenas *Crocuta crocuta*, Drea and Carter 2009). However, subordinates are sometimes better problem-solvers (i.e. meerkats *Suricata suricata*, Thornton and Samson 2012), supporting the BC hypothesis that predicts that their lower competitive ability makes them more likely to seek alternative resources through innovations.

To understand how individual characteristics, potentially in interaction with social dynamics cause within-population variation in problem-solving, it is paramount to clarify how personality influences problem-solving ability and performance. Two major gaps currently hinder our understanding of the impact of personality on problem-solving, namely 1) the ecological validity of experimental findings on captive animals, and how these translate to the wild (Amici et al. 2019) and 2) the current strong focus on certain taxa exhibiting specific characteristics (i.e. boldness, neophobia; see Griffin and Guez 2014).

Concerning the first gap, it is currently unclear whether problem-solving in nature is actually reflected in testing individuals under controlled laboratory conditions, in which animals are often kept in non-stimulating or even deprived conditions (Tomasello and Call 2011, Amici et al. 2019). These experimental conditions can have a significant effect on the results reported; e.g. whether the test condition itself imposes stress on the animals or not (Delacoux and Guenther 2023). Only a handful of studies have compared problem-solving under natural versus captive conditions (Webster and Lefebvre 2001, Benson-Amram et al. 2013) and, to our knowledge, none have done so using the same individuals in both conditions, despite this being identified as a methodological omission (Carere and Locurto 2011). Individuals in captive conditions have been shown to innovate more than their wild conspecifics (Reader and Laland 2003, Benson-Amram and Holekamp 2012, Amici et al. 2019), but a within-individual study design might allow us to test for the translatability of results between conditions. We believe that linking these two conditions is essential to improve the ecological validity of the studies of problem-solving in wild animals.

Concerning the second gap, a large body of literature focuses on birds, a taxon with well-known problem-solving abilities, which is mostly neophobic (Greenberg and Mettke-Hofmann 2001, Greenberg 2003, O'Hara et al. 2017). However, studying other, more neophilic taxa, may let us gain a more complete insight into how different behavioural traits interact with the risk-reward tradeoffs that are inherent to solving novel problems. As shown by the few studies on mammals, bold and friendly guinea pigs *Cavia porcellus* were better problem-solvers (Guenther and Brust 2017) and, similarly, bolder spotted hyenas *Crocuta crocuta* were more likely to innovate (Benson-Amram and Holekamp 2012). Bridging this taxonomic gap is paramount to draw generalizable conclusions, as in many cases the results can be species- and context-specific (Boogert et al. 2018, Dougherty and Guillelte 2018).

An ideal taxon to fill both aforementioned gaps should be known to innovate and exhibit stable individual differences in risk-taking. Multiple species of rodents have been shown to innovate both in the wild and in the laboratory, and they are also known to have behavioural traits that are easy to measure using standardised tests with validated ecological relevance (Krebs et al. 2019). Wild house mice *Mus musculus domesticus*, group-living rodents, are especially well-suited to study in this context (Vrbanec et al. 2021) and can be studied thoroughly in artificial semi-natural enclosures with high ecological validity, as human structures like farm buildings are their natural habitat (König et al. 2012).

In the present study, we investigated how risk-taking behavioural traits in a potentially threatening situation measured as exploration of an unknown area and time spent away from protection, interact with the problem-solving success of wild house mice *Mus musculus domesticus* across two different conditions. We used individuals living freely in large, semi-natural enclosures, and tested their performance in a

battery of voluntary problem-solving tasks presented directly in the enclosures. Then, after a period of being housed singly or in small, same-sex groups in cages, we tested them again for problem-solving under controlled conditions. By testing individuals in both conditions, we aimed to investigate whether the behavioural traits of exploration and risk-taking affect problem-solving across different conditions, as well as how the conditions themselves impact the problem-solving performance of wild house mice.

We predicted that risk-taking propensity influences the performance of individuals similarly across conditions. Specifically, higher exploration and risk-taking should improve problem-solving performance, as these traits affect different non-cognitive components of problem-solving (e.g. the likelihood of encountering a problem, motivation, and persistence). Additionally, the social environment is an important component of all behaviours of a group-living species such as the house mouse. As such, individuals living in large groups might have less time to devote to problem-solving when compared to individuals who are alone. Therefore, we expect problem-solving performance to be lower in the semi-natural enclosures compared to controlled conditions in which individuals can interact with the test setups alone and undisturbed.

Methods

Animals

We set up two replicated semi-natural enclosures (20 m² each) by releasing adult wild house mice ($n=35$ individuals in the first enclosure, 20 females and 15 males, $n=34$ individuals in the second enclosure, 20 females and 14 males, all aged 50–100 days). These founders of the colonies were descendants of wild-caught individuals from the Cologne/Bonn region, Germany. Each enclosure contained woodchip bedding, nesting material (cotton rolls, toilet paper) and thirteen nest boxes where mice were free to build nests and breed (see the Supporting information for a picture and a schematic representation of the enclosures). Equally distributed in each enclosure, we placed nine feeding stations where mice could access food (Altromin 1324, Germany) and water ad libitum. Temperature and light-dark cycle varied naturally but light was supplemented between 8 a.m. and 5 p.m., and underfloor heating prevented temperatures below 10°C. Enclosures were covered by a roof, inaccessible to predators and generally designed to resemble the natural conditions under which house mice would establish a colony, like for example a barn (König et al. 2012). These semi-natural enclosures are an established system for wild house mouse colonies, leading to close-to-natural population development and social- and genetic structure, allowing mice to exhibit their natural behaviours (estimated carrying capacity of 140 individuals per enclosure, for more details on the enclosures see Prabh et al. 2023 and the Supporting information therein). During monthly monitorings, all mice were caught,

weighed, and new individuals ($W > 10$ g) were fitted with RFID transponders (ISO Transponders, PlanetID) for individual identification.

After being kept under semi-natural conditions for six months, when the founding generation was 7–10 months old, we moved all individuals (for all sample sizes see Table 1) from the enclosures into Macrolon Type III cages ($L \times W \times H$: $382 \times 220 \times 150$ mm) for the second part of the experiment (described in section ‘Arenas’). Adult males were housed individually to prevent them from fighting, while juvenile males and females were housed in pairs/triads based on the location they were found, according to standard cage housing practices. The cages were lined with woodchip bedding and contained a shelter, nesting material and a running wheel for enrichment. The mice were kept in the same light-dark and temperature conditions as before, with ad libitum access to food and water.

Behavioural traits assay

Behavioural traits such as exploration and risk-taking are considered to favour innovative problem-solving, and to measure them we opted for the standardised and well-established Open field test. In this test, the focal animal can explore a novel, bright and open area (Gould et al. 2009, Peralas et al. 2017). This test was originally developed to study rodents’ responses to challenging and potentially dangerous situations (Hall and Ballachey 1932), as they exhibit thigmotaxis, i.e. avoiding open areas and preferring to stay in close contact with the walls of the apparatus. In mice, measurements in the Open field test have been well validated in laboratory strains, but also linked with ecological parameters for wild-caught individuals. For example, the distance covered by the animal in a set amount of time reflects its exploratory tendencies in larger and more complex semi-natural environments (Krackow 2003, Krebs et al. 2019), while the time spent in the centre reflects its risk-taking strategy (i.e. more risk-prone animals would spend more time in the central open area, while more risk-averse animals would spend more time close to the walls; Krebs et al. 2019).

Mice were filmed for five minutes while exploring a square (W 60 cm $\times$ L 60 cm $\times$ H 80 cm) Open field. A previous study using the same wild house mouse system has provided ecological validation of the measured behavioural traits (Krebs et al. 2019); as such, we used 1) the distance covered as a measure of exploratory tendencies and 2) the time spent in the central zone of the Open field as a measure of risk-taking strategy. Mice were caught opportunistically directly from the semi-natural enclosures using unbaited live traps. Care was taken to minimise the time each mouse spent in the trap to avoid unnecessary stress and any other adverse consequences (e.g. starvation, dehydration, endangering dependent offspring), therefore traps were checked often and no individual was trapped or out of the enclosure for more than 30 min. Trapping began at least 1 h after sunset and lasted for multiple hours. Each individual was tested twice, one month apart, to assess the repeatability of the measurements. The Open field arena was cleaned with disinfectant (35% propanol, 25% ethanol; neoform Rapid, Germany) between tests.

Problem-solving tests

Semi-natural enclosures

One month after the mice had been introduced to the semi-natural enclosures, we started to test for problem-solving by presenting four consecutive problem-solving setups that the mice could voluntarily access and solve. Each setup was presented for five nights (following a regime of two consecutive testing nights, then a break of five nights to boost motivation, then three more nights of testing), and there was a minimum two-week gap between setups (for a graphical representation of the testing timeline see the Supporting information). Overall, the setups were present in the enclosures for 40 h.

For each setup that was presented, five testing boxes were placed in predefined locations in each enclosure in the empty space between the nest boxes to minimise the chances that a box would be inside a mouse’s territory, and therefore mice not belonging to the family would be denied access (Supporting information). The floor of the testing boxes was

Table 1. Sample sizes of individuals that participated in different parts of the experiment. Each group represents a subset of the group shown above in the table. DS1-5 indicate the datasets that these groups create, some of which were used in the following analyses. *Seven individuals were born late into the problem-solving experiment and four individuals died early into the experiment. As such, they are not included in the $n=101$ individuals that were examined for problem-solving in the enclosures. They were used, however, to calculate repeatabilities because they still represent part of the phenotyped population (bringing the sample size for these analyses to $n=121$).

Description	n
All individuals (DS1)	193
<i>All individuals that lived in the semi-natural enclosures for the duration of the experiment</i>	
Problem-solving in the enclosures (DS2)	121
<i>Individuals who could be tracked for at least half the time of the problem-solving experiment</i>	
Open field (DS3)	101*
<i>Individuals that were tested in the Open field</i>	
Repeated Open field (DS4)	61
<i>Individuals that were tested twice in the Open field</i>	
Arenas (DS5)	50
<i>Individuals that were tested in the Arenas</i>	

covered with bedding from the location they were placed, in order to keep the olfactory cues of the environment consistent. The testing boxes had a single entrance, which was fitted with an RFID antenna (EuroID, Netherlands) to record the mice that entered/exited the box, and all activity was filmed from the top (Panasonic HC-V 180, for the camera view Fig. 1.2).

The boxes were placed in the enclosures in the evening (at least 1 h after sunset) for 2 h and all activity was video recorded under red light. Trapping for behavioural testing was not performed on the same nights in which problem-solving was tested. The boxes including the setups were removed from the enclosures after the end of the two hours. To aid with video analysis, the readings from each antenna were synchronised and embedded in the corresponding video in the form of subtitles. As such, during video analysis, the observer could easily identify each mouse in real-time.

The four test setups (previously used for small rodents; Mazza and Guenther 2021, including wild house mice; Vrbanec et al. 2021) were presented in the following order: 1) omnidirectional slider, 2) inverted cup, 3) one-directional slider and 4) petri-dish (Fig. 1.1). All setups required a simple movement to be solved, requiring minimal force that all mice, regardless of age, size or physical strength could solve. To solve setup 1), the mice could push, pull, or slide the metal lid in any direction, while to solve setup 3) they could do the same but only towards one direction. To solve setup 2), the mice could push, pull or lift the semi-transparent cup while the base remained stable, while setup 4) could be solved by lifting the lid of the petri dish. All setup

solutions could arise both from insightful attempts as well as trial-and-error.

Setups were baited with a dried mealworm (one *Tenebrio molitor* larva). The mealworms were a novel food source at the start of the experiment to increase the interest of the mice, that display neophilic tendencies towards food sources. All individuals could voluntarily access the testing boxes and setups. Videos were analysed using BORIS (Friard and Gamba 2016). We recorded the identities of individuals that (a) investigated the entrance of the testing box, (b) entered the testing box, (c) interacted with the setup, and (d) solved the setup. For (b) and (c) we considered only the instances where the full body of the mouse, including all four paws, reached the corresponding location. As numerous individuals were born in the enclosures, we included the generation (founders or born in the enclosures) in the subsequent analyses. Additionally, we obtained the number of times that each individual was detected by any of the antennas at the entrance of the problem-solving boxes for each round of testing.

Moreover, to estimate how potential social effects affect problem-solving performance, we examined the number of mice present inside the problem-solving boxes during recorded solving events ($n=87$), and during a random sample of unsuccessful attempts ($n=100$). In the events where when multiple mice were present in the box, we ranked their interactions as negative (aggressive interactions such as chasing), neutral (merely being present together, no socio-positive or aggressive interactions), or positive (mutual interest in the setup and/or preceding of allogrooming and similar behaviours).

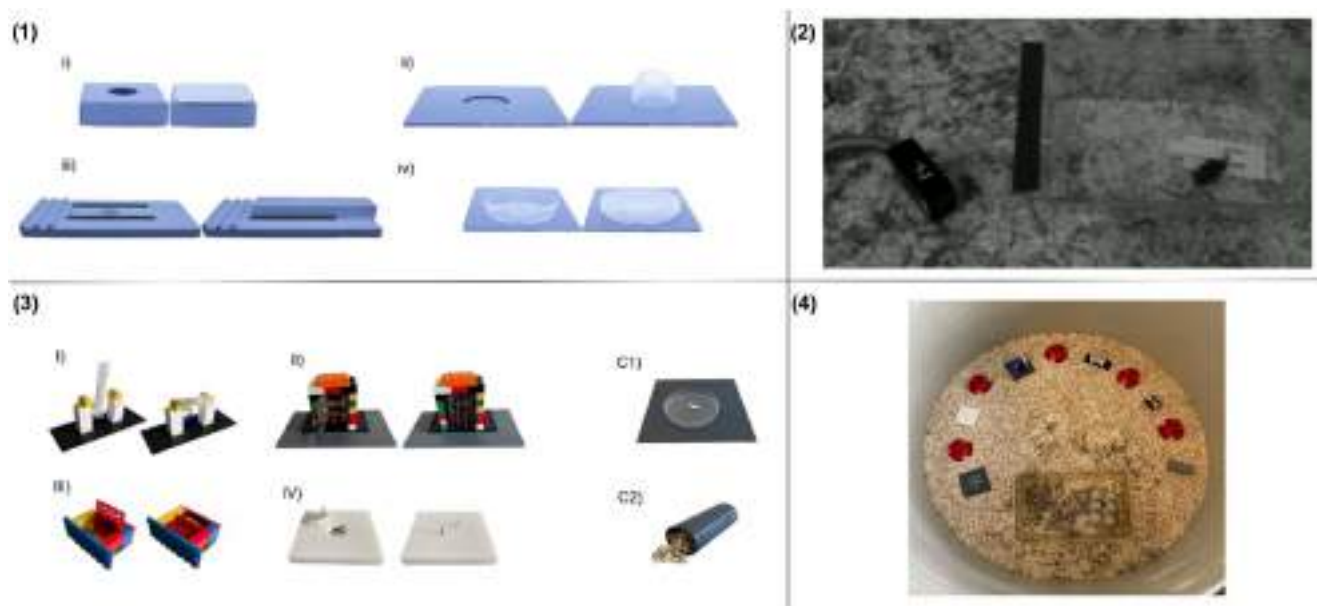


Figure 1. Presentation of the problem-solving setups in the semi-natural enclosures (top) and arenas (bottom). (1) The four problem-solving setups used in the semi-natural enclosures: (i) omnidirectional slider, (ii) inverted cup, (iii) one-directional slider and (iv) petri-dish. (2) The camera view of the problem-solving box with setup and RFID antenna. (3) The four problem-solving setups and two controls used in the arenas: (I) flip-tube, (II) barred-door, (III) trap-door, (IV) lift platform, (C1) open petri dish and (C2) tube with bedding. (4) The camera view of the arena, with the cage in and all the setups on opposite sides. In (1) and (3), the problem-solving setups are shown in their open (solved) configuration on the left and their closed (unsolved) configuration on the right.

Arenas

After one month of habituation to being housed in cages, we performed another round of problem-solving tests, in which each mouse participated individually. To understand the impact of risk-taking and exploration on problem-solving, we selected the 16 highest and lowest-ranking individuals in each of the two behavioural measurements from the Open field (using the average time in the centre and distance covered). As a few individuals belonged simultaneously in two of the above groups (e.g. in the group with the highest time in the centre and in the group with the highest distance covered), we ended up with $n=50$ individuals to test. Before testing, all individuals received mealworms to ensure that they were familiar with this food reward even if they had not encountered them in the semi-natural enclosures already.

Mice were placed overnight (3 p.m. – 8 a.m.; 17 hours total) in a type III cage with openings on both wide sides that contained a handful of bedding from their home cage. The cage was located on one side of a round arena ($\varnothing$ 120 cm); on the other side we placed four novel problem-solving setups (i.e. problems the animals could not know from the semi-natural enclosures): 1) trap-door, 2) lift platform, 3) barred-door and 4) flip-tube, as well as two controls: (C1) tube with bedding and (C2) open petri dish (Fig. 1.3). These setups were the same or slight alterations of setups previously used to measure problem-solving in rodents (Mazza and Guenther 2021, Vrbanc et al. 2021) and all posed the same level of difficulty as the setups used in the semi-natural enclosures, requiring minimal force and a simple movement to solve. To solve setup 1), the mice could lift the plastic door upwards, and to solve setup 3) the mice could pull the plastic door upwards. Both setups could be solved by rotating the whole structure so the reward would drop out. Setup 2) could be solved by lifting the plastic platform upwards, while setup 4) could be solved by pulling the tube downwards. Problem-solving setups were baited with a familiar reward (half a *T. molitor* larva) as well as an unfamiliar reward (a similarly sized piece of dried mango). One control was baited with the familiar reward (C1) and one with the unfamiliar (C2). The novel food reward served as an additional driver of novelty for the new testing conditions. All setups were arranged about 5 cm from the walls of the arena in a semi-circular fashion, with the controls always at the two edges and the problem-solving setups randomly in between. Between the setups, we placed semi-transparent red igloos to serve as shelters for the mice (Fig. 1.4). Food and water were available ad libitum from the lid of the cage, accessible even if a mouse did not exit the cage the entire night.

We video recorded all activity for the first 3 h after the mice were placed in the cages in the arenas, and also whether the problems had been solved or the controls eaten the next morning. During video analysis using BORIS, we recorded whether an individual (a) exited the cage, and (b) interacted with a setup.

Statistical analyses

All statistical analyses were performed in R ver. 4.2.0 (www.r-project.org).

Repeatability analysis

Repeatability for the behavioural measurements ($n=61$ with repeated measurements, $n=51$ with a single measurement) was tested with the *rptR::rpt* function (Nakagawa and Schielzeth 2010, Stoffel et al. 2017), adding the trial number as a fixed and the individual ID as a random effect. For the time spent in the centre, the values were square-root transformed to reach normality. We assessed the correlation between the two behavioural traits using the repeated measures correlation package 'rmcorr' (Bakdash and Marusich 2017). As they were not correlated ($r=0.06$, $df=60$, $p=0.67$, 95% CI=[-0.19, 0.30]), we used both as predictors in the following models. To test if the individuals visited the four setups in the semi-natural enclosures with similar frequency, we ran a repeatability analysis with the number of reads from the antennas at the entrance of the problem-solving box for each setup as the response variable (square-rooted), the type of setup as a fixed and the individual as a random effect. The response variable assumed a normal distribution.

To test if the likelihood to solve was repeatable across the four setups presented in the semi-natural enclosures, we ran a binomial mixed model with the setup type as a fixed and the individual as a random effect. This analysis was performed only using individuals who interacted with at least one setup, and using only one measurement per individual per type of setup. As a result, repeated solutions of the same setup did not contribute to this assessment. Overdispersion was checked using the 'DHARMA' package (Hartig and Lohse 2022). Additionally, we ran another binomial repeatability analysis with the likelihood to solve a problem in each condition (semi-natural versus arena) as a response variable, the testing condition as a fixed and the individual as a random effect. All repeatability analyses were run using 1000 bootstraps and 1000 permutations, and as such we report the p-values estimated from likelihood ratio tests [LRT] and permutations.

Problem-solving in the semi-natural enclosures

We run two models to understand the relationships between problem-solving performance, the number of visits in the problem-solving boxes and the behavioural traits measured in the Open field. The first model (model 1) was a generalised linear mixed model with a binomial distribution, the likelihood to solve each of the setups i-iv as the response variable, and included as fixed effects the distance (in meters), and the time in the centre (in seconds, square-rooted) in the first trial of the Open field, the generation (founding generation or born in the enclosures), the sex, the type of setup as well as the number of reads from the antennas at the entrance of the problem-solving box (square-rooted) for each setup. The individual ID was specified as a random effect, nested within the identity of the mother of

each individual to account for relatedness. If an individual solved a setup at least once, then it was characterised as a solver for that setup type; as such, repeated solutions of the same setup did not contribute to this measurement. The second model (model 2) was a similarly structured linear mixed model with the reads (square rooted) as the response variable, and all the above fixed and random effects (except the reads themselves). Both models were run in the subset DS3 (Table 1) of individuals, with $n=101$ and their results are presented in Table 3.

Finally, we used binomial proportion tests to examine potential differences in the social context of unsuccessful solving attempts versus successful events, as well as the proportion of positive versus negative social interactions in the cases where multiple mice were present.

Problem-solving in the arenas

First, we examined whether the individuals were equally likely to eat both the familiar and novel control, by performing a proportion test with the number of animals that ate each control. Then we built three models to understand the relationship between the measured behavioural traits and problem-solving performance in the arenas, using the subset DS4 (Table 2) of $n=50$ individuals.

Initially, we examined whether the extreme performers (16 high- and 16 low-performers for each behavioural trait) in the distance covered and the time in the centre of the Open field differed in their likelihood to solve problems in the arenas. For this, we built two generalised linear models (one for animals selected based on exploration; model SM1, and one for animals selected on time in the centre, model SM2), assuming a binomial distribution. The likelihood to solve was our response variable, the grouping based on the performance in the Open field (high or low exploration, time in centre respectively), the sex, and generation as explanatory variables. The output of these models is reported in the Supporting information.

In addition, we ran a generalised linear model with a binomial distribution (model 3), with the likelihood to solve (solvers or non-solvers) as the response variable, the actual performance in the first trial of the Open field, i.e. the distance (in meters) and the time in the centre (in seconds, square-rooted), the generation (founding generation or born in the enclosures) and the sex as fixed effects, using all the individuals that participated in the arenas. The output of this model is presented in Table 3.

Comparison of performance under the two conditions

We used Wilcoxon tests to detect differences in the problem-solving performance (number of problems solved) of males and females in each of the two conditions (Supporting information). This test was selected due to the heavily zero-inflated distribution of values (cf. Supporting information for more details). We also performed proportion tests to compare the number of individuals that visited, interacted and solved the problem-solving setups in each of the two conditions.

Results

Repeatability

We found significant repeatability estimates for both measurements of the Open field test (distance covered: $R=0.35$, $SE=0.11$, $p_{[LRT]}=0.002$, $p_{[permutation]}=0.003$; time spent in the centre: $R=0.2$, $SE=0.12$, $p_{[LRT]}=0.04$, $p_{[permutation]}=0.07$). Despite the marginal permutation p-values in the time spent in the centre, we used the variable for further analyses. The distance covered decreased in the second trial ($\beta=-7.42$, $SE=\pm 1.34$, $p<0.001$), whereas no such effect was found for the time in the centre ($\beta=0.1$, $SE=\pm 0.23$, $p=0.68$).

Semi-natural enclosures

Overall, $n=121$ individuals could potentially participate in the experiment; of those, $n=113$ (93.4%) visited the problem-solving boxes, $n=106$ (87.6%) interacted with a setup at least once, and $n=26$ (21.4%) solved at least one problem (mean $\pm$ SE: 3.88 ± 1.3 problems; Fig. 2a). Individuals were consistent in their frequency to visit across the four setups ($R=0.33$, $SE=\pm 0.06$, $p_{[LRT]}<0.001$, $p_{[permutation]}=0.001$). Individual solving performance exhibited low, but significant repeatability among setups ($R=0.137$, $SE=0.178$, $p_{[LRT]}=0.002$, $p_{[permutation]}=0.01$). The results of model 1 revealed that neither behavioural trait (distance covered or time in the centre of the Open field) had an effect on whether an individual would be a problem solver, and this was the case with generation (founding or born in the enclosures) or sex. However, individuals that visited the problem-solving boxes were more likely to solve them (see Fig. 3a and for the detailed results of this model, see Table 3 – model 1). Shyer individuals (lower time in the centre of the Open field) visited

Table 2. Social context of successful and unsuccessful problem-solving attempts, by recording the number of cases where multiple individuals were present inside the problem-solving box while a mouse was either unsuccessfully attempting or successfully solving a setup. Successful attempts represent all solving events recorded in the semi-natural enclosures, while unsuccessful attempts were randomly sampled through all setups in both enclosures.

	No. of observations	No. of observations where multiple individuals were present	No. of negative social interactions	No. of neutral social interactions	No. of positive social interactions
Unsuccessful attempts	100	12 (12%)	3 (3%)	7 (7%)	2 (2%)
	(randomly selected)				
Successful attempts	87	6 (7%)	4 (5%)	2 (2%)	0 (0%)

Table 3. Synthetic model outputs of the three main models used. Models used a different subset of data (DS3: n= 101 individuals with behavioural measurements, DS4: n = 50 individuals who were tested in the arenas, see). The type of model (GLMM or LMM) is also presented.

Predictors	Likelihood to solve (enclosures)				Number of reads (sqrt)				Likelihood to solve (arenas)			
	Log-Odds	SE	CI	p	Estimates	SE	CI	p	Log-Odds	SE	CI	p
(Intercept)	-2.67	1.07	-4.77 – -0.58	0.012	12.07	1.99	8.15 – 15.98	<0.001	-0.78	1.07	-2.88 – 1.31	0.465
distance (m)	0.001	0.03	-0.05 – 0.05	0.965	0.08	0.05	-0.01 – 0.17	0.085	0.05	0.03	-0.02 – 0.11	0.147
time in centre (sqrt)	-0.24	0.22	-0.67 – 0.18	0.266	-0.87	0.32	-1.49 – -0.24	0.007	-0.23	0.24	-0.70 – 0.23	0.326
generation	-0.76	0.55	-1.83 – 0.31	0.164	4.39	1.75	0.94 – 7.84	0.013	1.07	0.90	-0.70 – 2.83	0.237
sex [male]	0.35	0.50	-0.62 – 1.33	0.479	-2.24	0.92	-4.04 – -0.43	0.015	1.19	0.82	-0.42 – 2.79	0.147
reads [sqrt]	0.14	0.03	0.08 – 0.21	<0.001								
setup [setup2]	-2.14	0.63	-3.37 – -0.91	0.001	3.16	1.04	1.11 – 5.21	0.003				
setup [setup3]	-1.29	0.58	-2.42 – -0.16	0.026	-1.91	1.04	-3.95 – 0.13	0.067				
setup [setup4]	-1.78	0.64	-3.04 – -0.53	0.005	-1.06	1.06	-3.14 – 1.03	0.320				
Random effects												
σ^2	3.29				41.23				3.29			
τ_{00}	1.13	idshort:mother			2.06	idshort:mother			0.28	mother		
ICC	0.00	mother			12.55	mother			0.08			
N	101	idshort			0.26				23	mother		
Observations	28	mother			101	idshort			50			
Marginal R ² /	335				28	mother			0.138/			
conditional R ²	NA				335				0.206			

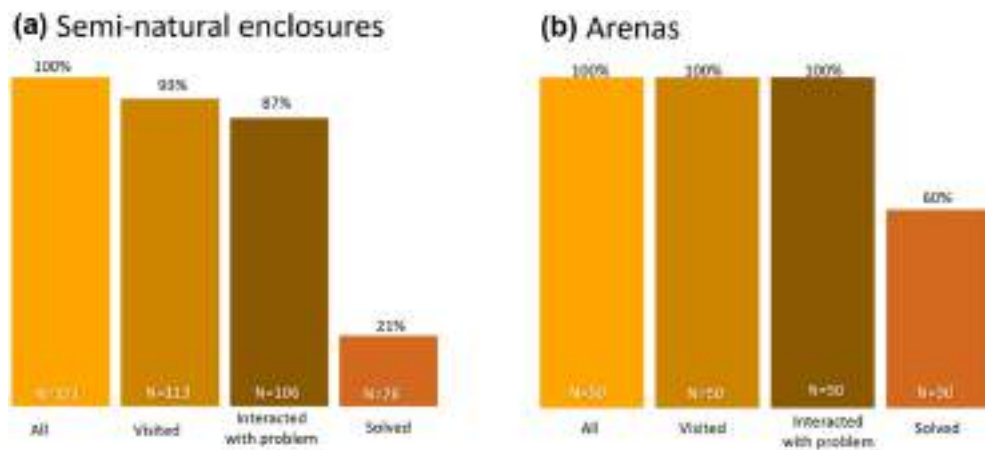


Figure 2. Barplot representing the number of individuals who successfully completed each step of solving a problem, with the seminatural enclosures on the left and the arenas on the right. The bars represent, from left to right: Individuals that participated in each phase of the experiment ('All'), individuals that visited the test box in the seminatural enclosures or exited the cage into the arenas ('Visited'), individuals that interacted with at least one problem ('Interacted with problem') and individuals that solved at least one problem ('Solved').

the problem-solving boxes more (Fig. 3b). Younger mice also visited the problems more, and so did females. There was no effect of the distance in the Open field (for the detailed results of this model see Table 3 – model 2).

When examining the social context of successful ($n=87$) versus unsuccessful ($n=100$, randomly sampled) solving attempts, the attempting/solving mice were predominantly alone in the box, in both cases (81/87 solving events; 93% and 88/100

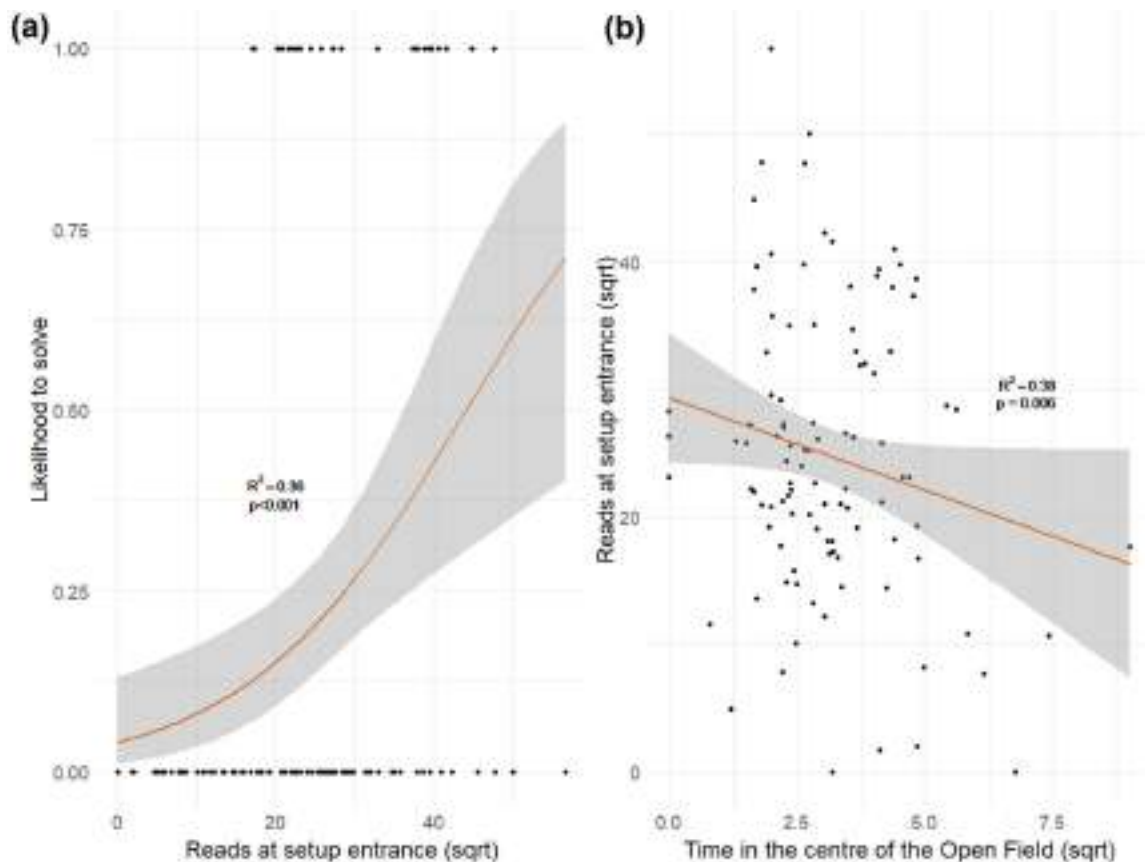


Figure 3. (a) Differences in the number of reads from the RFID antennas (square-rooted) between solvers ($n=26$) and non-solvers ($n=93$). (b) Linear regression between the time in the centre ($n=61$) in the Open field and the reads from the RFID antennas in the problem-solving boxes (both variables are squarerooted). Shading represents 95% confidence intervals in both plots.

attempts; 88%). Even when multiple mice were present in the box, it was usually pairs (17/18 cases, in one case there were three mice in the box), and their interactions were mostly either neutral (merely being present together, no socio-positive or aggressive interactions, 9/18 cases), or negative (aggressive interactions, mostly chasing, 7/18 cases), with a few being positive (mutual interest in the setup and/or preceding of allogrooming and similar behaviours, 2/18 cases). Binomial proportion tests revealed that the rates at which mice are alone when attempting but failing is not significantly different than when they solve successfully ($\chi^2=0.87$, $df=1$, $p=0.35$), and the proportion of negative social interactions is not significantly different than the proportion of positive social interactions ($\chi^2=2.37$, $df=1$, $p=0.12$). For a detailed report of the social context and types of interactions see [Table 2](#).

Arenas

In the arenas, $n=50$ individuals could participate in the experiment and, of those, all exited the cage and interacted with at least one setup, and $n=30$ (60%) solved at least one problem (average number of problems solved $\pm$ SE: 1.93 ± 0.2 ; [Fig. 2b](#)). Similar proportions of individuals ate the familiar (86%) and unfamiliar (74%) controls ($\chi^2_{(1)}=1.56$, $p=0.21$). We found that the likelihood to solve in the arenas was not affected by either behavioural trait, generation or sex (for the detailed results of this model see [Table 3](#) – model 3).

Comparing the two conditions

Overall, $n=17$ (34%) individuals did not solve any problem in either condition, $n=22$ (44%) individuals solved a problem in only one of the two conditions (three individuals only in the enclosures, 19 individuals only in the arenas), and $n=11$ (22%) individuals solved at least one problem in both conditions. Performance in one condition did not predict the performance in the other ($R=0.22$, $SE=0.18$, $p[LRT]=0.08$, $p[permutation]=0.04$). In the seminatural enclosures, a Wilcoxon's rank sum test showed no difference in the number of problems solved by either sex (Wilcoxon's rank sum test, $W=1825.5$, $p=0.99$). However, males in the arenas solved more problems than females ($W=389$, $p=0.001$, males: mean = 1.94, $SE=\pm 0.38$; females: mean = 0.79, $SE=\pm 0.16$).

Discussion

In the current study, we examined the performance of individual wild house mice in two conditions: first, free-living in large social groups in seminatural enclosures that closely approximate the conditions in the wild; and later in a controlled experiment, similar to how wild animals are commonly tested in captivity.

No direct impact of behavioural type on problem-solving, but there is an effect on steps towards problem-solving

Our results revealed that the behavioural traits exploration and risk-taking strategy are not directly linked to problem-solving

performance in wild house mice, regardless of the condition in which the mice were tested. However, when examining the steps preceding eventual solving, shyer individuals had improved chances of encountering the setups.

Boldness tended to affect the steps that led to the possibility of solving a setup, if not of finding the solution in itself. In the semi-natural enclosures, shyer individuals were visiting the setups more often, i.e. they were more likely to get recorded at the entrance of the problem-solving boxes. This finding contradicts the prediction that bolder and faster-exploring animals are more likely to encounter novel stimuli ([Sih and Del Giudice 2012](#)); in our case, we showed that it is the shyer individuals that are overall more likely to visit an 'out of the ordinary' stimulus, even when it has – presumably – lost its novelty. However, our results are in line with the 'bad competitor' hypothesis ([Laland and Reader 1999](#), [Reader and Laland 2003](#)). In mice, bold versus shy mice have been shown to use different strategies to resolve agonistic encounters ([Benus et al. 1992](#), [Koolhaas et al. 2010](#)). While bold males engage in aggressive interactions easily and frequently, shy males are more flexible and resort to aggressive behaviour only when absolutely necessary ([Koolhaas et al. 2010](#)). Thus, shy individuals might spend more time investigating low-competition resources.

We saw that shy individuals consistently visited the problem-solving boxes every time they were presented to them, making them more likely to reach the first step towards solving a problem, i.e. encountering the problem itself. Under the notion that problem-solving does not (solely) depend on the cognitive or motor abilities of the animal ([van Horik and Madden 2016](#), [Van Horik et al. 2017](#), [Rowell et al. 2021](#)), having more chances to approach and interact with a problem could lead to a successful solution. Our findings indicated the specific influence that individual variation might have on problem-solving: specific behavioural traits can influence which animals come across opportunities, but then it is other factors (i.e. cognitive ability, persistence, motor skills, time availability, perceived risk or even just chance) that eventually determine whether an individual successfully problem-solves. Furthermore, the behavioural traits we measured showed moderate repeatability, indicating consistent individual, i.e. personality differences in making use of problem-solving opportunities.

Consistent with our results that do not show a direct link between exploration and problem-solving success, other studies reported a similar lack of this hypothesised link ([Benson-Amram et al. 2013](#), [Guenther and Brust 2017](#), [Amici et al. 2019](#), [Rowell et al. 2021](#)). For example, in two species of mouse lemurs *Microcebus* spp., exploration did not affect their problem-solving performance ([Henke-von der Malsburg and Fichtel 2018](#)). Similarly, in chimango caracaras *Milvago caracara*, object exploration did not affect problem-solving performance in both adult and juvenile individuals ([Biondi et al. 2010](#)). Our results add to this body of evidence that exploration is not ubiquitously correlated with problem-solving performance.

In neophilic species like house mice, the potential risks of novel opportunities associated with problem-solving might not

have the same connotation as they have for neophobic species. In birds, for example, responses to novelty correlate with performance in novel foraging tasks (Webster and Lefebvre 2001, Reader 2003, Biondi et al. 2010). In a meta-analysis, Griffin and Guez (2014) found that, in multiple bird species, neophobia affects whether animals would interact with a novel situation. Additionally, another meta-analysis by Amici et al. (2019) that examined studies (mostly on birds) showed that neophilia and exploration had positive effects on innovation. On the one hand, different behavioural traits might affect innovation in neophobic and neophilic species. On the other hand, even if the same behavioural traits impact innovation in all species, measuring them clearly in neophobic species might be difficult as their aversion to novelty would dominate any encounter with a novel situation or problem. As such, by testing neophilic species whose risk avoidance does not impede any aspect of problem-solving, we can get valuable insights into the role of the other behavioural traits on the emergence and evolution of problem-solving (Amici et al. 2019).

In a complex socioecological context, such as the semi-natural enclosures of our experiment, each individual can only allocate a certain amount of time and effort to interacting with the problems we present. In the arenas, however, many of these constraints are lifted and the problem-solving opportunities are easily accessible to all individuals. By comparing how individuals performed across the two conditions, we can get a better understanding of the specific impact of the testing condition on the problem-solving measurements that we observe.

Different conditions cause differences in problem-solving: potential reasons and implications on how we measure problem-solving

Our findings revealed that the likelihood to solve problems was higher in the controlled conditions. When tested alone in the arenas, more individuals solved problems when compared to the semi-natural enclosures. Additionally, most individuals did not perform consistently across the two conditions, indicating that different conditions might not affect the performance of individuals in the same way.

Testing animals in large, socially complex groups can be very different from testing them alone. On the one hand, the possible anxiety stemming from isolation could interfere with the cognitive processes, or setups could fail to elicit interest in the subject in absence of conspecifics providing stimulus enhancement (Avarguès-Weber and Chittka 2014, Brandão et al. 2015, Lambert and Guillette 2021). On the other, individuals in groups can be distracted or compete with others (Benson-Amram et al. 2013), which has been hypothesised as a reason causing worse observed performance in problem-solving tests (Boere 2001, Krashennikova and Schneider 2014). Testing individuals alone for brief periods of time offers greater control and can provide better insight into the abilities of each individual, while also removing confounding effects from external stimuli (Amici et al. 2019).

Our results show the importance of testing the same animals in multiple conditions, as we saw that the problem-solving

performance of our mice was different in each of the two conditions. In the controlled experiment, 60% of individuals (30/50) solved at least one problem, which is similar to previous results in the same species (Vrbanec et al. 2021). In the semi-natural enclosures, however, we see a large contrast with 21.4% of individuals (26/121) successfully solving at least one setup. Our results are aligned with previous findings where captive individuals tested alone are more successful problem-solvers than wild individuals, typically tested in larger social groups. For example, more than 70% of captive and isolated spotted hyenas *Crocuta crocuta* were able to solve a problem-solving task at least once – a stark difference from the 14.5% of wild individuals (Benson-Amram et al. 2013). Captive hyenas also showed higher persistence. The presence of conspecifics has been hypothesised to distract the focal animal, thus reducing their performance in problem-solving tasks. On the one hand, wild common marmosets *Callithrix jacchus* made more errors in a problem-solving task due to distractions from conspecifics (Halsey et al. 2006), but on the other hand, there was no such effect in a study on range-winged amazons *Amazona amazonica* (Krashennikova and Schneider 2014). Thus, simply the presence of conspecifics may not necessarily explain these differences in problem-solving performance across the two conditions. In our case, we found that mice mostly attempted and solved problems alone, with conspecifics being very rarely present for those attempts. This highlights both the absence of direct competition (i.e. some individuals being actively excluded from the problem-solving boxes) but also the absence of opportunities for social learning through imitation to occur.

The two conditions that we focused on represent different experimental designs that subject the individuals to different challenges, and understanding these challenges might help us reveal why both individual and overall performance vary starkly between the two. The semi-natural enclosures offer a varied, dense socioecological environment where the mice live complex lives. While these conditions were not harsh, as food was available ad libitum and nesting opportunities were plentiful, individuals would still compete for space, potential mates, and defend their territories and offspring (König et al. 2012). All these factors distract the individuals from the task we expect them to perform, simply because they need to allocate the majority of their time performing these other functions (Kummer and Goodall 1985). The ‘excess of energy’ hypothesis (Kummer and Goodall 1985) suggests that individuals would solve more problems when they have more time and energy to devote to innovation. Indeed, in the controlled conditions, all individuals were alone for many hours, and there were virtually no external stimuli outside the problem-solving setups, so they could allocate all their time and energy into solving the problems. This is potentially one of the factors that allowed more individuals to problem-solve when tested in the arenas. Also, as we tested the same individuals in both conditions, there were no differences in their intrinsic abilities or affinity with the setups (i.e. higher familiarity with man-made objects for captive individuals, see van de Waal and Bshary 2011, Benson-Amram et al. 2013).

Therefore, as the experimental conditions cause this difference in performance, our findings can help distinguish the specific areas that testing in isolation/controlled conditions versus in a complex social group in (semi-)natural conditions really examines (Tomasello and Call 2011, Amici et al. 2019).

Each individual's performance in one condition is not correlated to their performance in the other condition with only 56% of individuals being equally likely to solve successfully in both conditions. A characteristic example is the case of the most prolific solver in the semi-natural enclosures; this individual successfully solved 34 problems while in the enclosures – almost nine times more than the average solver. However, the same individual solved only two of the four problems available in the arena. In contrast, multiple individuals that solved no problems in the enclosures eventually solved multiple problems while in the arenas. It becomes evident that, for most individuals, the conclusions drawn would have been starkly different had we investigated them under only one of the two conditions.

Our findings point towards the conclusion that different conditions allow us to measure different aspects of problem-solving. Studying individuals in controlled conditions (e.g. in the laboratory, socially isolated) allows us to determine whether an individual can solve problems (i.e. we measure their problem-solving ability). Studying individuals in uncontrolled conditions (e.g. in the wild or in semi-natural enclosures) allows us to observe whether an individual will problem-solve (i.e. we measure their problem-solving propensity).

To ultimately understand the evolutionary importance of problem-solving and its effects on individuals, it is important to additionally examine its fitness consequences. Problem-solving may help individuals achieve higher fitness either through direct benefits (accessing higher quality, quantity or novel food) or indirect (mate choice preference of solvers over non-solvers, high social status of solvers), and understanding if and how this happens is paramount to further our understanding of the evolution of problem-solving.

Conclusions

This study provides evidence that behavioural traits may influence problem-solving performance in wild house mice in an indirect way, affecting non-cognitive aspects of innovation, such as improving an individual's chances to discover these problems in their natural environment. In different conditions, however, each individual performed differently and non-consistently, revealing that researchers should be careful when transferring findings from the laboratory to the wild and vice versa. We suggest that inter-individual differences should be taken into account in a species-specific manner, as well as examining the ecological validity of the setting that problem-solving is observed.

Acknowledgements – We would like to thank Milan Jovicic for caring for the mice throughout their lives in the enclosures and in the cages,

Helena Höhnke for helping with the arena experiments, Ekaterina Gorshkova for helping with the video analysis, Karem Lopez Hervas and Fragkiskos Darmis for helpful comments during data analysis. We would also like to thank Denis Réale whose comments greatly improved our work.

Funding – This work was funded by the German Science Foundation (DFG – Deutschen Forschungsgemeinschaft – project number 493860801) with a joint grant to VM (MA 9757/2-1) and AG (GU 1665/5-1).

Permits – All experiments conformed to ASAB/ABS guidelines for animal experiments and were carried out under the licence V244 - 57612/2022(43-4/19) of the Ministerium für Energiewende, Landwirtschaftliche Räume und Umwelt, Kiel. Keeping and breeding of mice was approved and is regularly controlled by the Veterinäramt Plön under permit: 1401-144/PLO-004697.

Author contributions

Alexandros Vezyrakis: Conceptualization (equal), Data curation (lead), Formal analysis (lead), Methodology (equal), Project administration (equal), Validation (equal), Visualization (equal), Writing - original draft (lead), Writing - review and editing (lead). **Anja Guenther:** Conceptualization (equal), Funding acquisition (equal), Methodology (equal), Project administration (equal), Supervision (equal), Writing - review and editing (equal). **Valeria Mazza:** Conceptualization (equal), Funding acquisition (equal), Methodology (equal), Project administration (equal), Supervision (equal), Writing - review and editing (equal).

Data availability statement

Data are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.fbg79d2g> (Vezyrakis et al. 2024).

Supporting information

The Supporting information associated with this article is available with the online version.

References

- Amici, F., Widdig, A., Lehmann, J. and Majolo, B. 2019. A meta-analysis of interindividual differences in innovation. – *Anim. Behav.* 155: 257–268.
- Aplin, L. M., Sheldon, B. C. and Morand-Ferron, J. 2013. Milk bottles revisited: social learning and individual variation in the blue tit, *Cyanistes caeruleus*. – *Anim. Behav.* 85: 1225–1232.
- Avarguès-Weber, A. and Chittka, L. 2014. Local enhancement or stimulus enhancement? Bumblebee social learning results in a specific pattern of flower preference. – *Anim. Behav.* 97: 185–191.
- Bakdash, J. Z. and Marusich, L. R. 2017. Repeated measures correlation. – *Front. Psychol.* 8: 456.
- Benson-Amram, S. and Holekamp, K. E. 2012. Innovative problem solving by wild spotted hyenas. – *Proc. R. Soc. B* 279: 4087–4095.
- Benson-Amram, S., Weldele, M. L. and Holekamp, K. E. 2013. A comparison of innovative problem-solving abilities between wild and captive spotted hyenas, *Crocuta crocuta*. – *Anim. Behav.* 85: 349–356.

- Benus, R. F., Koolhaas, J. M. and Van Oortmerssen, G. A. 1992. Individual strategies of aggressive and non-aggressive male mice in encounters with trained aggressive residents. – *Anim. Behav.* 43: 531–540.
- Biondi, L. M., García, G. O., Bó, M. S. and Vassallo, A. I. 2010. Social learning in the caracara chimango, *Milvago chimango* (Aves: Falconiformes): an age comparison. – *Ethology* 116: 722–735.
- Boere, V. 2001. Environmental enrichment for neotropical primates in captivity. – *Cien. Rural* 31: 543–551.
- Boogert, N. J., Reader, S. M. and Laland, K. N. 2006. The relation between social rank, neophobia and individual learning in starlings. – *Anim. Behav.* 72: 1229–1239.
- Boogert, N. J., Madden, J. R., Morand-Ferron, J. and Thornton, A. 2018. Measuring and understanding individual differences in cognition. – *Philos. Trans. R. Soc. B* 373: 20170280.
- Brandão, M. L., Braithwaite, V. A. and Gonçalves-de-Freitas, E. 2015. Isolation impairs cognition in a social fish. – *Appl. Anim. Behav. Sci.* 171: 204–210.
- Carere, C. and Locurto, C. 2011. Interaction between animal personality and animal cognition. – *Curr. Zool.* 57: 491–498.
- Day, R. L., Laland, K. N. and Odling-Smee, F. J. 2003. Rethinking adaptation: the niche-construction perspective. – *Perspect. Biol. Med.* 46: 80–95.
- Delacoux, M. and Guenther, A. 2023. Stressfulness of the design influences consistency of cognitive measures and their correlation with animal personality traits in wild mice (*Mus musculus*). – *Anim. Cogn.* 26: 997–1009.
- Dougherty, L. R. and Guillelte, L. M. 2018. Linking personality and cognition: a meta-analysis. – *Philos. Trans. R. Soc. B* 373: 20170282.
- Drea, C. M. and Carter, A. N. 2009. Cooperative problem solving in a social carnivore. – *Anim. Behav.* 78: 967–977.
- Fisher, J. 1949. The opening of milkbottles by birds. – *Brit. Birds* 42: 347–357.
- Fox, R. J., Donelson, J. M., Schunter, C., Ravasi, T. and Gaitán-Espitia, J. D. 2019. Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. – *Philos. Trans. R. Soc. B* 374: 20180174.
- Friard, O. and Gamba, M. 2016. BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. – *Methods Ecol. Evol.* 7: 1325–1330.
- Gould, T. D., Dao, D. T. and Kovacsics, C. E. 2009. The open field test. – In Gould, T. D. (ed.), *Mood and anxiety related phenotypes in mice: characterization using behavioral tests*. Humana Press, pp. 1–20.
- Greenberg, R. 2003. The role of neophobia and neophilia in the development of innovative behaviour of birds. – In: Reader, S. M. and Laland, K. N. (eds), *Animal innovation*. Oxford Univ. Press.
- Greenberg, R. and Mettke-Hofmann, C. 2001. Ecological aspects of neophobia and neophilia in birds. – In: Nolan, V. and Thompson, C. F. (eds), *Current ornithology*. Springer, pp. 119–178.
- Griffin, A. S. and Guez, D. 2014. Innovation and problem solving: a review of common mechanisms. – *Behav. Processes* 109: 121–134.
- Griffin, A. S., Guillelte, L. M. and Healy, S. D. 2015. Cognition and personality: an analysis of an emerging field. – *Trends Ecol. Evol.* 30: 207–214.
- Guenther, A. and Brust, V. 2017. Individual consistency in multiple cognitive performance: behavioural versus cognitive syndromes. – *Anim. Behav.* 130: 119–131.
- Hall, C. and Ballachey, E. L. 1932. A study of the rat's behavior in a field. A contribution to method in comparative psychology. – Univ. of California publications in psychology.
- Halsey, L. G., Bezerra, B. M. and Souto, A. S. 2006. Can wild common marmosets (*Callithrix jacchus*) solve the parallel strings task? – *Anim. Cogn.* 9: 229–233.
- Hartig, F. and Lohse, L. 2022. Dharma: Residual diagnostics for hierarchical (multi-level/mixed) regression models 0.4.6. – <https://cran.r-project.org/web/packages/DHARMA/>.
- Henke-von der Malsburg, J. and Fichtel, C. 2018. Are generalists more innovative than specialists? A comparison of innovative abilities in two wild sympatric mouse lemur species. – *R. Soc. Open Sci.* 5: 180480.
- Kaiser, M. I. and Müller, C. 2021. What is an animal personality? – *Biol. Philos.* 36: 1.
- Klump, B. C., Martin, J. M., Wild, S., Hörsch, J. K., Major, R. E. and Aplin, L. M. 2021. Innovation and geographic spread of a complex foraging culture in an urban parrot. – *Science* 373: 456–460.
- Klump, B. C., Major, R. E., Farine, D. R., Martin, J. M. and Aplin, L. M. 2022. Is bin-opening in cockatoos leading to an innovation arms race with humans? – *Curr. Biol.* 32: R910–R911.
- König, B., Lindholm, A. K. and Pialek, J. 2012. The complex social environment of female house mice (*Mus domesticus*). – In: Macholan, M., Baird, S. J. E. and Munclinger, P. (eds), *Evolution of the house mouse*. Cambridge Univ. Press, pp. 114–134.
- Koolhaas, J. M., de Boer, S. F., Coppens, C. M. and Buwalda, B. 2010. Neuroendocrinology of coping styles: towards understanding the biology of individual variation. – *Front. Neuroendocrinol.* 31: 307–321.
- Krackow, S. 2003. Motivational and heritable determinants of dispersal latency in wild male house mice (*Mus musculus musculus*). – *Ethology* 109: 671–689.
- Krasheninnikova, A. and Schneider, J. M. 2014. Testing problem-solving capacities: differences between individual testing and social group setting. – *Anim. Cogn.* 17: 1227–1232.
- Krebs, R., Linnenbrink, M. and Guenther, A. 2019. Validating standardised personality tests under semi-natural conditions in wild house mice (*Mus musculus domesticus*). – *Ethology* 125: 761–773.
- Kummer, H. and Goodall, J. 1985. Conditions of innovative behaviour in primates. – *Philos. Trans. R. Soc. B* 308: 203–214.
- Laland, K. N. and Reader, S. M. 1999. Foraging innovation in the guppy. – *Anim. Behav.* 57: 331–340.
- Lambert, C. T. and Guillelte, L. M. 2021. The impact of environmental and social factors on learning abilities: a meta-analysis. – *Biol. Rev.* 96: 2871–2889.
- Mazza, V. and Guenther, A. 2021. City mice and country mice: innovative problem solving in rural and urban noncommensal rodents. – *Anim. Behav.* 172: 197–210.
- Morand-Ferron, J., Cole, E. F., Rawles, J. E. C. and Quinn, J. L. 2011. Who are the innovators? A field experiment with two passerine species. – *Behav. Ecol.* 22: 1241–1248.
- Nakagawa, S. and Schielzeth, H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. – *Biol. Rev.* 85: 935–956.
- Nicolakakis, N., Sol, D. and Lefebvre, L. 2003. Behavioural flexibility predicts species richness in birds, but not extinction risk. – *Anim. Behav.* 65: 445–452.
- O'Hara, M., Mioduszevska, B., von Bayern, A., Auersperg, A., Bugnyar, T., Wilkinson, A., Huber, L. and Gajdon, G. K. 2017.

- The temporal dependence of exploration on neotic style in birds. – *Sci. Rep.* 7: 1.
- Overington, S. E., Cauchard, L., Côté, K.-A. and Lefebvre, L. 2011. Innovative foraging behaviour in birds: what characterizes an innovator? – *Behav. Processes* 87: 274–285.
- Perals, D., Griffin, A. S., Bartomeus, I. and Sol, D. 2017. Revisiting the open-field test: what does it really tell us about animal personality? – *Anim. Behav.* 123: 69–79.
- Prabh, N., Linnenbrink, M., Jovicic, M. and Guenther, A. 2023. Fast adjustment of pace-of-life and risk-taking to changes in food quality by altered gene expression in house mice. – *Ecol. Lett.* 26: 99–110.
- Ramsey, G., Bastian, M. L. and Van Schaik, C. 2007. Animal innovation defined and operationalized. – *Behav. Brain Sci.* 30: 393–407.
- Reader, S. 2003. Innovation and social learning: individual variation and brain evolution. – *Anim. Biol.* 53: 147–158.
- Reader, S. M. and Laland, K. N. 2003. Animal innovation: an introduction. – In: Reader, S. M. and Laland, K. N. (eds), *Animal innovation*. Oxford Univ. Press, pp. 3–36.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T. and Dingemanse, N. J. 2007. Integrating animal temperament within ecology and evolution. – *Biol. Rev.* 82: 291–318.
- Rowell, M. K., Pillay, N. and Rymer, T. L. 2021. Problem solving in animals: proposal for an ontogenetic perspective. – *Animals* 11: 3.
- Seed, A. and Byrne, R. 2010. Animal tool-use. – *Curr. Biol.* 20: R1032–R1039.
- Sih, A. and Del Giudice, M. 2012. Linking behavioural syndromes and cognition: a behavioural ecology perspective. – *Philos. Trans. R. Soc. B* 367: 2762–2772.
- Sol, D. 2003. Behavioural innovation: a neglected issue in the ecological and evolutionary literature? – In: Reader, S. M. and Laland, K. N. (eds), *Animal innovation*. Oxford Univ. Press.
- Stamps, J. and Groothuis, T. G. G. 2010. The development of animal personality: relevance, concepts and perspectives. – *Biol. Rev.* 85: 301–325.
- Stoffel, M. A., Nakagawa, S. and Schielzeth, H. 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. – *Methods Ecol. Evol.* 8: 1639–1644.
- Tebich, S., Griffin, A. S., Peschl, M. F. and Sterelny, K. 2016. From mechanisms to function: an integrated framework of animal innovation. – *Philos. Trans. R. Soc. B* 371: 20150195.
- Thornton, A. and Samson, J. 2012. Innovative problem solving in wild meerkats. – *Anim. Behav.* 83: 1459–1468.
- Tomasello, M. and Call, J. 2011. Methodological challenges in the study of primate cognition. – *Science* 334: 1227–1228.
- van de Waal, E. and Bshary, R. 2011. Contact with human facilities appears to enhance technical skills in wild vervet monkeys (*Chlorocebus aethiops*). – *Folia Primatol.* 81: 282–291.
- van Horik, J. O. and Madden, J. R. 2016. A problem with problem solving: motivational traits, but not cognition, predict success on novel operant foraging tasks. – *Anim. Behav.* 114: 189–198.
- Van Horik, J. O., Langley, E. J. G., Whiteside, M. A. and Madden, J. R. 2017. Differential participation in cognitive tests is driven by personality, sex, body condition and experience. – *Behav. Processes* 134: 22–30.
- Vezyrakis, A., Guenther, A. and Mazza, V. 2024. Data from: Effects of behavioural types on the problem-solving performance of wild house mice under controlled and semi-natural conditions. – Dryad Digital Repository, <https://doi.org/10.5061/dryad.fbb-g79d2g>.
- Vrbanc, L., Matijević, V. and Guenther, A. 2021. Enhanced problem-solving ability as an adaptation to urban environments in house mice. – *Proc. R. Soc. B* 288: 20202504.
- Webster, S. J. and Lefebvre, L. 2001. Problem solving and neophobia in a columbiform–passeriform assemblage in Barbados. – *Anim. Behav.* 62: 23–32.
- Wright, T. F., Eberhard, J. R., Hobson, E. A., Avery, M. L. and Russello, M. A. 2010. Behavioral flexibility and species invasions: the adaptive flexibility hypothesis. – *Ethol. Ecol. Evol.* 22: 393–404.
- Young, J. K., Touzot, L. and Brummer, S. P. 2019. Persistence and conspecific observations improve problem-solving abilities of coyotes. – *PLoS One* 14: e0218778.