THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

Enhanced problem-solving ability as an adaptation to urban environments in house mice

Lara Vrbanec, Vanja Matijević and Anja Guenther

Article citation details
Proc. R. Soc. B **288**: 20202504.
http://dx.doi.org/10.1098/rspb.2020.2504

Review timeline

Original submission:	8 October 2020
1st revised submission:	9 December 2020
2nd revised submission:	13 January 2021
Final acceptance:	25 January 2021

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPB-2020-2504.R0 (Original submission)

Review form: Reviewer 1

Recommendation
Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Good

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
Yes

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
There is growing evidence that urban animals tend to be better at finding behavioral solutions to new problems than non-urban animals, but it is unclear whether these differences reflect adaptive evolution, plasticity or habitat-matching. This study addresses by gap by comparing problem solving ability of house mouse from three subspecies, which are presumed to vary in the period they have lived in urban areas as human commensals (from ~3000 years to ~11000 years). To tease apart plastic differences, individuals were allowed to breed for several generations in the laboratory before being tested for problem-solving performance.

I read the MS with great interest, as there is little (if any) evidence of adaptive divergence caused by urbanization. While I think these results are potentially important, I have a number of concerns regarding their evolutionary interpretations.

General issues

I found that the experimental assays were well-designed and elegant. The authors take good care to properly habituate individuals, present several problems and novel objects to assess individual's consistency and measure behaviors to interpret the results. However, the theoretical framework is poorly developed. Why should urban animals be better at problem-solving? Although the authors cite literature on the importance of innovation and problem-solving, these are based on the general importance of being plastic rather than a discussion of the role of problem-solving to cope with the challenges or urban life and how these can affect mouse. I would make it clearer that living close to humans may more frequently exposes individuals to novel problems that may be solved by means of behavioral constructs. I have also a summary sentence citing previous evidence thay urban animals are better at problem solving than non-urban animals. Ignoring previous evidence may make the impression than the study is more novel than it is really. I also do not understand the logic of citing reviews discussing similar issues in differences places of the text; citations should be used more careful. I recommend the authors to read the recent book "Urban Evolutionary Biology" edited by Marta Szulkin, Jason Munshi-South and Anne Charmantier.

The authors claim that they have found adaptive divergence in problem-solving ability caused by urbanization, but I do not see where this conclusion comes from. First, the fact that individuals were allowed to breed for several generations in the laboratory may tease apart plastic differences between populations, yet do not demonstrate that differences have been cause by

natural selection. Other mechanisms like matching habitat choice or genetic drift may also create genetic differences across populations. Maternal effects are a less likely explanation as these effects should be diluted in a few generations. Second, it is unclear to me how the period of human commensalism has been estimated and to what extent ~3000 years are not enough to cause adaptive responses. Finally, even when this period is accurately assessed, the three subspecies diverged ~350-500 thousand years ago and hence may have been exposed to many different selective pressures in addition of urbanization. It is thus unclear whether urbanization has been the main driver of genetic differences in problem solving. Without evidence of selection differentials or gradients, the adaptive function of problem solving remain unclear.

Still, the finding that differences in problem-solving among subspecies may have a genetic basis is interesting. However, I miss here an analysis of heritability that allows to assess the extent to which this variation may respond to selection. If individuals were kept for several generations, why do not present data on the heritability of problem-solving performance and the correlates that may either inhibit or facilitate it?

I finally find surprising that problem-solving differences cannot be explained by exploratory tendency or inhibition. The analyses presented are centered on success/failure (instead of latency to solve the problem), which may have obscured these effects. I also recommend the authors to follow previous studies and use structural equation modeling to tease apart effects. This is because if two confounds are correlated between them because of causal links, including both as predictors in the model my block the effect of one of these variables. This can be avoided by defining causal scenarios and test them with SEM (Sol et al., 2012). Likewise, the underlying causes of problem-solving differences between subspecies may be better analyzed by SEM (Sol et al., 2011).

Specific issues:

Line 13: The term “evolutionary novel environments” looks weird,

Line 33: What do you mean by “adjust”?

Lines 36-40: This sentence is unclear. The present study also reports differences between traits without studying their adaptive significance. Moreover, the authors have not yet presented the idea that behavior matters.

Lines 49-50: This sentence may make the impression that evidence for differences in problem-solving between urban and non-urban animals is scarce, when it is not. The authors also miss all early evidence and focus on a few recent papers, suggesting that the issue has only recently been studied.

Line 52-53: Differential survival is selection. I think you refer to habitat-matching choice, where individuals with certain behaviors are more likely to settle in cities (see ref 22 in the main text).

Line 55: Reference is missing here.

Line 56: Why do you use “morphology” here if this is not mentioned again in the paragraph.

Line 56: Better” There is also little information regarding which morphological...

Line 64: The term “trial and error associative learning ability” is weird.

Lines 69-71: I do not recall that this paper discusses it.

Line 70: Not necessarily; matching habitat choice may also create genetic differences across populations.

Line 71: Is this a result?

Lines 82-84: This is a key assumption of the study, it is not enough to cite papers: 1) evidence to supports differential exposure to urbanization needs to be critically discussed; 2) whether differences in exposure are enough to generate divergence in problem solving also needs a better justification.

Line 91: Do you mean "spatial exploration"?

Line 104: Were they tested in pairs? Why?

Line 104: What is a type III cage?

Line 141: Interpreting the results of open-field test is not easy as some behaviors may reflect neophilia and others neophobia (Perals et al., 2017).

Line 164: The reference of the Package is missing.

Line 167: I do not see the logic of not using the same confounds here than in the analyses for setups with multiple solving strategies. And what about sex, why this was not included in the analysis? If sexes differ, maybe you are not observing the effect of natural selection but that of sexual selection.

Line 170: Where are the results of this comparison?

Lines 193-195: Did you test whether individuals that performed better in one setup also did so in the other setup? You could also report variation among individuals using a plot of the random factors.

Line 195: Why not simply "$P < 0.0001$". Moreover, you mention in the methods that these individuals were not used in the analyses.

Line 195: Do all analyses include the effect of confounding factors?

Fig. 1. Why do not add an arrow in the x-axis indicating the degree of exposure to urbanization. And why do not split the data among populations. Replication is important to convince the reader that the pattern is solid.

Line 204: What do you mean by third setup?

Lines 203-209: This section is a bit confusing, and not well integrated with the previous section. Help the reader interpreting these finding in the light of differences in problem-solving of the previous section.

Line 209: I'm lost here!

Line 224: How do you know that this is neophilia? And what about neophobia, which has been previously found to correlate with problem solving in urban animals (Sol et al., 2012)-

Line 225: What about the possibility that the lack of significance reflects insufficient power of the test?

Line 244: When was neophobia measured? Neophilia and neophobia are not opposed terms, they depend on different processes.

Line 244: Do the problems that mouse have to solve in the lab really relevant for solving problems in the wild?

Line 248: The first studies reporting differences have almost 10 years, so maybe evidence is not so new.

Line 257: I do not recall that these chapters provide evidence for fitness benefits; If I remember correctly, Reader et al. even found no relationship between innovativeness and environmental variability.

Lines 261-262: Not necessarily, as explained before.

Line 274: This conclusion is based on a cherry-picked number of papers, there are papers that show clear support for the effect of neophobic responses on problem-solving (see for example studies in common mynah). Others do not find support because the individuals have been allowed to habituate to the testing apparatus, which could be the present case. On the other hand, measuring problem-solving as failure/success reduces the probability to find effects of motivation or neophobia as an individual that solve the problem is considered successful regardless that it has done so in 1 or 10 min.

Perals, D., Griffin, A.S., Bartomeus, I. & Sol, D. 2017. Revisiting the open-field test: what does it really tell us about animal personality? Anim Behav 123: 69–79.

Sol, D., Griffin, A.S. & Bartomeus, I. 2012. Consumer and motor innovation in the common myna: The role of motivation and emotional responses. Anim Behav 83: 179–188.

Sol, D., Griffin, A.S., Bartomeus, I. & Boyce, H. 2011. Exploring or Avoiding Novel Food Resources? The Novelty Conflict in an Invasive Bird. PLoS One 6: e19535.

Review form: Reviewer 2

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Excellent

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Excellent

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
In their paper Vrbanec et al explore the problem solving abilities in three subspecies of house mice. The observed differences are interpreted as different adaptations to urban environments. Overall this is a very interesting and well written paper. I thoroughly enjoyed reading it and I think it provides and important contribution to the field. Below are some comments, suggestions for improvement and concerns.
1) Figure A2 shows some striking differences between the two populations of Mus musculus musculus (Austria vs Kazakhstan), whereby the Austrian population has the lowest likelihood to solve problems. Unless I missed it, this was not sufficiently emphasized in the Discussion. I think this should be carefully discussed as it clearly show how there's something else going on - which is not surprising given how many additional factors will likely interplay with the evolution of cognitive abilities.
2) Differences in the lenghth of commensalisms among the subspecies are indicated as the likely force shaping the observed differences. While I have no elements to dispute this, I would encourage more caution in interpreting results - it could be valuable to add a paragraph or so discussing potential factors that may also contribute to this? Would the kind of environment or type of urbanization contributed too?
1) More information should be provided about the origins of animals. I can see the ESM has some info, however it would be important to understand more such as if these were caught in cities vs villages, more information about trapping dates etc
2) While ESM 2 provides some information on the tests it would be very helpful to make some videos available so that the reader can better understand the dynamics of the tests.
3)along the same lines it would be good to provide images and videos of the neophilia test. With regards to this test I was confused by why you did not relocate the individual in the Type III (i.e. in a fashion similar to when conducting an open field test). What is the possible effect of time-to-become-aware of the novel environment (particularly important since latency to enter the environment is the key dependent variable)?
LL164 mixed effects model
4) While I am not a big fan of Likelihood ratio tests - I also acknowledge that such decisions are at the authors discretion and I think the effects reported appear to be sufficiently strong. I will suggest however to report some overall measure of fit such as squared R (which is possible to obtain through lme4). It would also be useful to include the parameter estimates (betas and SE).
Overall congratulations for an excellent and very interesting series of experiments!

Decision letter (RSPB-2020-2504.R0)

04-Nov-2020

Dear Dr Guenther:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors have raised some concerns with your manuscript and we would like to invite you to revise your manuscript to address them.

We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Associate Editor, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers. Please note that we cannot guarantee eventual acceptance of your manuscript at this stage.

To submit your revision please log into http://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions", click on "Create a Revision". Your manuscript number has been appended to denote a revision.

When submitting your revision please upload a file under "Response to Referees" - in the "File Upload" section. This should document, point by point, how you have responded to the reviewers' and Editors' comments, and the adjustments you have made to the manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Your main manuscript should be submitted as a text file (doc, txt, rtf or tex), not a PDF. Your figures should be submitted as separate files and not included within the main manuscript file.

When revising your manuscript you should also ensure that it adheres to our editorial policies (https://royalsociety.org/journals/ethics-policies/). You should pay particular attention to the following:

Research ethics:
If your study contains research on humans please ensure that you detail in the methods section whether you obtained ethical approval from your local research ethics committee and gained informed consent to participate from each of the participants.

Use of animals and field studies:
If your study uses animals please include details in the methods section of any approval and licences given to carry out the study and include full details of how animal welfare standards were ensured. Field studies should be conducted in accordance with local legislation; please include details of the appropriate permission and licences that you obtained to carry out the field work.

Data accessibility and data citation:
It is a condition of publication that you make available the data and research materials supporting the results in the article. Please see our Data Sharing Policies (https://royalsociety.org/journals/authors/author-guidelines/#data). Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the

article (https://royalsociety.org/journals/ethics-policies/data-sharing-mining/). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should also be fully cited and listed in the references.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link
http://datadryad.org/submit?journalID=RSPB&manu=(Document not available), which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

For more information please see our open data policy http://royalsocietypublishing.org/data-sharing.

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Please try to submit all supplementary material as a single file.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

Please submit a copy of your revised paper within three weeks. If we do not hear from you within this time your manuscript will be rejected. If you are unable to meet this deadline please let us know as soon as possible, as we may be able to grant a short extension.

Thank you for submitting your manuscript to Proceedings B; we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes,
Dr Robert Barton
mailto: proceedingsb@royalsociety.org

Associate Editor
Comments to Author:
Two reviewers have provided feedback on this article, and both reviewers praised merits of this study both in terms of the questions asked and the experimental design. However, both reviewers had a number of questions and concerns that must be addressed - not just via responses to the reviewers, but through revisions to the article itself. In addition to some requests for greater clarity from reviewer 2, reviewer 1 has proposed modifications to the analytical approach and refinement of the interpretations of the data. I suggest that these are warranted. Furthermore, I would ask that the authors also speak to their ability with this study to differentiate a difference between these subspecies' plasticity in responding to urban environments versus simply differences in flexibility in adapting to a lab environment (where they are now housed and tested).

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
There is growing evidence that urban animals tend to be better at finding behavioral solutions to new problems than non-urban animals, but it is unclear whether these differences reflect adaptive evolution, plasticity or habitat-matching. This study addresses by gap by comparing problem solving ability of house mouse from three subspecies, which are presumed to vary in the period they have lived in urban areas as human commensals (from ~3000 years to ~11000 years). To tease apart plastic differences, individuals were allowed to breed for several generations in the laboratory before being tested for problem-solving performance.

I read the MS with great interest, as there is little (if any) evidence of adaptive divergence caused by urbanization. While I think these results are potentially important, I have a number of concerns regarding their evolutionary interpretations.

General issues

I found that the experimental assays were well-designed and elegant. The authors take good care to properly habituate individuals, present several problems and novel objects to assess individual's consistency and measure behaviors to interpret the results. However, the theoretical framework is poorly developed. Why should urban animals be better at problem-solving? Although the authors cite literature on the importance of innovation and problem-solving, these are based on the general importance of being plastic rather than a discussion of the role of problem-solving to cope with the challenges or urban life and how these can affect mouse. I would make it clearer that living close to humans may more frequently exposes individuals to novel problems that may be solved by means of behavioral constructs. I have also a summary sentence citing previous evidence thay urban animals are better at problem solving than non-urban animals. Ignoring previous evidence may make the impression than the study is more novel than it is really. I also do not understand the logic of citing reviews discussing similar issues in differences places of the text; citations should be used more careful. I recommend the authors to read the recent book "Urban Evolutionary Biology" edited by Marta Szulkin, Jason Munshi-South and Anne Charmantier.

The authors claim that they have found adaptive divergence in problem-solving ability caused by urbanization, but I do not see where this conclusion comes from. First, the fact that individuals were allowed to breed for several generations in the laboratory may tease apart plastic differences between populations, yet do not demonstrate that differences have been cause by natural selection. Other mechanisms like matching habitat choice or genetic drift may also create genetic differences across populations. Maternal effects are a less likely explanation as these effects should be diluted in a few generations. Second, it is unclear to me how the period of human commensalism has been estimated and to what extent ~3000 years are not enough to cause adaptive responses. Finally, even when this period is accurately assessed, the three subspecies diverged ~350-500 thousand years ago and hence may have been exposed to many different selective pressures in addition of urbanization. It is thus unclear whether urbanization has been the main driver of genetic differences in problem solving. Without evidence of selection differentials or gradients, the adaptive function of problem solving remain unclear.

Still, the finding that differences in problem-solving among subspecies may have a genetic basis is interesting. However, I miss here an analysis of heritability that allows to assess the extent to which this variation may respond to selection. If individuals were kept for several generations, why do not present data on the heritability of problem-solving performance and the correlates that may either inhibit or facilitate it?

I finally find surprising that problem-solving differences cannot be explained by exploratory tendency or inhibition. The analyses presented are centered on success/failure (instead of latency to solve the problem), which may have obscured these effects. I also recommend the authors to

follow previous studies and use structural equation modeling to tease apart effects. This is because if two confounds are correlated between them because of causal links, including both as predictors in the model my block the effect of one of these variables. This can be avoided by defining causal scenarios and test them with SEM (Sol et al., 2012). Likewise, the underlying causes of problem-solving differences between subspecies may be better analyzed by SEM (Sol et al., 2011).

Specific issues:

Line 13: The term "evolutionary novel environments" looks weird,

Line 33: What do you mean by "adjust"?

Lines 36-40: This sentence is unclear. The present study also reports differences between traits without studying their adaptive significance. Moreover, the authors have not yet presented the idea that behavior matters.

Lines 49-50: This sentence may make the impression that evidence for differences in problem-solving between urban and non-urban animals is scarce, when it is not. The authors also miss all early evidence and focus on a few recent papers, suggesting that the issue has only recently been studied.

Line 52-53: Differential survival is selection. I think you refer to habitat-matching choice, where individuals with certain behaviors are more likely to settle in cities (see ref 22 in the main text).

Line 55: Reference is missing here.

Line 56: Why do you use "morphology" here if this is not mentioned again in the paragraph.

Line 56: Better" There is also little information regarding which morphological...

Line 64: The term "trial and error associative learning ability" is weird.

Lines 69-71: I do not recall that this paper discusses it.

Line 70: Not necessarily; matching habitat choice may also create genetic differences across populations.

Line 71: Is this a result?

Lines 82-84: This is a key assumption of the study, it is not enough to cite papers: 1) evidence to supports differential exposure to urbanization needs to be critically discussed; 2) whether differences in exposure are enough to generate divergence in problem solving also needs a better justification.

Line 91: Do you mean "spatial exploration"?

Line 104: Were they tested in pairs? Why?

Line 104: What is a type III cage?

Line 141: Interpreting the results of open-field test is not easy as some behaviors may reflect neophilia and others neophobia (Perals et al., 2017).

Line 164: The reference of the Package is missing.

Line 167: I do not see the logic of not using the same confounds here than in the analyses for setups with multiple solving strategies. And what about sex, why this was not included in the analysis? If sexes differ, maybe you are not observing the effect of natural selection but that of sexual selection.

Line 170: Where are the results of this comparison?

Lines 193-195: Did you test whether individuals that performed better in one setup also did so in the other setup? You could also report variation among individuals using a plot of the random factors.

Line 195: Why not simply "P < 0.0001". Moreover, you mention in the methods that these individuals were not used in the analyses.

Line 195: Do all analyses include the effect of confounding factors?

Fig. 1. Why do not add an arrow in the x-axis indicating the degree of exposure to urbanization. And why do not split the data among populations. Replication is important to convince the reader that the pattern is solid.

Line 204: What do you mean by third setup?

Lines 203-209: This section is a bit confusing, and not well integrated with the previous section. Help the reader interpreting these finding in the light of differences in problem-solving of the previous section.

Line 209: I'm lost here!

Line 224: How do you know that this is neophilia? And what about neophobia, which has been previously found to correlate with problem solving in urban animals (Sol et al., 2012)-

Line 225: What about the possibility that the lack of significance reflects insufficient power of the test?

Line 244: When was neophobia measured? Neophilia and neophobia are not opposed terms, they depend on different processes.

Line 244: Do the problems that mouse have to solve in the lab really relevant for solving problems in the wild?

Line 248: The first studies reporting differences have almost 10 years, so maybe evidence is not so new.

Line 257: I do not recall that these chapters provide evidence for fitness benefits; If I remember correctly, Reader et al. even found no relationship between innovativeness and environmental variability.

Lines 261-262: Not necessarily, as explained before.

Line 274: This conclusion is based on a cherry-picked number of papers, there are papers that show clear support for the effect of neophobic responses on problem-solving (see for example studies in common mynah). Others do not find support because the individuals have been allowed to habituate to the testing apparatus, which could be the present case. On the other hand, measuring problem-solving as failure/success reduces the probability to find effects of motivation or neophobia as an individual that solve the problem is considered successful regardless that it has done so in 1 or 10 min.

Perals, D., Griffin, A.S., Bartomeus, I. & Sol, D. 2017. Revisiting the open-field test: what does it really tell us about animal personality? Anim Behav 123: 69–79.

Sol, D., Griffin, A.S. & Bartomeus, I. 2012. Consumer and motor innovation in the common myna: The role of motivation and emotional responses. Anim Behav 83: 179–188.

Sol, D., Griffin, A.S., Bartomeus, I. & Boyce, H. 2011. Exploring or Avoiding Novel Food Resources? The Novelty Conflict in an Invasive Bird. PLoS One 6: e19535.

Referee: 2

Comments to the Author(s)
In their paper Vrbanec et al explore the problem solving abilities in three subspecies of house mice. The observed differences are interpreted as different adaptations to urban environments. Overall this is a very interesting and well written paper. I thoroughly enjoyed reading it and I think it provides and important contribution to the field. Below are some comments, suggestions for improvement and concerns.
1) Figure A2 shows some striking differences between the two populations of Mus musculus musculus (Austria vs Kazakhstan), whereby the Austrian population has the lowest likelihood to solve problems. Unless I missed it, this was not sufficiently emphasized in the Discussion. I think this should be carefully discussed as it clearly show how there's something else going on - which is not surprising given how many additional factors will likely interplay with the evolution of cognitive abilities.
2) Differences in the lenghth of commensalisms among the subspecies are indicated as the likely force shaping the observed differences. While I have no elements to dispute this, I would encourage more caution in interpreting results - it could be valuable to add a paragraph or so discussing potential factors that may also contribute to this? Would the kind of environment or type of urbanization contributed too?
1) More information should be provided about the origins of animals. I can see the ESM has some info, however it would be important to understand more such as if these were caught in cities vs villages, more information about trapping dates etc
2) While ESM 2 provides some information on the tests it would be very helpful to make some videos available so that the reader can better understand the dynamics of the tests.
3)along the same lines it would be good to provide images and videos of the neophilia test. With regards to this test I was confused by why you did not relocate the individual in the Type III (i.e. in a fashion similar to when conducting an open field test). What is the possible effect of time-to-become-aware of the novel environment (particularly important since latency to enter the environment is the key dependent variable)?
LL164 mixed effects model
4) While I am not a big fan of Likelihood ratio tests - I also acknowledge that such decisions are at the authors discretion and I think the effects reported appear to be sufficiently strong. I will suggest however to report some overall measure of fit such as squared R (which is possible to obtain through lme4). It would also be useful to include the parameter estimates (betas and SE).
Overall congratulations for an excellent and very interesting series of experiments!

Author's Response to Decision Letter for (RSPB-2020-2504.R0)

See Appendix A.

RSPB-2020-2504.R1 (Revision)

Review form: Reviewer 1

Recommendation
Accept as is

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Good

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
Comments made directly to the editor

Review form: Reviewer 2

Recommendation
Accept as is

Scientific importance: Is the manuscript an original and important contribution to its field?
Excellent

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Excellent

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
Overall I am happy with the revisions of the manuscript, congratulations for an excellent paper.

Decision letter (RSPB-2020-2504.R1)

08-Jan-2021

Dear Dr Guenther

I am pleased to inform you that your manuscript RSPB-2020-2504.R1 entitled "Enhanced problem-solving ability as an adaptation to urban environments in house mice" has been accepted for publication in Proceedings B.

The referee(s) have recommended publication, but also suggest some minor revisions to your manuscript. Therefore, I invite you to respond to the referee(s)' comments and revise your manuscript. Because the schedule for publication is very tight, it is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let us know.

To revise your manuscript, log into https://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision. You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you will be able to respond to the comments made by the referee(s) and upload a file "Response to Referees". You can use this to document any changes you make to the original manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Before uploading your revised files please make sure that you have:

1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

2) A separate electronic file of each figure (tiff, EPS or print-quality PDF preferred). The format should be produced directly from original creation package, or original software format. PowerPoint files are not accepted.

3) Electronic supplementary material: this should be contained in a separate file and where possible, all ESM should be combined into a single file. All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

4) A media summary: a short non-technical summary (up to 100 words) of the key findings/importance of your manuscript.

5) Data accessibility section and data citation
It is a condition of publication that data supporting your paper are made available either in the electronic supplementary material or through an appropriate repository.

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should be fully cited. To ensure archived data are available to readers, authors should include a 'data accessibility' section immediately after the acknowledgements section. This should list the database and accession number for all data from the article that has been made publicly available, for instance:
• DNA sequences: Genbank accessions F234391-F234402
• Phylogenetic data: TreeBASE accession number S9123
• Final DNA sequence assembly uploaded as online supplemental material
• Climate data and MaxEnt input files: Dryad doi:10.5521/dryad.12311
NB. From April 1 2013, peer reviewed articles based on research funded wholly or partly by RCUK must include, if applicable, a statement on how the underlying research materials – such as data, samples or models – can be accessed. This statement should be included in the data accessibility section.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link
http://datadryad.org/submit?journalID=RSPB&manu=(Document not available) which will take you to your unique entry in the Dryad repository. If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link. Please see https://royalsociety.org/journals/ethics-policies/data-sharing-mining/ for more details.

6) For more information on our Licence to Publish, Open Access, Cover images and Media summaries, please visit https://royalsociety.org/journals/authors/author-guidelines/.

Once again, thank you for submitting your manuscript to Proceedings B and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Sincerely,
Dr Robert Barton
Editor, Proceedings B
mailto:proceedingsb@royalsociety.org

Associate Editor:
Comments to Author:
I thank the authors for their thorough response to all reviewer feedback and for providing considerable extra detail in the article and ESM, which I believe has strengthened and clarified the article.

Both of the reviewers who reviewed the original submission reviewed this revision. Both were satisfied with the revisions and again noted the methodological strengths of this study. Neither reviewer raised outstanding concerns regarding the methods or analyses, however reviewer 1 still takes issue with the interpretation of the results but it is my opinion that the author has provided a fair response to all concerns raised in the previous round of revisions.

I appreciate the addition of repeatability analyses (i.e. ESM 5) and suggest moving these results from the supplementary materials to the main article as I think reporting these is important and will be of interest to many. The accompanying interpretation should also be moved to the main article (Discussion).

Minor comments
On lines 100 and 102 where you introduce the study subjects and their housing respectively, it might be worth referencing ESM 2 here in both places.

I believe that the style of this journal is to use a comma instead of a period to break up numbers over 1000 e.g., on line 19 3.000 should be 3,000 (ditto for 11.000), see also line 68.

Reviewer(s)' Comments to Author:
Referee: 1

Comments to the Author(s)
Comments made directly to the editor

Referee: 2

Comments to the Author(s)
Overall I am happy with the revisions of the manuscript, congratulations for an excellent paper.

Author's Response to Decision Letter for (RSPB-2020-2504.R1)

See Appendix B.

Decision letter (RSPB-2020-2504.R2)

25-Jan-2021

Dear Dr Guenther

I am pleased to inform you that your manuscript entitled "Enhanced problem-solving ability as an adaptation to urban environments in house mice" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

Open Access
You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700. Corresponding authors from member institutions (http://royalsocietypublishing.org/site/librarians/allmembers.xhtml) receive a 25% discount to these charges. For more information please visit http://royalsocietypublishing.org/open-access.

Your article has been estimated as being 7 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Paper charges
An e-mail request for payment of any related charges will be sent out after proof stage (within approximately 2-6 weeks). The preferred payment method is by credit card; however, other payment options are available

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,
Dr Robert Barton
Editor, Proceedings B

mailto: proceedingsb@royalsociety.org

Associate Editor:
Board Member
Comments to Author:
(There are no comments.)

Appendix A

Associate Editor
Comments to Author:
Two reviewers have provided feedback on this article, and both reviewers praised merits of this study both in terms of the questions asked and the experimental design. However, both reviewers had a number of questions and concerns that must be addressed - not just via responses to the reviewers, but through revisions to the article itself. In addition to some requests for greater clarity from reviewer 2, reviewer 1 has proposed modifications to the analytical approach and refinement of the interpretations of the data. I suggest that these are warranted. Furthermore, I would ask that the authors also speak to their ability with this study to differentiate a difference between these subspecies' plasticity in responding to urban environments versus simply differences in flexibility in adapting to a lab environment (where they are now housed and tested).

Response: *We have now revised the article in the light of both reviewers´ comments. To improve the clarity of methods, we have expanded our electronic supplementary material and now provide more details as to the origin and dates of capture of mice; pictures and videos of all test setups, including the novel environment and inhibitory control test. In addition, we also provide goodness of fit (R^2) measures for statistical models to show how much variation can actually be explained by our findings and we expended the discussion slightly to acknowledge also other factors (e.g. degree of urbanisation, geographical origin or population origin) to potentially influence problem-solving performance.*

We also included a number of suggestions made by reviewer 2. For example, we added an analysis based on latency to solve in addition to the solving outcome. The reviewer also raised some concerns about the interpretation of our neophilia test. These were in part apparently raised by a misunderstanding of how the test was conducted (i.e. he/she was under the impression that we forced the animals into an unknown area to measure their response which would indeed not measure neophilia). By including a picture and a video of the test setup and re-wording in the methods section, we think that this has become much clearer now.

Lastly, regarding adaptation to a commensal life style versus differences in adjustments to a laboratory environment, we have several lines of arguments supporting that the differences we see are not due to differences in adjusting to a cage environment. A) We keep all of these populations under laboratory conditions for several generations now. All of them reproduce easily and there are only little differences in reproductive output, which are mainly caused by size-differences. B) A frequent sign of suffering (i.e. of not being well adjusted) in lab-housed rodents is the development of stereotypies, in mouse mainly excessive food chewing, flipping and running in circles. We recorded the occurrence of stereotypies in our animals (see also raw data submitted to Dryad) to test if they would influence our experiments. This was not the case and we also did not find any differences in occurrence of stereotypies between different populations. C) We made sure that populations actually differ genetically before testing for phenotypic differences (for a detailed analysis of genetic differences see Harr et al. 2016 – "Genomic resources for wild populations of the house mouse, Mus musculus and its close relative Mus spretus). Based on all of these facts, we are confident that any differences we see are not merely attributed to differences in adjustment to a laboratory environment.

We hope that our improvements to the manuscript and our detailed responses to reviewers (see below) can convince you to accept our article for puplication.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
There is growing evidence that urban animals tend to be better at finding behavioral solutions to new problems than non-urban animals, but it is unclear whether these differences reflect adaptive evolution, plasticity or habitat-matching. This study addresses by gap by comparing problem solving ability of house mouse from three subspecies, which are presumed to vary in the period they have lived in urban areas as human commensals (from ~3000 years to ~11000 years). To tease apart plastic differences, individuals were allowed to breed for several generations in the laboratory before being tested for problem-solving performance.

I read the MS with great interest, as there is little (if any) evidence of adaptive divergence caused by urbanization. While I think these results are potentially important, I have a number of concerns regarding their evolutionary interpretations.

General issues

I found that the experimental assays were well-designed and elegant. The authors take good care to properly habituate individuals, present several problems and novel objects to assess individual's consistency and measure behaviors to interpret the results. However, the theoretical framework is poorly developed. Why should urban animals be better at problem-solving? Although the authors cite literature on the importance of innovation and problem-solving, these are based on the general importance of being plastic rather than a discussion of the role of problem-solving to cope with the challenges or urban life and how these can affect mouse. I would make it clearer that living close to humans may more frequently exposes individuals to novel problems that may be solved by means of behavioral constructs.

Response: *Thank you for the positive assessment of our experimental setup. We have re-phrased several parts of the introduction to better highlight the importance of problem-solving not just as a measure for flexibility but as a strategy to cope with challenges presented by a human-altered environment.*

I have also a summary sentence citing previous evidence thay urban animals are better at problem solving than non-urban animals. Ignoring previous evidence may make the impression than the study is more novel than it is really.

Response: *We explicitly state that several previous studies found urban populations to outperform their rural conspecifics in problem-solving, see for example lines 50-52. We do claim that our results are completely novel in this aspect but that they are amongst the first to actually show that such differences lead to an evolutionary (rather than just a plastic) response.*

I also do not understand the logic of citing reviews discussing similar issues in differences places of the text; citations should be used more careful. I recommend the authors to read the recent book "Urban Evolutionary Biology" edited by Marta Szulkin, Jason Munshi-South and Anne Charmantier.

Response: *In several places we chose to cite summarising reviews instead of several empirical articles to be able to adhere to the word restrictions of PRSB. We wanted to prevent leaving the impression of "cherry-picking" papers for citations. However, we acknowledge that this may not have been ideal and instead refer to empirical articles in several places now. Thank you for pointing us to the exciting book that was published earlier this year, we were not aware of that book yet.*

The authors claim that they have found adaptive divergence in problem-solving ability caused by urbanization, but I do not see where this conclusion comes from. First, the fact that individuals were allowed to breed for several generations in the laboratory may tease apart plastic differences between populations, yet do not demonstrate that differences have been cause by natural selection. Other mechanisms like matching habitat choice or genetic drift may also create genetic differences across populations. Maternal effects are a less likely explanation as these effects should be diluted in a few generations.

Response: *We agree that maternal effects and developmental plasticity cannot influence our results because animals have been bred randomly in an outbreeding scheme for several generations under common environment, i.e. lab conditions. Random evolutionary mechanisms such as genetic drift are unlikely to have caused the observed differences among subspecies because we a) tested always 2 populations from very different geographical areas per subspecies and b) all founding individuals were caught in different places mostly several kilometres apart from each other. All our populations were founded by non-related individuals that did not share a common environment. In addition, we already knew that the populations differ in genetic background (see Harr et al. 2016). And lastly, from most places in which our mice were caught, we know from other studies that house mice in these areas only occur in human settlements and not under more natural conditions. For example, mice from Taiwan: Chou et al. 1998 and Wu et al. 2006 show that house mice usually only occur in human dwellings but not under natural conditions and Yu and Peng (2002) show that Taiwanese populations in different areas are not isolated and gene flow is happing between animals of different geographical places. Such evidence also exists for mice populations in other areas, in fact, MD and MM are much better studied than MC populations. In essence, such detailed studies about house mice all over the world show that the majority of populations only occur commensally with humans and that populations have a fast turn-over rate with a high degree of gene-flow. Therefore, we are quite confident that the observed differences in problem-solving across populations and subspecies are caused by natural selection.*

Chou, C. W., Lee, P. F., Lu, K. H., & Yu, H. T. (1998). A population study of house mice (Mus musculus castaneus) inhabiting rice granaries in Taiwan. *Zoological Studies*, *37*(3), 201-212.

Wu, S. Y., Lin, Y. T., & Yu, H. T. (2006). Population ecology of the Southeast Asian house mouse (Muridae: Mus musculus castaneus) inhabiting rice granaries in Taiwan. *Zoological Studies*, *45*(4), 467-474.

Yu, H. T., & Peng, Y. H. (2002). Population differentiation and gene flow revealed by microsatellite DNA markers in the house mouse (Mus musculus castaneus) in Taiwan. *Zoological science*, *19*(4), 475-483.

Second, it is unclear to me how the period of human commensalism has been estimated and to what extent ~3000 years are not enough to cause adaptive responses. Finally, even when this period is accurately assessed, the three subspecies diverged ~350-500 thousand years ago and hence may have been exposed to many different selective pressures in addition of urbanization. It is thus unclear whether urbanization has been the main driver of genetic differences in problem solving. Without evidence of selection differentials or gradients, the adaptive function of problem solving remain unclear.

Response: *the period of commensalism is estimated from fossils found in archaeological remains of human settlements across Eurasia as well as mtDNA control regions and haplotype analyses (see also citations in main manuscript).*

The genus Mus is a very divers group of species with more than 600 recognised species worldwide. It has undergone several periods of adaptive radiation processes that were driven by changes to the environment. All members of the Mus musculus species were adapted to life in a relatively open habitat, including woodland, shrubland, grasslenad and steppe see e.g. Suzuki et al. (2004). Commensalism is described as one adaptation occurring only in Mus musculus (but not for example in close relatives Mus spicilegus or Mus spretus) to allow animals to expand their range. It is known that commensalism is facilitated by some heritable property which is among the genus Mus unique to Mus musculus (see Suzuki and Aplin, Phylogeny and biogeography of the genus Mus in Eurasia, 2012). The transition towards commensalisms in Mus musculus has been described to lead to the evolution of several morphological, physiological and behavioural adaptations (Suzuki and Aplin, 2012). So, although several ecological characteristics probably influenced the evolution of subspecies, commensalism is described as one of the main sources of adaptive shifts in characteristics such as reproductive cycles, behaviour (e.g. territoriality) or physiology (see also Berry 1981 – The biology of the house mouse). Hence, it is sensible to assume that it would also affect cognitive behaviours such as problem-solving in concert with other behavioural, physiological or morphological adaptations.

Still, the finding that differences in problem-solving among subspecies may have a genetic basis is interesting. However, I miss here an analysis of heritability that allows to assess the extent to which this variation may respond to selection. If individuals were kept for several generations, why do not present data on the heritability of problem-solving performance and the correlates that may either inhibit or facilitate it?

Response: *Conducting these experiments with the high number of animals needed if population and/or subspecies estimates should be compared took in total 14 months with nearly every day experimental work. This amount of time and workload would not have been feasible for us to conduct on several generations, therefore, we cannot provide estimates of*

heritability. However, we have now added a new ESM including estimates of repeatability, which is commonly used in behavioural ecology studies to give an upper boundary for heritability. Repeatability includes heritable variation as well as long-term stable phenotypic variation, i.e., parental or permanent environment effects. The reviewer him/herself acknowledges that parental effects are in this case unlikely to have caused population-differences in problem-solving ability, which we fully agree with. Similarly, permanent environment effect are unlikely to cause population differences since all populations have experienced the same environment (laboratory cages) for several generations. Therefore, we belief that our repeatability estimates reflect very closely real heritability estimates. For a more detailed discussion and explanation please see new ESM 6.

I finally find surprising that problem-solving differences cannot be explained by exploratory tendency or inhibition. The analyses presented are centered on success/failure (instead of latency to solve the problem), which may have obscured these effects. I also recommend the authors to follow previous studies and use structural equation modeling to tease apart effects. This is because if two confounds are correlated between them because of causal links, including both as predictors in the model my block the effect of one of these variables. This can be avoided by defining causal scenarios and test them with SEM (Sol et al., 2012). Likewise, the underlying causes of problem-solving differences between subspecies may be better analyzed by SEM (Sol et al., 2011).

Response: *In part, we were also surprised about these findings. We have taken up the suggestion of the reviewer to analyse effects based on solving latency in addition to solving outcome based on success/ failure. Please see additions to the results section of the main MS, new Tables A13 and A14 in ESM 4 and addition to statistical code (also ESM4). Generally, if we take the whole time to solve starting when test setups were first introduced to an individual, the results strongly reflect results of the success/failure analysis. M.m.domesticus outperforms M.m.musculus and M.m.castaneus. This analysis is in concert with what other studies have done. However, the interpretation of results changes somewhat if we take solving latency from the first approach of a setup to the actual solving of the setup. In this case, there is only a statistical difference between M.m. musculus and M.m.castaneus. We excluded animals that did not solve because we already knew that proportion of solvers differ among subspecies and taking the maximum latency would have obscured effects. We interpret this finding that there is indeed a subspecies difference in the likelihood to solve but for those individuals that do actually solve, the time they need to solve is quite similar. We have added this discussion point to the main MS.*

Importantly, even if we use the latency to solve as response variable, exploration and inhibitory control still do not explain significant proportions of variance. We belief that exploration (i.e. risk-taking or boldness) does not play a main role here due to our extensive habituation phase prior to the experiments. All participating individuals had learned that whenever a researcher enters the room, there will be an option to find a treat. Why, however, the inhibitory control also did not explain any differences in solving outcome (or latency) despite showing a significant difference between subspecies, is equally puzzling to us. This finding certainly deserves more intense investigation in the future but this beyond the scope of this project. Other studies using an equally extensive habituation, reach similar conclusions, see for example (Mazza & Guenther, in press, Animal Behaviour).

Finally, as for the use of SEM´s, we would like to keep our way of analysis for two reasons. As the reviewer correctly points out, SEM´s are a great tool to investigate underlying causes influencing a trait (for example a behaviour). However, the causality of effects is usually the main purpose for an analysis. It would for example be a great method if our main aim was to test if inhibitory control, exploration etc. was causing differences in solving success. But our main aim was to test for differences between subspecies, a very straight forward question, for which hypothesis-testing statistical methods such as linear models (and their extension, linear mixed effects models) have specifically been developed. Second, equally importantly, SEM´s are usually applied to data assuming a Gaussian distribution. It seems that there are new methods such as "piecewiseSEm" (https://cran.r-project.org/web/packages/piecewiseSEM/vignettes/piecewiseSEM.html) methods have become available recently, but we have not yet seen them being used in any ecological/behavioural study and hence cannot judge their merit or potential pitfalls. Therefore, we have decided to stick to the more classic way of analysing our data.

Specific issues:

Line 13: The term "evolutionary novel environments" looks weird,

Response: *We changed the sentence to: "The ecological effects of human-altered environments are being increasingly recognised and understood but their effects on evolution are largely unknown".*

Line 33: What do you mean by "adjust"?

Response: *We changed to "adapt"*

Lines 36-40: This sentence is unclear. The present study also reports differences between traits without studying their adaptive significance. Moreover, the authors have not yet presented the idea that behavior matters.

Lines 49-50: This sentence may make the impression that evidence for differences in problem-solving between urban and non-urban animals is scarce, when it is not. The authors also miss all early evidence and focus on a few recent papers, suggesting that the issue has only recently been studied.

Response: *We rephrased to: "The behavioural repertoire of animals is crucial for coping with environmental challenges and growing evidence reports behavioural differences between populations living in close contact to humans over naturally occurring populations across taxa"*

Line 52-53: Differential survival is selection. I think you refer to habitat-matching choice, where individuals with certain behaviors are more likely to settle in cities (see ref 22 in the main text).

Response: *We changed the sentence to:* ***"****The underlying mechanisms of problem-solving differences are poorly understood. It is not yet known if the human environment acts as a*

filter in which individuals with certain behavioural types are more likely to settle in urban areas, if differential selection pressures lead to microevolution or if plastic changes in behaviour cause differences in problem-solving ability."

Line 55: Reference is missing here.

Response: We added a reference.

Line 56: Why do you use "morphology" here if this is not mentioned again in the paragraph.

Response: *We deleted "morphological".*

Line 56: Better" There is also little information regarding which morphological...

Response: *We changed the sentence accordingly.*

Line 64: The term "trial and error associative learning ability" is weird.

Response: *we changed to trial-and-error learning.*

Lines 69-71: I do not recall that this paper discusses it.

Response: *We choose a reference here that in general discusses different mechanisms for behavioural adjustments towards living in urbanised areas. They provide a detailed overview and explanations of processes that can generate genetic adjustments. Therefore, we believe that this is a correct citation here.*

Line 70: Not necessarily; matching habitat choice may also create genetic differences across populations.

Response: *We changed the wording from "differential selection pressures" to the broader term "genetic adaptations", which could include both processes.*

Line 71: Is this a result?

Response: *We moved this sentence to the discussion.*

Lines 82-84: This is a key assumption of the study, it is not enough to cite papers: 1) evidence to supports differential exposure to urbanization needs to be critically discussed; 2) whether differences in exposure are enough to generate divergence in problem solving also needs a better justification.

Response: *House mice are generally considered one of a handful of animal species being archetypes of early urban evolution (see for example Johnson and Munshi-South, 2017) because their history with humans is well known in the literature. Several studies from different teams of authors investigate the phylogeny, biogeography and synanthrophy of house mice (for example reviewed in "Biology of the house mouse, Eds. Berry, 1981"; "The evolution of the house mouse, Eds. By Macholán, Baird, Munclinger, Piálek, 2012" and several others). As a consequence of these studies, the onset of synanthrophy, as well as the spread of house mice around the globe in human company, is well documented.*

In addition, studying behavioural and life history differences between house mice subspecies and facultative versus obligate commensal populations, have been a major research field some decades ago. From these studies, we know that house mice have changed several aspects of their phenotype, including life history decisions (e.g. dispersal, breeding cycles), morphological (e.g. tail length) and behavioural traits (e.g. territorial behaviour, activity, exploration) to adjust to a commensal life-style. These studies show that mice are able to and actually do alter aspects of their phenotype when exposed to human-altered environments. This already demonstrated capability in adaptation to human-altered environments has led us to test the hypothesis that innovative problem-solving differs between subspecies which have been exposed to human environments for different times. A series of studies investigating problem-solving in rural versus urban populations across different species strongly suggested that exposure is enough to facilitate changes in problem-solving (as we have mentioned in the introduction).

We have reworded the second part of this paragraph to better highlight the well documented known phenotypic adaptations in house mice and included two more specific references about house mice.

Johnson, M. T., & Munshi-South, J. (2017). Evolution of life in urban environments. *Science, 358*(6363).

Line 91: Do you mean "spatial exploration"?

Response: *No, we mean specifically exploration of elevated areas, i.e. areas which animals need to climb up to. We rephrased the sentence to say: "Frynta et al. [27] found commensal MM populations to show increased exploration of elevated places over non-commensal populations".*

Line 104: Were they tested in pairs? Why?

Response: *For experiments, pars were separated as mentioned at the end of this section. We kept them in pairs because mice (especially wild mice) are very social and suffer when being kept alone. To enable animals living in company but the opportunity to test them separately without interfering effects of being transferred to an unknown environment, we used the setup of 2 connected cages that could be separated briefly by inserting a wire-mesh. We now provide a picture of the cage setup in ESM 2. We hope that this makes it clearer.*

Line 104: What is a type III cage?

Response: *We changed the wording to "standard cage" and describe 2 sentences before that we used Makrolon type III standard mouse cages for the experiments. Cage type III refers to conventional, standardised sized rodent cages of 425 x 265 x 150 mm.*

Line 141: Interpreting the results of open-field test is not easy as some behaviors may reflect neophilia and others neophobia (Perals et al., 2017).

Response: *We absolutely agree with this statement. This is exactly why we did not use an Open Field test where the animal is forced into an unknown environment. Instead, we attached an unknown environment directly to the home cage of the animal and gave it a free*

choice to enter and explore the environment or to stay in its home cage, i.e. to stay in safety. Hence, the response we see if only towards the novelty of the situation and excludes potential confounding effects of a forced component. Furthermore, since the animal can choose not to react to the unknown environment at all, if it chooses to investigate it, the response is neophilia and not neophobia. We have now provided images (ESM 2) and a video about the Neophilia test which we hope makes the distinction to an Open Field test clearer.

Line 164: The reference of the Package is missing.

Response: This sentence is just a general description of how models were build. Since we used multiple different R packages, we give the references to the specific packages when describing specific models in more detail.

Line 167: I do not see the logic of not using the same confounds here than in the analyses for setups with multiple solving strategies. And what about sex, why this was not included in the analysis? If sexes differ, maybe you are not observing the effect of natural selection but that of sexual selection.

Response: *There are two reasons why we did not include some confounding factors such as interaction time or approach latency in the second set of setups (those with a specific solving strategy). First, when we confronted animals with this second set of setups, they were already highly trained by the first batch of setups and were both, approaching much faster and interacting less long. This led to very little variance (compared to the first batch of setups) and to differences in approach latency, interaction time etc. between both types of setup. We wanted to avoid potentially different interpretations and instead decided to focus on the most reliable readout, the outcome of the solving attempt throughout. Please note that we have run all models also for the first batch with and without these potentially confounding variables and we arrived at the same conclusions.*

Regarding sex effects, we did not include sex since we did not have any prior expectations for sexes to differ in foraging innovations. Both sexes search for food in similar fashion and are thus equally likely to encounter food that can only be obtained by inventing a new strategy for extraction. However, we acknowledge that sex effects have often been overlooked in the past and re-run all our initial models including sex and the two-way interaction of sex and subspecies. Neither the interaction, nor the main effect of sex showed statistical significance in problem-solving (setups with multiple solving strategies, all $p > 0.20$; setups with a specific solving strategy, all $p > 0.22$; control, all $p > 0.1$), novel environment exploration (all $p > 0.51$) or inhibitory control (all $p > 0.59$).

In an attempt to keep the manuscript most concise and focus on the main results in a hypothesis-driven way, we decided not to include these additional aspects.

Line 170: Where are the results of this comparison?

Response: *the results of all models, with and without participating animals, can be found in ESM 4. We added a reference to the ESM in the main manuscript.*

Lines 193-195: Did you test whether individuals that performed better in one setup also did so

in the other setup? You could also report variation among individuals using a plot of the random factors.

Response: *No, we did not. We do not quite understand how this would be important to answer our question. Nevertheless, as a response to another of your comments, we now report "repeatability", which basically carries this information. It suggests that there is a level of consistent individual differences comparable to other behaviour traits. Please see new ESM 6 for further information.*

Line 195: Why not simply "P < 0.0001". Moreover, you mention in the methods that these individuals were not used in the analyses.

Response: *We have changed to p< 0.001 throughout. We say in the statistical analyses that we performed all models with and without the participating animals. Since we describe our main result here, we wanted to report both versions to show the robustness of the result.*

Line 195: Do all analyses include the effect of confounding factors?

Response: *yes, all results are based on the confounding factors described in the statistical analyses part.*

Fig. 1. Why do not add an arrow in the x-axis indicating the degree of exposure to urbanization. And why do not split the data among populations. Replication is important to convince the reader that the pattern is solid.

Response: *We added arrows to the x-axes of figures 1 and 2, indicating the onset of synanthrophy between ~3000 y to ~11.000 years ago. The population-level analysis, including detailed statistical pair-wise comparisons, can be found in ESM 3. We would like to keep this analysis, as well as the detailed analyses across setups in the ESM to keep the main manuscript concise and focused.*

Line 204: What do you mean by third setup?

Response: *Thank you for noticing this oversight, we changed to setup S3, as well as S1 and S4 in the next sentence. This is in accordance with description of setups given in the methods section now.*

Lines 203-209: This section is a bit confusing, and not well integrated with the previous section. Help the reader interpreting these finding in the light of differences in problem-solving of the previous section.

Response: *We rephrased this part to: "Particularly setup S3 was more difficult to solve than the remaining setups (all pairwise comparisons $p < 0.001$, see ESM 3). However, even when excluding this setup, subspecies differences in solving-success were consistent ($df = 2$, $\chi^2 = 25.7$, $p < 0.001$). Subspecies performed equally well in the control setup with more than 80% of all animals participating (Fig. 2a)." We hope that this change and the proper references to setups S3 (previous comment) improve the readability of this section.*

Line 209: I'm lost here!

Response: *please see reply to the previous 2 comments, we re-phrased this section.*

Line 224: How do you know that this is neophilia? And what about neophobia, which has been previously found to correlate with problem solving in urban animals (Sol et al., 2012)-

Response: *As explained in the methods section, the unknown cage including unknown objects, was attached to the individuals home cage but the animal could freely choose to enter and explore this novelty or to remain in the home cage, i.e. in the safe area. Therefore, if an animal chooses to explore this novelty, the reaction cannot be out of fear (i.e. neophobia) but should be based on curiosity and hence represent neophilia.*

We did not test for neophobia because we think that it will not influence our results due to our test design. Neophobia, the fear of the unknown, should play a role primarily when animals try to avoid contact with unknown (objects or places). Due to our extensive habituation phase (which for some populations lasted only one week, for others up to 6 weeks), we trained the animals to expect our test setups and to associate these setups with a food reward. We therefore tried our best to reduce any signs of neophobia in these animals.

Line 225: What about the possibility that the lack of significance reflects insufficient power of the test?

Response: *We conducted several power analyses before we started our series of tests to estimate the number of animals needed to achieve statistical significance based on different effect sizes. With the number of animals we choose, we are able to detect statistically significant differences up to E = 0.25 which is slightly lower than most reported effect sizes for ecologically relevant behavioural traits. Thus, if we lack statistical power in this test, it means that the differences between subspecies would be less than 25%. In such cases, we would wonder if such small differences could actually have an impact the life of animals under natural conditions.*

Line 244: When was neophobia measured? Neophilia and neophobia are not opposed terms, they depend on different processes.

Response: *Thank you for having an amazing eye for detail, this was a mistake in wording, it should have been neophilia. We changed accordingly.*

Line 244: Do the problems that mouse have to solve in the lab really relevant for solving problems in the wild?

Response: *We build our experimental setups based on solving actions of which we thought they would resemble techniques that mice could need to invent in the "wild" (please note that wild for a house mouse usually means a barn, stable or another kind of human made housing). Opening setups in different ways, e.g. by pulling, pushing etc. should resemble mechanisms by which humans try to save food or food for animals. We acknowledge however, that we cannot be certain of that. However, the alternative would be to mainly rely on innovations observed in the "wild" but since these happen only very rarely (and are observed even more rarely), this method would not be suited for an experimental design in our point of view.*

Line 248: The first studies reporting differences have almost 10 years, so maybe evidence is not so new.

Response: *We deleted the word "recent".*

Line 257: I do not recall that these chapters provide evidence for fitness benefits; If I remember correctly, Reader et al. even found no relationship between innovativeness and environmental variability.

Response: *We originally cited chapters discussing the potential adaptive value of innovative problem-solving here. We have now changed this to 2 citations which provide experimental evidence for the effect of problem-solving on fitness.*

Lines 261-262: Not necessarily, as explained before.

Response: *We do believe that we are able to show an evolutionary response in problem-solving for several reasons explained in different responses. In summary:*

a) by raising animals under standard laboratory conditions for several generations, we exclude that differences in problem solving are the result of any form of plasticity (flexibility, learning, developmental plasticity or maternal effects).

b) founding animals of all populations were caught in several different places at each trapping location to make sure to capture an ecologically relevant variation in mice geno- and phenotypes for each population (please see also ESM 1).

c) there is already published evidence that mice from the populations we use show genome wide differences in genetic makeup and gene expression.

d) We acknowledge that also other factors that exposure to human-altered environments may influence problem-solving (see for example factors discussed in ESM 3 and briefly in the discussion of the main manuscript), however, for MC and MD, we could show that populations from very different geographical and climatic regions do not differ in problem-solving performance.

Line 274: This conclusion is based on a cherry-picked number of papers, there are papers that show clear support for the effect of neophobic responses on problem-solving (see for example studies in common mynah). Others do not find support because the individuals have been allowed to habituate to the testing apparatus, which could be the present case. On the other hand, measuring problem-solving as failure/success reduces the probability to find effects of motivation or neophobia as an individual that solve the problem is considered successful regardless that it has done so in 1 or 10 min.

Response: *We fully agree that there are studies showing effects of neophilia and studies showing effects of neophobia on problem-solving. There are however mixed reports in the literature, as you probably point out correctly, due to differences in experimental design. We here merely say that we did not find such effects in our study, we do not claim that there is no effect in general. To point out that differences in methodology might lead to differences in findings, we changed our concluding sentence to: "Here, we found no evidence for an influence of object or place neophilia on problem-solving, potentially due to our intensive habituation phase.". The papers we cited are merely meant as examples, they will not reflect*

the full spectrum of findings. However, since investigating the effects of neophilia on problem-solving was not our main aim here, we´d like to avoid a lengthy discussion on this topic.

Regarding the scoring of problem-solving as 1/0 success versus failure, it is of course true that we would have a higher chance of finding an effect of neophilia if we consider latencies to solve. Based on one of your previous comments, we included an analysis of the latency to solve (from first approach of setup to solving) and, although MM needed longer to solve than MC or MD, there was no direct influence of neophilia. Interestingly, if we consider as latency to solve the time from introducing the test setup to actual solving (as done in many studies), there is an effect of neophilia. This is due to the fact that subspecies differ in approach latencies (see line 222) and MM need longest to approach the setup. Nevertheless, we were primarily interested in the solving capability rather than in the approach of the setup.

Perals, D., Griffin, A.S., Bartomeus, I. & Sol, D. 2017. Revisiting the open-field test: what does it really tell us about animal personality? Anim Behav 123: 69–79.

Sol, D., Griffin, A.S. & Bartomeus, I. 2012. Consumer and motor innovation in the common myna: The role of motivation and emotional responses. Anim Behav 83: 179–188.

Sol, D., Griffin, A.S., Bartomeus, I. & Boyce, H. 2011. Exploring or Avoiding Novel Food Resources? The Novelty Conflict in an Invasive Bird. PLoS One 6: e19535.

Referee: 2

Comments to the Author(s)
In their paper Vrbanec et al explore the problem solving abilities in three subspecies of house mice. The observed differences are interpreted as different adaptations to urban environments. Overall this is a very interesting and well written paper. I thoroughly enjoyed reading it and I think it provides and important contribution to the field. Below are some comments, suggestions for improvement and concerns.

Response: *Thank you for your kind words and the overall positive evaluation of our work!*

1) Figure A2 shows some striking differences between the two populations of Mus musculus musculus (Austria vs Kazakhstan), whereby the Austrian population has the lowest likelihood to solve problems. Unless I missed it, this was not sufficiently emphasized in the Discussion. I think this should be carefully discussed as it clearly show how there's something else going on - which is not surprising given how many additional factors will likely interplay with the evolution of cognitive abilities.

Response: *We have analysed the respective data more closely to investigate this population difference. We found that it is mainly driven by WI animals failing to solve in more complex setups. While they are equally likely to solve the two most easy setups (S1 and S2, compared with the second population of M.musculus musculus, KH), none of them could solve the two slightly more complex setups (S3 and S4) of the setups offering multiple solving strategies. It indeed suggests that there is something going on. We have added a split analysis of setups*

and a short discussion to ESM 2 and now refer to these population differences in M.musculus musculus also in the discussion of the main MS. We have not added a whole section about population effects into the discussion in an emphasis to keep the manuscript concise.

2) Differences in the length of commensalisms among the subspecies are indicated as the likely force shaping the observed differences. While I have no elements to dispute this, I would encourage more caution in interpreting results - it could be valuable to add a paragraph or so discussing potential factors that may also contribute to this? Would the kind of environment or type of urbanization contributed too?

Response: *We completely agree that several other environmental factors are likely to play a role in the evolution of problem-solving and we have added a brief discussion about this to the manuscript. As for the type of urbanisation: This will likely play a major role since selection factors will probably strongly differ between rural sites, including villages and farms versus highly urbanised areas as for example big, fast changing cities. All our mice originated from more rural areas, which also present more closely the habitats in which commensal house mice will have evolved during the last 10.000 years. We have at the moment no data to compare problem-solving capabilities between rural and highly urban house mice although this would be extremely interesting. Recent studies in a closely related mouse species (Apodemus agrarius), for example indicate that mice living in a big city (Berlin, Germany) are able to outperform their rural conspecifics (Mazza & Guenther, 2020, in press, Animal Behaviour).*

1) More information should be provided about the origins of animals. I can see the ESM has some info, however it would be important to understand more such as if these were caught in cities vs villages, more information about trapping dates etc

Response: *Done. We have restructured ESM 1 to now include more details about the trapping sites as well as the time of trapping. Specifically, we have added the years in which wild mice were originally caught and provide detailed explanation of the trapping locations, including GPS coordinates wherever available. All mice were trapped in villages or at the suburbs of towns, mostly in small shops or horse stables. Thus, all mice are descendants of commensal populations and trapped in human settlements. Furthermore, all mice were derived from low intense urbanised areas (i.e. no big cities), representing habitats in which commensally living mice have evolved in the last centuries/millennia.*

2) While ESM 2 provides some information on the tests it would be very helpful to make some videos available so that the reader can better understand the dynamics of the tests.

Response: *Thank you for suggesting this, this is indeed an excellent method to show how tests work. We have uploaded several short video files showing a) the habituation, b) three problem-solving apparatuses, c) novel environment exploration and d) part of the inhibitory control test.*

3)along the same lines it would be good to provide images and videos of the neophilia test. With regards to this test I was confused by why you did not relocate the individual in the Type III (i.e. in a fashion similar to when conducting an open field test). What is the possible

effect of time-to-become-aware of the novel environment (particularly important since latency to enter the environment is the key dependent variable)?

Response: *We now provide images and a video of the test setups for the neophilia test as well as the inhibitory control test in addition to the images already provided for the problem-solving tests (see ESM 2).*

We did not relocate (force) individuals into the novel environment since this would mix responses to an uncontrollable (stressful) situation with reactions towards novelty. Instead, we decided to give the animals a free choice to voluntarily enter (or not) an unknown environment. Hence, the reaction we see should be driven by curiosity and the urge to explore the unknown but should not include "fear" reactions.

To attach the novel environment to the home cage, we needed to touch the home cage, thereby inevitably producing some noise and movement. The animals were thus aware that we had changed something on their home cage and also new the direction immediately. The time to become aware of the unknown area was thus negligible and, most importantly, comparable across all animals.

LL164 mixed effects model

4) While I am not a big fan of Likelihood ratio tests - I also acknowledge that such decisions are at the authors discretion and I think the effects reported appear to be sufficiently strong. I will suggest however to report some overall measure of fit such as squared R (which is possible to obtain through lme4). It would also be useful to include the parameter estimates (betas and SE).

Response: *We followed your suggestion and now provide R_2 for all models, please see updated results section of the main MS.*

Overall congratulations for an excellent and very interesting series of experiments!

Response: *Thank you very much!*

Appendix B

Dear Dr Guenther

I am pleased to inform you that your manuscript RSPB-2020-2504.R1 entitled "Enhanced problem-solving ability as an adaptation to urban environments in house mice" has been accepted for publication in Proceedings B.

Sincerely,

Dr Robert Barton
Editor, Proceedings B

Dear Dr. Barton,

thank you for accepting our manuscript for publications in Proceedings B. We have made the all required revisions and uploaded new versions of our manuscript and associated files.

Sincerely on behalf of all co-authors,

Dr. Anja Guenther

Associate Editor:
Comments to Author:
I thank the authors for their thorough response to all reviewer feedback and for providing considerable extra detail in the article and ESM, which I believe has strengthened and clarified the article.
Both of the reviewers who reviewed the original submission reviewed this revision. Both were satisfied with the revisions and again noted the methodological strengths of this study. Neither reviewer raised outstanding concerns regarding the methods or analyses, however reviewer 1 still takes issue with the interpretation of the results but it is my opinion that the author has provided a fair response to all concerns raised in the previous round of revisions.

Thank you very much for the positive evaluation. We have included all your comments in the new version of the manuscript.

I appreciate the addition of repeatability analyses (i.e. ESM 5) and suggest moving these results from the supplementary materials to the main article as I think reporting these is important and will be of interest to many. The accompanying interpretation should also be moved to the main article (Discussion).

We have added the repeatability analysis to the methods and results section and included a short discussion.

Minor comments
On lines 100 and 102 where you introduce the study subjects and their housing respectively, it might be worth referencing ESM 2 here in both places.

done

I believe that the style of this journal is to use a comma instead of a period to break up numbers over 1000 e.g., on line 19 3.000 should be 3,000 (ditto for 11.000), see also line 68.

We have changed the numbers throughout the manuscript.