THE ROYAL SOCIETY
PUBLISHING

PROCEEDINGS B

Symbolic representation of numerosity by honeybees (*Apis mellifera*): matching characters to small quantities

Scarlett R. Howard, Aurore Avarguès-Weber, Jair E. Garcia, Andrew D. Greentree and Adrian G. Dyer

Article citation details
Proc. R. Soc. B **286**: 20190238.
http://dx.doi.org/10.1098/rspb.2019.0238

Review timeline

Original submission:	29 January 2019
1st revised submission:	9 April 2019
2nd revised submission:	11 May 2019
Final acceptance:	13 May 2019

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

Review History

RSPB-2019-0238.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?
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
In their study 'Symbol and numerosity matching in the honeybee (Apis mellifera)' Howard et al. investigate the honeybee's capability of associating number value with abstract symbols and vice versa in a previously established delayed-match-to-sample assay. While the bees manage to associate both symbol with number (2 or 3) and number with symbol, they seem to fail in the subsequent reversal test. Pushing the boundaries of associative task complexity in insects is best done in honeybees, and the authors already published a couple of fascinating papers in this regard in recent years, where they challenged the cognitive - and especially the numerical - capabilities of bees.
Here, Howard et al. provide an exciting study that adds a novel nuance to our knowledge about animal cognition and thus has general significance to biologists. The manuscript is well written and sets the question in adequate biological context; the results would benefit from showing individual data points, but the experiments are mostly well-designed and controlled. The manuscript does fall short on providing some relevant information about the time course of the experiments (see below). Maybe related to this is my biggest concern: this work completely disregards the effect of extinction by unreinforced choices during testing. While this is unavoidable for individual tests, it could heavily affect the outcome of different test paradigms that are performed sequentially (as done here). In this light I recommend the manuscript for publication after addressing this point - either by convincing me that my concern is unfounded or by additional experiments (see below).

Minor comments:
line 68-72: please comment on the time course of the experiment.
line 128: showed → shown
line 137: What was the time frame of the experiments? How long did each part last? Please also add information about the delay between training and testing the bees, and between the different test situations. (either here or in line 68-72)

Figure 2:
* spelling error : 'radndom → random' in both a) and b)
* please spell out 'SA' in figure

Major comments:
Figure 3:
Only 20 bees were tested altogether, which is consistent with their previous papers in the same setup, but in stark contrast to the amount of bees the same authors tested in their recent paper in Current Zoology for one figure alone (n=138). This low number however allows for showing individual choice probabilities for the 10 bees for each data point in 3a)-d) (i.e. with beeswarm plot in R).
If the authors also connect the data for each animal across trials in 3a) and 3b), individual acquisition performances can be observed (ideally in different colors). Taken together, this adds another layer of information and represents the acquired data more transparently. Along those lines, why did the authors use a linear fit? Although this was done in previous publications too, it was and still is misleading. Acquisition of an association is not necessarily linear (more often exponential or sigmoidal), and a nonlinear fit would also be more appropriate, given the low number of tested individuals. Please use a higher-order polynomial or other, more suitable fit to describe the data.
If I interpret the experimental layout in Fig. 2 correctly, the two groups were further split to test learning with equal surface area or with equal element size. This would result in an n=5 for bars 2 & 3 in fig. 3c) and d), respectively. Please show individual data points for each animal's proportion of correct choices, and add the corresponding count of animals to the legend of 3c) and d).

As mentioned before, please add more information about the time course of the experiments. As soon as bees are tested, the learned association also becomes extinguished, since each unreinforced choice represents an extinction event. In case of the reversal test, the bees already underwent 20 extinction trials (during transfer test and learning test), which would inevitably influence performance.

Please demonstrate that this particular appetitive-aversive differential delayed-match-to-sample training paradigm leads to robust long-lasting memory. Also, please demonstrate that there was a sufficient period for spontaneous recovery of memory between test types. Otherwise, the lack of performance in the reversal test could be explained by extinction.
Since the authors put sufficient emphasis on the lack of reversal learning performance in bees in their discussion, they should 1) either provide the aforementioned information or 2) add more experiments where reversal tests are performed immediately after the 50 training trials.

Review form: Reviewer 2

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.
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
This is an interesting paper. However, some issues need clarification before any final decision about the suitability of this manuscript for publication can be reached. In particular, it seems to me that a control experiment with non-numerical stimuli would be necessary to a proper interpretation of the bees failure at reversing the symbol-number vs. number/symbol task.

Specific comments

-P 5, line 92: the 20% ethanol concentration to clean the apparatus might not be enough to remove all the olfactory cues. Literature supports a 30% or higher ethanol concentration:
https://royalsocietypublishing.org/doi/full/10.1098/rspb.2017.2278
https://link.springer.com/article/10.1007/s10071-017-1086-6
http://science.sciencemag.org/content/360/6393/1124.abstract?casa_token=i6BqQcIG03cAAAA A:tDQxpxNPfKiplbIitH8_lfMauE4jE90PaWpQm4-39ilfwYd81LilG9zI2qPnuQ9B2aO3grz93zNGZg

-P 6, line 115: It is written that the 3) set was used as a control for the continuous variables that may covary with numerosity. However, how these controls were made is not specified. In particular, it would be useful to know whether all the stimuli in this group were simultaneously controlled for all the features or not. In the first case, it would be very difficult to be able to control all the continuous variables at a time (for example, it is very difficult to simultaneously control for the density and the convex hull, as well as for the area and the line length, since these two variables are inversely related).

-P 8: Looking at the graphs and according to the variability showed by the error bars, my concern is whether there are differences between the performances of the two groups, both during the training phase and the test phase.

-P 10, Discussion: before arguing for an inability of honeybees to reverse a symbol-number association, a control might be needed. Bees should be trained with a DMTS where non-numerical stimuli will be used (colors and shapes or different symbols) in order to establish a causal association (e.g. if yellow than circle; if blue than triangle): Then, subjects should be tested in a reverse test. The hypothesis is that if bees are able to do the reverse task when presented with non-numerical DMTS, then their inability to reverse the symbol-number association is strictly connected to the number matching. In contrast, if bees are not able to reverse their response in a non-numerical DMTS, then a general inability to deal with operational schemes can be hypothesized.

Line 32: I would not use ‘numbers’ here but ‘numerosities’. In general, I think the authors should make an effort to provide the reader (who is not necessarily a specialist on numerical cognition) that the type of evidence available for non-human animals for quantity (discrete numerousness) is relative to an approximate, analogous representation (so-called ANS, Approximate Number System) and that the association which has been studied in work with e.g. monkeys (see for instance reference 19) refers to a link between these non-symbolic, approximate numerousness representation and symbols. See e.g. for a recent review Vallortigara, G. (2017). An animal’s sense of number. In “The nature and Development of Mathematics. Cross Disciplinary Perspective on Cognition, Learning and Culture” (Adams, J.W., Barmby P., Mesoudi, A., eds.), pp. 43-65, Routledge, New York.

Line 53-54: Here, again I am worried that a wrong message would be conveyed to the non-specialist reader, for evidence for “zero” in bees and other non-human animals is not relative to number zero (that would be possible only in association with a symbol) but for an approximate representation of an empty set as a quantity (evidence for this being based on the presence of a “distance effect”).

• Line 279: References here are mostly primate-centric, please add also something more comparative (e.g. Ferrigno, S. & Cantlon, J.F. (2017). Evolutionary constraints on the emergence of human mathematical concepts. In J. Kaas (Ed.), Evolution of Nervous Systems (2nd ed. vol. 3) (pp. 511-521). Oxford: Elsevier; Vallortigara, G. (2014). Foundations of Number and Space Representations in Non-Human Species. In “Evolutionary Origins and Early Development of Number Processing”, pp. 35-66 (Eds., D.C. Geary, D.B. Bearch, K. Mann Koepke), Elsevier, New York; Vallortigara, G. (2012). Core knowledge of object, number, and geometry: A comparative and neural approach. Cognitive Neuropsychology, 29: 213-236; Brannon, E.M. & Merritt, D. (2011) Evolutionary foundations of the Approximate Number System. In Space, Time, and Number in the Brain: searching for the foundations of mathematical thought. Editors, Dehaene, S., & Brannon, E.M. Elsevier (2011).

Review form: Reviewer 3

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
Manuscript ID RSPB-2019-0238
"Symbol and numerosity matching in the honeybee (Apis mellifera)"

SUMMARY
The authors trained half of the bees to associate specific quantities with symbols, and the other half to associate symbols with a specific quantity. They used two symbols (N and an inverted T) and two quantities (2 and 3). The appearance of the symbols was always identical, whereas the item shapes and sizes of the quantity displays varied. Both groups of bees learned the task; the bees were able to transfer the rule to differently colored quantity displays, as well as to numerical controls (equal surface area and equal item size). Finally, the bees were tested on the reverse of the learned association, which the bees in both groups failed. The authors conclude that an insect brain is able to build a symbolic representation of quantity.
This is another interesting and timely study by Howard et al. about the numerical capabilities of bees. The data are novel and the manuscript is well written. It would have been nice to demonstrate the bees' semantic association capabilities beyond the minimum number of quantities (two and three), but this might be difficult to achieve in bees.

MAJOR CONCERN
1. In the Methods the Testing paragraph, the authors write that the non-reinforced transfer,

numerical controls test, and reversal tests were directly following each other. In other papers from the lab, they wrote that they inserted a few reinforced trials with the known training stimuli to keep the bees motivated. If this was the case in these experiments too, please state it in the manuscript. If this was not the case, the negative results from the reversal test could potentially be a result of unmotivated bees.
2. I am not convinced that bees (or any animal, for that matter) use “symbols” in a strict sense. In semiotics, a symbol is the most complex type of a sign, such as words or numbers that are part of a combinatorial symbol system. Other, less demanding signs, would be “icons” and “indices” (according to C.S. Peirce and T. Deacon). The bees (just as any other animal) learned to use visual shapes as indices (not as symbols). If the authors want to avoid this semiotic issues (e.g. Smith and Harper, J. Theor. Biol., 1995; Nieder, CurrOpNeurobiol, 2009), I suggest they replace “symbol” with “sign”, because a sign would be the overarching term for any object/event that stands for something else, in this study for set size. This also applies to the title.

MINOR COMMENTS
- Line 108: it should be “had” not “has”
- Line 144: the term “Learning test” is confusing (is it learning, or is it testing?). Same goes for the figures in which the same term is used. I would name it something along the lines of “Numerical control”
- Fig 3 a,b) a GLMM through 5 data points seems a bit overkill to me; it is not adding value.

Decision letter (RSPB-2019-0238.R0)

18-Mar-2019

Dear Ms Howard:

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. 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,
Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
This manuscript has now been critiqued in detail by three expert reviewers. I was favorable towards the potential of this manuscript, and so were all three reviewers. However, all reviewers recommended revisions, with two out of three reviewers recommending major revisions. Overall, there is a number of requests for sometimes deep revisions – and which I would encourage the authors to take seriously.
I will not repeat all reviewers' comments here (but all should be taken into account). The majority of reviewers considers the potential of the manuscript high, too, but the majority also considered the current manuscript lacking in some aspects. R1 asked for more data to clarify between current interpretational possibilities (esp. with regards to extinction effects), and R1 also asks for the inclusion of more data detail in the MS. R2 also asks for more data, in this case for a control experiments with non-numerical stimuli. R2 also points out a couple of other problems, including potential problems with the cleaning procedure used. R3 notes the absence of a few explanations, which should be provided as these may interfere with the correct interpretation of the data. R3 also urges to use “sign” instead of “symbol”.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
In their study ‘Symbol and numerosity matching in the honeybee (Apis mellifera)’ Howard et al. investigate the honeybee’s capability of associating number value with abstract symbols and vice versa in a previously established delayed-match-to-sample assay. While the bees manage to associate both symbol with number (2 or 3) and number with symbol, they seem to fail in the subsequent reversal test. Pushing the boundaries of associative task complexity in insects is best done in honeybees, and the authors already published a couple of fascinating papers in this regard in recent years, where they challenged the cognitive - and especially the numerical - capabilities of bees.
Here, Howard et al. provide an exciting study that adds a novel nuance to our knowledge about animal cognition and thus has general significance to biologists. The manuscript is well written and sets the question in adequate biological context; the results would benefit from showing individual data points, but the experiments are mostly well-designed and controlled. The manuscript does fall short on providing some relevant information about the time course of the experiments (see below). Maybe related to this is my biggest concern: this work completely

disregards the effect of extinction by unreinforced choices during testing. While this is unavoidable for individual tests, it could heavily affect the outcome of different test paradigms that are performed sequentially (as done here). In this light I recommend the manuscript for publication after addressing this point - either by convincing me that my concern is unfounded or by additional experiments (see below).

Minor comments:
line 68-72: please comment on the time course of the experiment.
line 128: showed → shown
line 137: What was the time frame of the experiments? How long did each part last? Please also add information about the delay between training and testing the bees, and between the different test situations. (either here or in line 68-72)

Figure 2:
* spelling error : 'radndom → random' in both a) and b)
* please spell out 'SA' in figure

Major comments:
Figure 3:
Only 20 bees were tested altogether, which is consistent with their previous papers in the same setup, but in stark contrast to the amount of bees the same authors tested in their recent paper in Current Zoology for one figure alone (n=138). This low number however allows for showing individual choice probabilities for the 10 bees for each data point in 3a)-d) (i.e. with beeswarm plot in R).
If the authors also connect the data for each animal across trials in 3a) and 3b), individual acquisition performances can be observed (ideally in different colors). Taken together, this adds another layer of information and represents the acquired data more transparently. Along those lines, why did the authors use a linear fit? Although this was done in previous publications too, it was and still is misleading. Acquisition of an association is not necessarily linear (more often exponential or sigmoidal), and a nonlinear fit would also be more appropriate, given the low number of tested individuals. Please use a higher-order polynomial or other, more suitable fit to describe the data.
If I interpret the experimental layout in Fig. 2 correctly, the two groups were further split to test learning with equal surface area or with equal element size. This would result in an n=5 for bars 2 & 3 in fig. 3c) and d), respectively. Please show individual data points for each animal's proportion of correct choices, and add the corresponding count of animals to the legend of 3c) and d).

As mentioned before, please add more information about the time course of the experiments. As soon as bees are tested, the learned association also becomes extinguished, since each unreinforced choice represents an extinction event. In case of the reversal test, the bees already underwent 20 extinction trials (during transfer test and learning test), which would inevitably influence performance.

Please demonstrate that this particular appetitive-aversive differential delayed-match-to-sample training paradigm leads to robust long-lasting memory. Also, please demonstrate that there was a sufficient period for spontaneous recovery of memory between test types. Otherwise, the lack of performance in the reversal test could be explained by extinction.
Since the authors put sufficient emphasis on the lack of reversal learning performance in bees in their discussion, they should 1) either provide the aforementioned information or 2) add more experiments where reversal tests are performed immediately after the 50 training trials.

Referee: 2

Comments to the Author(s)
This is an interesting paper. However, some issues need clarification before any final decision about the suitability of this manuscript for publication can be reached. In particular, it seems to me that a control experiment with non-numerical stimuli would be necessary to a proper interpretation of the bees failure at reversing the symbol-number vs. number/symbol task.

Specific comments

-P 5, line 92: the 20% ethanol concentration to clean the apparatus might not be enough to remove all the olfactory cues. Literature supports a 30% or higher ethanol concentration:
https://royalsocietypublishing.org/doi/full/10.1098/rspb.2017.2278
https://link.springer.com/article/10.1007/s10071-017-1086-6
http://science.sciencemag.org/content/360/6393/1124.abstract?casa_token=i6BqQcIG03cAAAA A:tDQxpxNPfKiplbIitH8_lfMauE4jE90PaWpQm4-39ilfwYd81LilG9zI2qPnuQ9B2aO3grz93zNGZg

-P 6, line 115: It is written that the 3) set was used as a control for the continuous variables that may covary with numerosity. However, how these controls were made is not specified. In particular, it would be useful to know whether all the stimuli in this group were simultaneously controlled for all the features or not. In the first case, it would be very difficult to be able to control all the continuous variables at a time (for example, it is very difficult to simultaneously control for the density and the convex hull, as well as for the area and the line length, since these two variables are inversely related).

-P 8: Looking at the graphs and according to the variability showed by the error bars, my concern is whether there are differences between the performances of the two groups, both during the training phase and the test phase.

-P 10, Discussion: before arguing for an inability of honeybees to reverse a symbol-number association, a control might be needed. Bees should be trained with a DMTS where non-numerical stimuli will be used (colors and shapes or different symbols) in order to establish a causal association (e.g. if yellow than circle; if blue than triangle): Then, subjects should be tested in a reverse test. The hypothesis is that if bees are able to do the reverse task when presented with non-numerical DMTS, then their inability to reverse the symbol-number association is strictly connected to the number matching. In contrast, if bees are not able to reverse their response in a non-numerical DMTS, then a general inability to deal with operational schemes can be hypothesized.

Line 32: I would not use ‘numbers’ here but ‘numerosities’. In general, I think the authors should make an effort to provide the reader (who is not necessarily a specialist on numerical cognition) that the type of evidence available for non-human animals for quantity (discrete numerousness) is relative to an approximate, analogous representation (so-called ANS, Approximate Number System) and that the association which has been studied in work with e.g. monkeys (see for instance reference 19) refers to a link between these non-symbolic, approximate numerousness representation and symbols. See e.g. for a recent review Vallortigara, G. (2017). An animal’s sense of number. In “The nature and Development of Mathematics. Cross Disciplinary Perspective on Cognition, Learning and Culture” (Adams, J.W., Barmby P., Mesoudi, A., eds.), pp. 43-65, Routledge, New York.

Line 53-54: Here, again I am worried that a wrong message would be conveyed to the non-specialist reader, for evidence for “zero” in bees and other non-human animals is not relative to number zero (that would be possible only in association with a symbol) but for an approximate representation of an empty set as a quantity (evidence for this being based on the presence of a "distance effect").

• Line 279: References here are mostly primate-centric, please add also something more comparative (e.g. Ferrigno, S. & Cantlon, J.F. (2017). Evolutionary constraints on the emergence of human mathematical concepts. In J. Kaas (Ed.), Evolution of Nervous Systems (2nd ed. vol. 3) (pp. 511-521). Oxford: Elsevier; Vallortigara, G. (2014). Foundations of Number and Space Representations in Non-Human Species. In "Evolutionary Origins and Early Development of Number Processing", pp. 35-66 (Eds., D.C. Geary, D.B. Bearch, K. Mann Koepke), Elsevier, New York; Vallortigara, G. (2012). Core knowledge of object, number, and geometry: A comparative and neural approach. Cognitive Neuropsychology, 29: 213-236; Brannon, E.M. & Merritt, D. (2011) Evolutionary foundations of the Approximate Number System. In Space, Time, and Number in the Brain: searching for the foundations of mathematical thought. Editors, Dehaene, S., & Brannon, E.M. Elsevier (2011).

Referee: 3

Comments to the Author(s)
Manuscript ID RSPB-2019-0238
"Symbol and numerosity matching in the honeybee (Apis mellifera)"

SUMMARY
The authors trained half of the bees to associate specific quantities with symbols, and the other half to associate symbols with a specific quantity. They used two symbols (N and an inverted T) and two quantities (2 and 3). The appearance of the symbols was always identical, whereas the item shapes and sizes of the quantity displays varied. Both groups of bees learned the task; the bees were able to transfer the rule to differently colored quantity displays, as well as to numerical controls (equal surface area and equal item size). Finally, the bees were tested on the reverse of the learned association, which the bees in both groups failed. The authors conclude that an insect brain is able to build a symbolic representation of quantity.
This is another interesting and timely study by Howard et al. about the numerical capabilities of bees. The data are novel and the manuscript is well written. It would have been nice to demonstrate the bees' semantic association capabilities beyond the minimum number of quantities (two and three), but this might be difficult to achieve in bees.

MAJOR CONCERN
1. In the Methods the Testing paragraph, the authors write that the non-reinforced transfer, numerical controls test, and reversal tests were directly following each other. In other papers from the lab, they wrote that they inserted a few reinforced trials with the known training stimuli to keep the bees motivated. If this was the case in these experiments too, please state it in the manuscript. If this was not the case, the negative results from the reversal test could potentially be a result of unmotivated bees.
2. I am not convinced that bees (or any animal, for that matter) use "symbols" in a strict sense. In semiotics, a symbol is the most complex type of a sign, such as words or numbers that are part of a combinatorial symbol system. Other, less demanding signs, would be "icons" and "indices" (according to C.S. Peirce and T. Deacon). The bees (just as any other animal) learned to use visual shapes as indices (not as symbols). If the authors want to avoid this semiotic issues (e.g. Smith and Harper, J. Theor. Biol., 1995; Nieder, CurrOpNeurobiol, 2009), I suggest they replace "symbol" with "sign", because a sign would be the overarching term for any object/event that stands for something else, in this study for set size. This also applies to the title.

MINOR COMMENTS
- Line 108: it should be “had” not “has”
- Line 144: the term “Learning test” is confusing (is it learning, or is it testing?). Same goes for the figures in which the same term is used. I would name it something along the lines of “Numerical control”
- Fig 3 a,b) a GLMM through 5 data points seems a bit overkill to me; it is not adding value.

Author's Response to Decision Letter for (RSPB-2019-0238.R0)

See Appendix A.

RSPB-2019-0238.R1 (Revision)

Review form: Reviewer 1

Recommendation
Accept with minor revision (please list 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.
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
The authors addressed my raised concerns and answered my questions in a sufficient way, and I can recommend publication of this manuscript. One small typo is still hiding in Figure 2a,b: 'Transfer' instead of 'Tranfer'.

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?
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?
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
I think the authors have addressed adequately my concerns.
Please check Ref 57, the correct reference is
Vallortigara, G. (2015). Foundations of Number and Space Representations in Non-Human Species. In "Evolutionary Origins and Early Development of Number Processing", pp. 35-66 (Eds., D.C. Geary, D.B. Bearch, K. Mann Koepke), Elsevier, New York.

Decision letter (RSPB-2019-0238.R1)

10-May-2019

Dear Ms Howard

I am pleased to inform you that your manuscript RSPB-2019-0238.R1 entitled "Symbolic representation of numerosity by honeybees (*Apis mellifera*): Matching characters to small quantities" 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,

Proceedings B
mailto:proceedingsb@royalsociety.org

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
The authors addressed my raised concerns and answered my questions in a sufficient way, and I can recommend publication of this manuscript. One small typo is still hiding in Figure 2a,b: 'Transfer' instead of 'Tranfer'.

Referee: 2

Comments to the Author(s)
I think the authors have addressed adequately my concerns.
Please check Ref 57, the correct reference is
Vallortigara, G. (2015). Foundations of Number and Space Representations in Non-Human Species. In “Evolutionary Origins and Early Development of Number Processing”, pp. 35-66 (Eds., D.C. Geary, D.B. Bearch, K. Mann Koepke), Elsevier, New York.

Author's Response to Decision Letter for (RSPB-2019-0238.R1)

See Appendix B.

Decision letter (RSPB-2019-0238.R2)

13-May-2019

Dear Ms Howard

I am pleased to inform you that your manuscript entitled "Symbolic representation of numerosity by honeybees (*Apis mellifera*): Matching characters to small quantities" 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

Your article has been estimated as being 10 pages long. Our Production Office will be able to confirm the exact length at proof stage.

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.

Paper charges
An e-mail request for payment of any related charges will be sent out shortly. 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,

Proceedings B
mailto: proceedingsb@royalsociety.org

Appendix A

Dear Editor,

We would like to thank you for the opportunity to resubmit our paper after receiving the following reviews. We also thank you and the three reviewers for your time and effort during the review process. The comments we have received are insightful and we believe have improved the quality of our manuscript. We have incorporated most suggested changes and where we have not made changes, we ensured to explain why this was the case. Specifically, we have given more information about the use of refresher trials to counteract the extinction effect mentioned by two of the reviewers, we have included more detail on the methods section as requested, we have edited our statement about the cleaning of equipment, and edited the entire manuscript as requested by R3 using new terminology. While we did not conduct the suggested experiment by R2, we believe it is an excellent suggestion and have thus included a detailed discussion on their comment in the discussion section. Indeed we believe our study, if published in Proceedings B, will promote a lot of new research on this contemporary question of numeric processing in different animals, including honeybees. Our responses are indented and numbered following each comment from the reviewers.

We look forward to hearing from you about the suitability of our edited manuscript for publication in Proceedings of the Royal Society B.

Kind regards,

Scarlett Howard, Aurore Avarguès-Weber, Jair Garcia, Andrew Greentree, and Adrian Dyer.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

In their study 'Symbol and numerosity matching in the honeybee (Apis mellifera)' Howard et al. investigate the honeybee's capability of associating number value with abstract symbols and vice versa in a previously established delayed-match-to-sample assay. While the bees manage to associate both symbol with number (2 or 3) and number with symbol, they seem to fail in the subsequent reversal test. Pushing the boundaries of associative task complexity in insects is best done in honeybees, and the authors already published a couple of fascinating papers in this regard in recent years, where they challenged the cognitive - and especially the numerical - capabilities of bees.

Here, Howard et al. provide an exciting study that adds a novel nuance to our knowledge about animal cognition and thus has general significance to biologists. The manuscript is well written and sets the question in adequate biological context; the results would benefit from showing individual data points, but the experiments are mostly well-designed and controlled. The manuscript does fall short on providing some relevant information about the time course of the experiments (see below). Maybe related to this is my biggest concern: this work completely disregards the effect of extinction by unreinforced choices during testing. While this is unavoidable for individual tests, it could heavily affect the outcome of different test paradigms that are performed sequentially (as done here). In this light I recommend the manuscript for publication after addressing this point - either by convincing me that my concern is unfounded or by additional experiments (see below).

Author response 1:

We thank the reviewer for their kind comments and support of our study. We have attempted to adequately address all of the reviewer's concerns regarding the manuscript and provided more details. As for their major concern of an extinction effect during testing, we unfortunately omitted to state within the original manuscript that refresher trials were conducted between each test as is standard procedure for experiments with multiple tests on honeybees. This point is detailed below and we thank the reviewer for raising the important issue.

Minor comments:

line 68-72: please comment on the time course of the experiment.

Author response 2:

Each experiment lasted for approximately 3 – 5 hours per bee. We have now added this information into the methods section (lines 87-88; 148-149).

line 128: showed → shown

Author response 3:

We thank the reviewer for bringing this typo to our attention, this has now been corrected.

line 137: What was the time frame of the experiments? How long did each part last? Please also add information about the delay between training and testing the bees, and between the different test situations. (either here or in line 68-72)

Author response 4:

We have now included the requested information:

Lines 171-172: "*The testing phase lasted approximately 30 minutes to one hour per bee.*"

Lines 148-149: "*The training phase lasted for approximately two to four hours per bee.*"

Lines 87-88: "*Each experiment lasted for approximately 3 – 5 hours per bee.*"

Lines 152-156: "*After the 50 training choices were completed, bees were removed from the apparatus following their final choice and given sucrose until they became satiated and returned to the hive. During the bee's absence from the experimental area, we prepared for the testing phase by placing test stimuli and clean landing poles into the Y-maze. Once bees returned from the hive (typically < 5 minutes), individuals underwent four tests in the following order:*"

Figure 2:
* spelling error : 'radndom → random' in both a) and b)5 Author response 2:

Author response 5:

We thank the reviewer for bringing this to our attention and we have amended Figure 2.

* please spell out ‘SA’ in figure

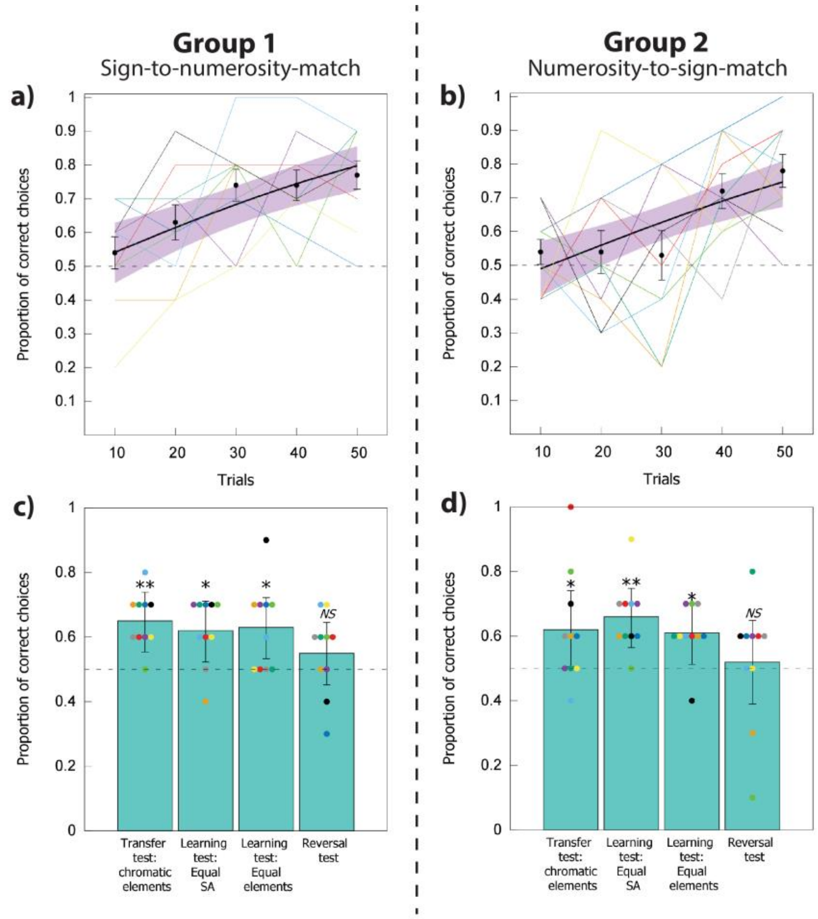
Author response 6:

We thank the reviewer for bringing this to our attention and we have amended Figure 2.

Major comments:
Figure 3:
Only 20 bees were tested altogether, which is consistent with their previous papers in the same setup, but in stark contrast to the amount of bees the same authors tested in their recent paper in Current Zoology for one figure alone (n=138). This low number however allows for showing individual choice probabilities for the 10 bees for each data point in 3a)-d) (i.e. with beeswarm plot in R).

Author response 7:

We thank the reviewer for this suggestion. We have now plotted the individual choices per bee using a bee swarm plot. See Figure S2. The colour of each line in Figure S2ab and each dot in S2cd corresponds to an individual within each group. See below:



*"Figure S2: Results of the training and testing phases for the group of bees trained to match a sign with a quantity (Group 1; n = 10) and for the group of bees trained to match a quantity with a sign (Group 2; n = 10). a-b) Performance during the training phase for (a) Group 1 and (b) Group 2. Dashed line at 0.5 indicates chance level performance. Solid black line represents a function describing the training phase of n = 10 bees as modelled by a generalised linear mixed model (GLMM). Points (closed circles) along the curve indicate the observed mean ± 95 % CIs (purple) of correct choices for the bees. Increase in performance during the training phase was significant. Individual performances of bees are shown as coloured lines. Colours correspond to individuals in the same group in the test phase directly below. c-d) Performance during the testing phases for both (c) Group 1 and (d) Group 2. N = 10 for both test groups. Dashed line at 0.5 indicates chance level performance. Significance from chance level performance is indicated by * ≥ 0.05, ** ≥ 0.01, *** ≥ 0.001, NS > 0.05. Data shown are means ± 95 % CI boundaries for all tests. Coloured dots represent proportion of correct choices during the test for each individual bee. Colour corresponds to the specific individual bee in each group for the training and test phases."*

The very large number of bees tested in the Current Zoology study was because that research investigated spontaneous preferences which are typically weak but only require very short protocol to test; and our number of tested bees was

informed by pilot experiments and statistical power modelling. For numerosity studies, very long and meticulous experiments training individual bees are required, and our sample size is chosen to be consistent with previous studies on bees, and other animal psychophysics studies, so that results from respective studies are comparable.

If the authors also connect the data for each animal across trials in 3a) and 3b), individual acquisition performances can be observed (ideally in different colors). Taken together, this adds another layer of information and represents the acquired data more transparently. Along those lines, why did the authors use a linear fit? Although this was done in previous publications too, it was and still is misleading. Acquisition of an association is not necessarily linear (more often exponential or sigmoidal), and a nonlinear fit would also be more appropriate, given the low number of tested individuals. Please use a higher-order polynomial or other, more suitable fit to describe the data.

Author response 8:

There are several problems with the solution proposed by the reviewer. Firstly, although a polynomial or non-linear model can apparently fit the overall shape of the acquisition data, such a technique still assumes that the responses modelled by the function follow a normal distribution which is not true for the type of response, i.e. correct or incorrect choice, obtained from our experiments. Choice data (correct or incorrect) is accurately modelled by a Bernoulli distribution, a special case of the binomial distribution. Differently from the normal distribution, which is unbounded thus allowing negative numbers and probability values higher than 1, the Bernoulli distribution\accurately models the probability of an outcome as it is bounded between 0 and 1.

The generalised linear model we implemented uses an exponential function, the logit, to link the Bernoulli distributed response with the linear covariates. As the reviewer pointed out, the exponential terms in the logit function produces a typical sigmoidal function. Finally, the adequacy of our analysis to model the data can be assessed by the value of the model over dispersion (Group 1: overdispersion value = 1.197; Group 2: overdispersion value = 1.278) which is about 0.2 units above the theoretical values of 1.0 expected for an ideal Bernoulli distribution. These results suggest a good fit between the resulting model and the observed data.

If I interpret the experimental layout in Fig. 2 correctly, the two groups were further split to test learning with equal surface area or with equal element size. This would result in an n=5 for bars 2 & 3 in fig. 3c) and d), respectively. Please show individual data points for each animal’s proportion of correct choices, and add the corresponding count of animals to the legend of 3c) and d).

Author response 9:

The groups were not split to train to equal surface area or equal element size, rather all bees were trained with both Sets 1 (equal surface area) and Set 2 (equal element size). Thus, bees experienced both of these types of stimuli during training which were randomised per choice. Thus there is still *n* = 10 for both groups and they are not split any further. Please refer to lines 145-150. The figure legend has been edited to further detail the number of bees per group.

We have also used a bee swarm plot to show the individual proportions for both training and test phases, where the colour of a line in Figure S2ab corresponds to the colour of the same individual in Figure S2cd (see above in Author Response 7).

As mentioned before, please add more information about the time course of the experiments. As soon as bees are tested, the learned association also becomes extinguished, since each unreinforced choice represents an extinction event. In case of the reversal test, the bees already underwent 20 extinction trials (during transfer test and learning test), which would inevitably influence performance.

Author response 10:

The reviewer raises an important issue here. Motivating refresher trials using the training set stimuli were performed between each test for a single bout (consisting of 3 – 6 choices) per test. We did not include this in the initial manuscript and apologise for the omission. The use of refresher trials in between unconditioned tests counters the effect of extinction or extinguished learning during testing phases. We have included this in the methods section (lines 170-177):

"*The tests were not randomised for order as the reversal test could have potentially confused bees with the switching of stimuli location, thus it was conducted last so that this would not impact the bees' performance on other tests. The testing phase lasted approximately 30 minutes to one hour per bee. Between each test we conducted refresher choices to maintain bee motivation and counter extinction effects of the unconditioned tests. The refresher choices consisted of bees being presented with the training set stimuli with appetitive and aversive outcomes (as in the training phase) for 3 – 6 choices (a single bout).*"

We have also now included the relevant time information as requested, which was detailed above in Author Response 4.

Please demonstrate that this particular appetitive-aversive differential delayed-match-to-sample training paradigm leads to robust long-lasting memory. Also, please demonstrate that there was a sufficient period for spontaneous recovery of memory between test types. Otherwise, the lack of performance in the reversal test could be explained by extinction.
Since the authors put sufficient emphasis on the lack of reversal learning performance in bees in their discussion, they should 1) either provide the aforementioned information or 2) add more experiments where reversal tests are performed immediately after the 50 training trials.

Author response 11:

As mentioned above (Author response 10), refresher tests were conducted between each test for one bout (between 3 – 6 choices), thus counteracting the extinction effect and keeping bees motivated for the experiment. It has previously been established in both bumblebees [Dyer and Chittka 2004 Naturwissenschaften 91 (5), 224-227] and honeybees [Dyer and Garcia 2014 Insects 5 (3), 629-638] that differential conditioning leads to the formation of a long term memory that lasts for greater than one day.

Referee: 2

Comments to the Author(s)

This is an interesting paper. However, some issues need clarification before any final decision about the suitability of this manuscript for publication can be reached. In particular, it seems to me that a control experiment with non-numerical stimuli would be necessary to a proper interpretation of the bees failure at reversing the symbol-number vs. number/symbol task.

Author response 12:

We thank the reviewer for their suggestions and kind comment. We have detailed our responses and changes based on their review below.

Specific comments

-P 5, line 92: the 20% ethanol concentration to clean the apparatus might not be enough to remove all the olfactory cues. Literature supports a 30% or higher ethanol concentration:
https://royalsocietypublishing.org/doi/full/10.1098/rspb.2017.2278
https://link.springer.com/article/10.1007/s10071-017-1086-6
http://science.sciencemag.org/content/360/6393/1124.abstract?casa_token=i6BqQcIG03cAAAAA:tDQxpxNPfKiplbIitH8_lfMauE4jE90PaWpQm4-39ilfwYd81LilG9zI2qPnuQ9B2aO3grz93zNGZg

Author response 13:

The reviewer is correct and we apologise for this, the correct amount of ethanol was 30 % as in our previous studies which the reviewer cites above. We have amended this in the manuscript.

-P 6, line 115: It is written that the 3) set was used as a control for the continuous variables that may covary with numerosity. However, how these controls were made is not specified. In particular, it would be useful to know whether all the stimuli in this group were simultaneously controlled for all the features or not. In the first case, it would be very difficult to be able to control all the continuous variables at a time (for example, it is very difficult to simultaneously control for the density and the convex hull, as well as for the area and the line length, since these two variables are inversely related).

Author response 14:

As noted by the reviewer, it is not possible to simultaneously control all continuous variables at the same time with complex numerical stimuli; thus the approach we used was a standard psychophysics methodology of both

randomising between the sets of training stimuli and then also employing the completely novel chromatic factor to show true transfer of numerical ability. We have thus amended the description in the 'Stimuli' subsection of the Methods section.

-P 8: Looking at the graphs and according to the variability showed by the error bars, my concern is whether there are differences between the performances of the two groups, both during the training phase and the test phase.

Author response 15:

We thank the reviewer for this suggestion. We have run the analyses and put the results in to the paper. See lines 196-202 and lines 232-237. There were no significant differences between the groups in training or tests:

"We also fitted a model with trial number, group, and an interaction between trial number and group as continuous predictors, and subject as a random factor to account for repeated choices of individual bees to determine if the learning curve of the groups were significantly different. There was no significant effect of an interaction between trial number and group (z = -0.160, P = 0.873), nor a significant effect group in this model (z = -0.700, P = 0.484) demonstrating that both groups learnt at an equal rate."

"We fitted a model for each of the four tests to determine whether bees within the two groups performed significantly different from each other. We employed a GLMM including the intercept term and group as fixed factors and subject as a random term to account for the repeated measures. There was no effect of the group factor on any of the four tests ($P \geq 0.556$), thus demonstrating that performance on each test between the groups was not statistically different."

-P 10, Discussion: before arguing for an inability of honeybees to reverse a symbol-number association, a control might be needed. Bees should be trained with a DMTS where non-numerical stimuli will be used (colors and shapes or different symbols) in order to establish a causal association (e.g. if yellow than circle; if blue than triangle): Then, subjects should be tested in a reverse test. The hypothesis is that if bees are able to do the reverse task when presented with non-numerical DMTS, then their inability to reverse the symbol-number association is strictly connected to the number matching. In contrast, if bees are not able to reverse their response in a non-numerical DMTS, then a general inability to deal with operational schemes can be hypothesized.

Author response 16:

This is an excellent idea and one we believe is a very valid question. We do not currently have the possibility to test it and we believe this is a follow-up question to our study. We nevertheless have now included a discussion of this factor in the discussion. We do not know from this experiment whether the bee's inability to reverse the association is from the task of reversing an association itself or specifically the numerical element of the task. Thus we give this important point a paragraph of discussion in our manuscript (see lines 278-307):

"*The results demonstrate that while bees can perform both a sign-to-numerosity-match and a numerosity-to-sign-match, they are unable to reverse the task if it is only experienced in one direction. Interestingly, we do not know whether this inability to reverse the task is due to the numerical aspect of the task, or whether they are unable to reverse an association between any two stimuli within a DMTS framework. Interestingly, a similar result is observed in Gross et al. (2009), where honeybees were trained to match two patterns based on only the quantity cue using a DMTS procedure, as in the current study. In their study, bees were given training to match patterns containing either two or three elements, and thus the bees experienced the task in both directions. They were trained to match a two element pattern to a second two element pattern in the presence of a three element pattern. Bees were also trained to match a three element pattern to a separate three element pattern in the presence of a two element pattern. Bees experienced matching patterns containing either two or three elements, thus performing the matching task in both directions and leading to the ability to match novel patterns containing two or three elements, even when in the presence of an incorrect novel pattern containing four elements. However, bees in this previous study were unable to consistently match novel patterns of four, five, or six. Perhaps this was due to their inability to discriminate between numbers in the AMS/ANS range when trained with appetitive differential conditioning, or perhaps this was because bees are not able to perform a novel DMTS task without first learning the specific association in the correct direction. This suggests that the inability of bees to perform the reversal test in the current study may be due to the numerical nature of the task and the storage of numerical information, as discussed below, but it could also be that bees are unable to reverse any association which is presented in a DMTS task, thus an inability to process operational schemes or understand a reversible association [39]. Giurfa et al. (2001) demonstrated that in a DMTS task bees were able to learn to match a sample visual or olfactory cue. Furthermore, bees could extrapolate this task in tests demonstrating neural flexibility in DMTS tasks. For example, if bees had been trained to match odours and were then tested to match novel visual stimuli such as colours, they succeeded despite having no training on colour matching [50]. The results from Giurfa et al. (2001) suggest that the inability of the bees in the current study to perform a reversal test is due to the numerical nature of the task and not the DMTS task itself.*"

Line 32: I would not use ‘numbers’ here but ‘numerosities’. In general, I think the authors should make an effort to provide the reader (who is not necessarily a specialist on numerical cognition) that the type of evidence available for non-human animals for quantity (discrete numerousness) is relative to an approximate, analogous representation (so-called ANS, Approximate Number System) and that the association which has been studied in work with e.g. monkeys (see for instance reference 19) refers to a link between these non-symbolic, approximate numerousness representation and symbols. See e.g. for a recent review Vallortigara, G. (2017). An animal’s sense of number. In “The nature and Development of Mathematics. Cross Disciplinary Perspective on Cognition, Learning and Culture” (Adams, J.W., Barmby P., Mesoudi, A., eds.), pp. 43-65, Routledge, New York.

Author response 17:

The reviewer raises an important issue here and thus we have changed ‘number’ to ‘numerosity’ in this place and also throughout the manuscript where it is required.

We have included a brief introduction to the ANS and symbolic/non-symbolic number systems as suggested as well (lines 50-62):

“*There are two well accepted mechanisms for numerical discrimination in humans and non-human animals which are involved for different number ranges [24-28]. The first is known as subitizing or the object file system, the quick and accurate estimation of four and fewer objects. The second mechanism is the approximate number system (ANS), also known as the analog magnitude system (AMS) for quantifying numerosities above four objects. Animals which are able to process quantities above four are considered to be using the ANS/AMS mechanism of numerosity judgement. In humans, the ANS is thought to be an ancient evolutionary foundation, shared with non-human animals, for our ability to perform mathematics and use symbolic numerositiy mechanisms [29-32]. While symbolic number systems, such as the use of Arabic numerals allows us to easily discriminate between two numbers, non-symbolic number systems, such as the ANS, show that ratio-dependent number discrimination exists. Thus, humans and non-human animals share the evolutionary ancient quantification system known as the approximate number system.*”

Line 53-54: Here, again I am worried that a wrong message would be conveyed to the non-specialist reader, for evidence for “zero” in bees and other non-human animals is not relative to number zero (that would be possible only in association with a symbol) but for an approximate representation of an empty set as a quantity (evidence for this being based on the presence of a “distance effect”).

Author response 18:

We have now changed this line to "Interestingly bees spontaneously placed an empty set at the lower end of the numerical scale without specific training on the task, thereby demonstrating the ability to quantify an empty set as being lower than numbers one to six." – see lines 68-70.

Line 279: References here are mostly primate-centric, please add also something more comparative (e.g. Ferrigno, S. & Cantlon, J.F. (2017). Evolutionary constraints on the emergence of human mathematical concepts. In J. Kaas (Ed.), Evolution of Nervous Systems (2nd ed. vol. 3) (pp. 511-521). Oxford: Elsevier; Vallortigara, G. (2014). Foundations of Number and Space Representations in Non-Human Species. In "Evolutionary Origins and Early Development of Number Processing", pp. 35-66 (Eds., D.C. Geary, D.B. Bearch, K. Mann Koepke), Elsevier, New York; Vallortigara, G. (2012). Core knowledge of object, number, and geometry: A comparative and neural approach. Cognitive Neuropsychology, 29: 213-236; Brannon, E.M. & Merritt, D. (2011) Evolutionary foundations of the Approximate Number System. In Space, Time, and Number in the Brain: searching for the foundations of mathematical thought. Editors, Dehaene, S., & Brannon, E.M. Elsevier (2011).

Author response 19:

We thank the reviewer for their suggestion and have now included these references in line 348.

<u>Referee: 3</u>

Comments to the Author(s)
Manuscript ID RSPB-2019-0238
“Symbol and numerosity matching in the honeybee (Apis mellifera)”

SUMMARY
The authors trained half of the bees to associate specific quantities with symbols, and the other half to associate symbols with a specific quantity. They used two symbols (N and an inverted T) and two quantities (2 and 3). The appearance of the symbols was always identical, whereas the item shapes and sizes of the quantity displays varied. Both groups of bees learned the task; the bees were able to transfer the rule to differently colored quantity displays, as well as to numerical controls (equal surface area and equal item size). Finally, the bees were tested on the reverse of the learned association, which the bees in both groups failed. The authors conclude that an insect brain is able to build a symbolic representation of quantity.
This is another interesting and timely study by Howard et al. about the numerical capabilities of bees. The data are novel and the manuscript is well written. It would have been nice to demonstrate the bees’ semantic association capabilities beyond the minimum number of quantities (two and three), but this might be difficult to achieve in bees.

> **Author response 20:**
>
> **We thank the reviewer for their support and kind comments on our study. We have hopefully answered all questions and made the requested changes, which we detail below.**

MAJOR CONCERN
1. In the Methods the Testing paragraph, the authors write that the non-reinforced transfer, numerical controls test, and reversal tests were directly following each other. In other papers from the lab, they wrote that they inserted a few reinforced trials with the known training stimuli to keep the bees motivated. If this was the case in these experiments too, please state it in the manuscript. If this was not the case, the negative results from the reversal test could potentially be a result of unmotivated bees.

> **Author response 21:**
>
> **The reviewer raises an important issue here. Motivating refresher trials using the training set stimuli were performed between each test for a single bout (consisting of 3 – 6 choices) per test. We forgot to include this in the initial manuscript and apologise for the omission. We have now included this in the methods section (lines 172-177):**

"The testing phase lasted approximately 30 minutes to one hour per bee. Between each test we conducted refresher choices to maintain bee motivation and counter extinction effects of the unconditioned tests. The refresher choices consisted of bees being presented with the training set stimuli with appetitive and aversive outcomes (as in the training phase) for 3 – 6 choices (a single bout)."

2. I am not convinced that bees (or any animal, for that matter) use "symbols" in a strict sense. In semiotics, a symbol is the most complex type of a sign, such as words or numbers that are part of a combinatorial symbol system. Other, less demanding signs, would be "icons" and "indices" (according to C.S. Peirce and T. Deacon). The bees (just as any other animal) learned to use visual shapes as indices (not as symbols). If the authors want to avoid this semiotic issues (e.g. Smith and Harper, J. Theor. Biol., 1995; Nieder, CurrOpNeurobiol, 2009), I suggest they replace "symbol" with "sign", because a sign would be the overarching term for any object/event that stands for something else, in this study for set size. This also applies to the title.

Author response 22:

We have addressed this comment and replaced the use of "symbol" with "sign" in all places throughout the manuscript and Figures. We have edited the title to bridge the gap between symbolic representation and what our study shows: "Symbolic representation of numerosity by honeybees (*Apis mellifera*): Matching characters to small quantities".

MINOR COMMENTS

- Line 108: it should be "had" not "has"

Author response 23:

We thank the reviewer for bring the typo to our attention, we have now corrected it.

- Line 144: the term "Learning test" is confusing (is it learning, or is it testing?). Same goes for the figures in which the same term is used. I would name it something along the lines of "Numerical control"

Author response 24:

We use the phrase 'Learning test' as it is typical in the field of honeybee training and testing and describes the test which is performed using stimuli familiar to the bee. However, following your suggestion we have changed the term to 'Numerical control test' throughout the manuscript and in the figures.

- Fig 3 a,b) a GLMM through 5 data points seems a bit overkill to me; it is not adding value.

Author response 25:

The GLMM is not just through five data points, in fact the GLMM is calculated using a complex model of 100 predicted data points between 0 and 50 trials (using the information from the 50 data points/choices recorded per bee during training). The five data points which the reviewer refers to are actually the means of each block of 10 choices with CIs shown as this is commonly requested to be displayed alongside the GLMM model. The models shown in Figure 3ab are displaying the model of the analysis which was conducted to statistically analyse the data in the training phase. Additionally, the GLMM used is the correct framework due to the nature of the response rather than an optional choice, as discussed above (see Author Response 8 above).

Appendix B

Dear Editor,

Thank you for the acceptance of our manuscript (RSPB-2019-0238.R1) entitled "Symbolic representation of numerosity by honeybees (*Apis mellifera*): Matching characters to small quantities". We have responded to the minor final comments from the reviewers below and the tracked changed manuscript document is included below for easy reference.

We look forward to seeing our study published with Proceedings of the Royal Society B.

Thank you for all your time and effort on our manuscript.

Sincerely,

Scarlett Howard, Aurore Avarguès-Weber, Jair Garcia, Andrew Greentree, and Adrian Dyer.

Responses to referee comments:

Referee: 1

The authors addressed my raised concerns and answered my questions in a sufficient way, and I can recommend publication of this manuscript. One small typo is still hiding in Figure 2a,b: 'Transfer' instead of 'Tranfer'.

> Author response:
>
> We thank the referee for their support and for highlighting this error. It has been now amended in Figure 2ab.

Referee: 2

I think the authors have addressed adequately my concerns.

Please check Ref 57, the correct reference is

Vallortigara, G. (2015). Foundations of Number and Space Representations in Non-Human Species. In “Evolutionary Origins and Early Development of Number Processing”, pp. 35-66 (Eds., D.C. Geary, D.B. Bearch, K. Mann Koepke), Elsevier, New York.

Author response:

We thank the referee for their support of our manuscript and for bringing this error to our attention. We have amended the reference as seen in our tracked changes manuscript below at reference [57].