Summary

Azra et al. have greatly improved the manuscript, the text is significantly clearer in many sections and many new figures have been added giving greater insight to the model and its results. The authors addressed many of my previous comments. However, a large number of my comments remain unaddressed, or the authors' responses were unsatisfactory. Additionally, some of the added figures and information have raised new issues which were not apparent previously and will now also need to be addressed.

Based on the authors' replies to my previous comments the primary goal of the current model and manuscript is to demonstrate the effectiveness of simplified modelling approaches in studying 3D spatial encoding in rodents. Unfortunately, the current model deviates from the experimental data in every tested environment, but these errors are not addressed or discussed, resulting in an incomplete analysis of the model's effectiveness.

The head direction (HD) "resolution" analysis has been clarified and is much more understandable. However, reviewing this section of the manuscript now, I am not sure how the reader is supposed to interpret this analysis or how it relates to the HD experimental literature. This manipulation involves tuning the modelled HD cell resolution and has no apparent physiological basis.

The classification of the place and grid cells continues to raise concerns: the example grid cells shown are often unconvincing and are often indistinguishable from the example place cells. Based on the authors' responses to my previous comments the distinction between these cell types seems to largely depend on an arbitrarily chosen spiking threshold. The authors rely on an unspecified hexagonal grid score to define and justify their classification of grid cells, they describe this process inconsistently and with no detail. A single grid score alone is a very poor classification method, especially when it is unknown if we would expect to see any grid cells at all. I have provided possible solutions to these issues below where possible.

The authors undertook a lot of work to rerun the model, including new 2D environments, but in terms of analyses and discussion the current manuscript is only an incremental improvement over the previous draft. I think the model and many analyses need to be significantly improved before the manuscript would be suitable for publication.

Major comments

1. Previous major comment 1: I did not find the responses to this comment very convincing. It is still not entirely clear to me what advancements the current model provides. However, if my other comments were addressed this would not remain a major issue as I think at that point the current model would provide a lot of value to the literature and would be interesting to many readers.
    a. "By using an autoencoder in place of LAHN, the current approach can be expanded to model multisensory Integration" but Soman et al. (2017; https://doi.org/10.1109/TCDS.2017.2752369) also used their lateral anti-Hebbian network to model multisensory integration in the hippocampal formation. This is not unique to the current model.

b. “Another key advantage of using an encoder-based approach is the robustness of the model, posited as an agent trainable using reinforcement learning” but other models using lateral anti-Hebbian connectivity have also incorporated reinforcement signals using a multi-Hebbian learning rule (https://doi.org/10.1016/j.pneurobio.2003.12.001). This is not unique to the current model. Additionally, neither of these mechanisms are explored in the current manuscript so they are not advancements made by the current model.
c. “the current paper also elaborates on the impact of HD units (resolution of Head direction)” now that this analysis is better described and I understand it, I’m not convinced that manipulating the resolution of the HD system is a meaningful advancement (see major comment 4 below).
d. “For example, an evenly distributed trajectory with complete access to all dimensions, such as in bats but with lower HD resolution to train on, can exhibit anisotropic firing patterns. Similarly, vice versa is true, which means an unevenly distributed trajectory but with higher resolution of pitch (n2) and azimuth (n1) units for HD tuning can lead to more isotropic firing fields” this does not seem to be shown in the current manuscript, i.e. page 30, line 19 states that “The observed place fields were anisotropic […] for all observed combinations of Azimuth and pitch resolution” suggesting the authors did not find an effect of HD “resolution” on anisotropy. The authors also use the same trajectory throughout their tests, this was modelled on a rat trajectory, so the authors did not compare bat-like and rat-like trajectory distributions.
e. In response to previous minor comment 31: “the only experimental study replicated for rat is the response of place cell and grid cell when the rat moves on the vertical wall, which is not strictly a 3D environment.” I agree that a strength of the current manuscript is that it closely replicates the rat experiments and analyses. However, there are many discrepancies between the current model and the experiment data, both in the methodology (e.g. rats move diagonally in the aligned and tilted lattice, HD cells are discretized) and in results (e.g. the distinction between place and grid cells is unclear, place cells in the lattice mazes are periodic) which undermines this. I disagree that the current model adequately describes the experimental data. Furthermore, the authors’ investigation of 3D environments is a partial one because they do not analyze the activity of grid cells in the lattice maze environments.
f. The authors stated in their response to major comment 2 that “the primary goal of this model is to show the effectiveness of simplified modelling approaches to study 3D spatial encoding in rodents”, in this case I think the manuscript abstract needs to be reworked as it heavily emphasizes comparisons between rats and bats but no attempt has been made to model bat spatial responses or compare them to rat responses.

2. The authors have endeavored to clarify their cell categorization procedures; however, I still have serious concerns about the classification of cells as grid or place cells. Generally, the descriptions of the methods used to analyze the spatial cells lack a great deal of detail and are not informative or consistent, these need to be reworked for clarity.
    a. Page 32, line 29: “The 50 neurons in the encoder layer are analyzed for the grid and place cells (Markus et al., 1994; Hafting et al., 2005)” for the convenience of the reader and for clarity please specify the methods used for cell classification in the current manuscript rather than referring to a secondary source.

b. "To classify the place cells, we consider the neurons whose spatial information is greater than 0.3 bits/spike" please reformulate for clarity. Are only the cells with spatial information >0.3 tested to see if they are place cells? Or are all cells with spatial information >0.3 classified as place cells? In either case, this is not the method used by Markus et al., why did the authors choose 0.3 bits/spike (which is a low value in rat experimental data)? Why do the authors not use the spike train shuffle procedure described by Markus et al.?
c. Hafting et al. (2005) did not describe a grid score metric in their paper but is cited by the authors in reference to grid score. What grid score metric did the authors use? The authors should also include spike train shuffles to assess the statistical significance of a cell's grid score (i.e. see supplementary methods: https://doi.org/10.1126/science.1201685). Additionally, please describe the autocorrelation approach used in the methods/supplementary information.
d. In the authors' replies to my previous comments they stated that "For our analysis we have considered only the cells with positive Hexagonal Grid Score", a grid score cutoff of >0 is far too liberal. No experiment has used such a low grid score cutoff and grid cell grid scores are typically >0.5 (see Grieves at al. Figure 4C or https://www.nature.com/articles/nn.2602 Figure 4C). The authors need to clarify exactly what their classification process for grid cells was as this is an extremely important analysis.
e. "The neurons that pass the criteria of having a high Hexagonal grid score (HGS) (Hafting et al., 2005) for grid cells and criteria of high spatial information (Markus et al., 1994) and low HGS for place cells in a 2D environment are also tested on the 3D pegboard environment." Can the authors include the actual cutoff values (e.g. what is a 'high' spatial information score)? This description is given for the pegboard maze, are these the same or different criteria as described for the other experiments? For clarity and consistency the authors should clearly define their classification criteria for place and grid cells once in their Methods section and apply the same criteria to all of the modelled experiments.

3. The authors have added many plots of spatial responses, the authors also followed my previous suggestion of including a 2D arena which allows for easier classification of the spatial neurons. However, having now seen the responses of many of these cells in a 2D arena, I have concerns about the classification of cells as grid or place cells and the consistency of the model throughout the manuscript.
    a. Compare Figure S5 (pegboard place cells) to Figure S4 (pegboard grid cells). Many of the grid cells in Figure S4 have only 3 fields and their grid pattern does not persist across the arena. Some of the cells (i.e. bottom left and right) simply do not exhibit a grid pattern at all and should not be classified as grid cells. Many of the place cells have multiple, identically shaped fields. The spacing between fields is also remarkably consistent across place and grid cells suggesting a common underlying mechanism of field placement, unlike what is observed in experimental data (see e.g. Harland et al., 2021 https://doi.org/10.1016/j.cub.2021.03.003). One possibility is that the grid cells categorized by the authors are multi-field place cells that have fields in a triangular array by chance. I do not think the authors can rely on a single grid score metric to classify their cells as grid cells. I think the authors need to include multiple 2D environments – a true grid cell should exhibit a hexagonal grid patten in every environment. Because many of the authors' grid cells

exhibit incomplete local grid firing patterns, I would also suggest including a much larger environment (i.e. 4 x the size of the current 2D arena) to assess if the hexagonal firing pattern is indeed continuous. Without these additional tests, and based on the examples shown, I don't think the authors can convincingly label their cells as grid cells.

b. I pointed out previously that some of the plotted grid cells exhibited square firing patterns and that square firing patterns were reported using this same model previously (Aziz et al. 2022; https://doi.org/10.1002/hipo.23461) making it very likely they will arise in the current dataset. The authors responded that "we do not consider square grid cells in the current study" which is not a satisfactory response. If the authors want to make conclusions about hexagonal grid cells, square firing patterns will need to be excluded from the current study. Please include a square grid score, analyze square grid cells separately or exclude them. Please also report the distributions of hexagonal and square grid scores so that the reader can clearly understand the composition of the network's cells.

c. For each environment, the authors also model the activity of their cells in a 2D arena. These 2D arenas allow the authors to classify cells as place or grid cells. However, for the lattice maze experiment the authors did not test if neurons exhibited a hexagonal grid response in the 2D arena and did not analyze the activity of putative grid cells in the lattice mazes. I raised this issue previously and the authors state that "This has been shown by different models previously and hence does not fall under the scope of this study". However, the authors also stated in their response to major comment 2 that "the primary goal of this model is to show the effectiveness of simplified modelling approaches to study 3D spatial encoding in rodents" so it is hard to see how grid cell activity in the lattice mazes is not directly within the scope of the current study.

d. Almost every place cell shown in the aligned and tilted lattice maze seems to exhibit periodic firing (Figure S1.1). Some cells exhibit cubic firing patterns (i.e. 7th cell down, left column, Figure S1.1) while others exhibit columns arranged in a triangular grid (i.e. 3rd cell down, right column, Figure S1.1). These responses are not consistent with experimental rat or bat data, but this is not addressed by the authors. I raised this point previously in major comment 5 and the authors did not address it. The authors need to describe how the model deviates from the experimental data and discuss the implications of this. They should also conduct analyses to determine the periodicity of these cells. Please also include a figure showing the responses of these place cells (and grid cells if included) in the 2D arena.

e. Many of the helical maze place cells are periodic in the 2D arena and it is very difficult to see how they differ from helical maze grid cells in the 2D arena (i.e. compare the 4 fields of the place cell Neuron 7, Figure 9, A1 which are arranged in a triangular grid, to 4 fields of the grid cell Neuron 14, Figure 10, A2 which are arranged in a similar triangular grid). The example grid cells shown in Figure S2 are the most convincing in the manuscript, although they are also often missing firing fields. For some reason, the responses of the helical maze place cells are not shown for the 2D arena so it is impossible to determine how distinct the place and grid cell populations are for this experiment, can the authors add these responses to figure S3? The authors were dismissive of these qualitative assessments previously; however, we do not know a priori that their model will produce place or grid cells, therefore the grid cell classification measure used in the current manuscript appears to not be sufficient to conclude that these neurons are in fact grid cells.

f. Previous minor comment 14: In response to a previous question the authors provided a figure (Figure A) which shows the effect of firing rate map resolution and a spiking threshold on the proportion of place cells and grid cells found. This figure shows that when the spike threshold is low neurons are more likely to be classified as grid cells. This suggests that the place and grid cells in the current manuscript are not distinct populations, but instead their activity deviates purely due to the arbitrary spike threshold chosen by the authors. Thresholding is often used in computational modelling to isolate prominent receptive fields; however, a neuron's classification type should not be dependent on the level of thresholding used. Again, one explanation is that as the threshold is lowered, cells are likely to exhibit more fields and are thus more likely to exhibit fields in a triangular array by chance. See section 2a above for possible solutions to this issue.

g. Previous minor comment 25: Can the authors specify what hexagonal grid score they used (one is not given in the Hafting paper cited). I do not see how many of the plotted grid cells could pass any grid score metric – for example the top right 3 cells in Figure S4 do not exhibit a hexagonal pattern in their autocorrelogram at all. What are the grid scores for these cells? I suggested that grid scores be provided in these plots previously (previous minor comment 28) and they should be added for clarity.

4. I now understand better the analysis and results the authors describe related to the head direction resolution manipulation. Thank you for clarifying this text. Instead of modelling head direction cells with uniform azimuth preferred directions and a distribution of preferred pitch directions, the authors discretize these into n1 or n2 angles respectively. This procedure is sometimes used to reduce the computational complexity of a model and increase its speed, although the practice is less common in recent models. However, the discretization would normally use a large number of angles so as not to introduce artefacts/aliasing into the model (e.g. Page et al. use 500 unique azimuthal angles https://doi.org/10.1152/jn.00501.2017) while the current model includes a maximum of 20 angles. I'm not sure what effect this has on this model's responses.
    a. It is not clear to me why the authors manipulate the "resolution" of the HD system in this way. Experimental data has not shown that the head direction system is discretized in any dimension, HD cell preferred firing directions form a continuous distribution. Because this manipulation is not physiologically grounded it is very difficult to interpret its significance here beyond a model tuning parameter. What does this manipulation correspond to in the HD system?
    b. "This reinforces the hypothesis that the inaccessibility of specific dimensions leads to non-uniform HD tuning" there is no evidence that the uniformity of the HD network changes due to environmental affordances. If a HD cell exhibits a preferred azimuth x pitch angle, evidence strongly suggests that preferred angle will remain consistent relative to other HD cells regardless of the animal's surroundings. Can the authors clarify how they hypothesize inaccessibility leads to non-uniform tuning?
    c. "only when the chosen n1 and n2 values satisfy a specific inverse linear relationship. We speculate that this might be due to the uneven trajectory distribution with lower accessibility to the Z axis." I don't really understand this explanation. Why does altering the resolution of the HD system in the azimuthal plane change the likelihood of cells exhibiting vertical elongation? How is the HD resolution related to the trajectory distribution? Because

the manipulation is unorthodox the authors need to be much clearer with their interpretation and explanation of this.

5. In my previous comments I pointed out that many of the results presented in the current manuscript deviate significantly from the experimental data (Major comment 3). The authors' replies to these comments were generally unsatisfactory. Each inconsistency taken separately does not represent a serious issue, and of course computational models can rarely, if ever, fully replicate experimental findings. However, the current model displays errors in every environment. In their response to major comment 2 the authors stated that "the primary goal of this model is to show the effectiveness of simplified modelling approaches to study 3D spatial encoding in rodents" so these errors, at the very least, need to be described and discussed in the manuscript as they point to the effectiveness of the modelling approach, but they are instead ignored.
    a. Page 36, line 29: "We observed that in an aligned lattice, a higher proportion of fields are elongated and oriented towards the Z axis" Page 37, line1: "…we hypothesize that to observe the significant axial elongation and orientation as observed in experiments, inaccessibility of that axis is necessary". Grieves et al. (2020) did **not** observe a greater proportion of fields elongated vertically (see their Fig. 7C). The X, Y and Z axes "shared a similar number of fields". Thus, the experimental results are not in agreement with the current model. The authors stated in their response to Major comment 3a that this depends on the choice of model parameters: "But for another set of pitch and azimuth units, proportion of fields oriented along all three axes would be more than expected by chance." It is hard to assess just from Figure 8.2, but I do not think there is a combination of azimuth and pitch units which is in agreement with the experimental results. If there is a combination which recreates the experimental results more closely this needs to be described clearly in the text, and why is that not the parameter combination used when generating Figure 6? Because there is no evidence that HD "resolution" changes between environments, a parameter combination would need to be found which recreates the experimental results consistently across all the tested environments.
    b. The authors' response to previous Major comment 3b (that fields in the pegboard maze are not solid stripes but are instead multiple fields stacked vertically) did not address the issue. The authors also addressed this phenomenon in minor comment 26: "The fields modelled are indeed stripes but look like individual fields linearly stacked up. This is probably due to higher window of smoothening used to generate firing rate maps" The modelled pegboard fields shown in Figure 12 are **not** stripes. They are very clearly separated, individual fields arranged in a vertical line. This is not consistent with Hayman et al. (2011), see their Figure 2. This difference is apparent in the trajectory and spike plots, so this effect is not due to firing rate map smoothing. The authors need to describe how the model deviates from the experimental data and discuss the implications of this.
    c. In response to previous major comment 3d about the inconsistency of fields in the helical maze: "As we have trained the entire model again, those results are completely changed and we don't see above mentioned disruptions in our new results" but the example cells in Figure S2 and S3 clearly still show this effect. Most cells have fields that repeat only on a subset of coils, additionally, the majority of place cells in Figure S3 have many vertically repeating fields. These responses are not consistent with Hayman et al. (2011), see their

Figure 3A and 4A. The authors need to describe how the model deviates from the experimental data and discuss the implications of this.

d. I previously pointed out that the modelled place cells are not elongated parallel to the maze boundaries in the 2D arena (Major comment 3e). This was an important finding of Grieves et al. (2020) because place cells in the lattice mazes also exhibited elongated place fields along the 3D maze boundaries. This suggests place cells use the same mechanism for generating place fields in 2D and 3D. The authors' response was that "In the new results, we show the average length of firing fields in the minor and major axis in Fig. 11:B and Fig. 12: B. The data indeed shows elongated firing fields in 2D arena" but Figure 11B and 12B do not show this, the minor and major axes are virtually identical. The aspect ratio values in 11C and 12C are also very close to 1.0 and the spatial information in 11D and 12D are also virtually identical in each axis. These results are consistent with the fields being circular. Note that by definition the major axes will always be consistently larger, so the authors need to instead calculate field elongation, an analysis they use for the lattice mazes in Figure 5A1 & B1 but do not apply to the 2D arena. The authors also do not test if the fields in the 2D arena are oriented parallel to the walls, for this they would need to repeat the analysis shown in Figure 6 which was also not applied to the 2D arena. The authors also added text to page 34, line 19: "Firing fields in the flat arena were slightly oval, typical for place cells but were highly elongated in the pegboard. This is deduced by calculating the aspect ratios for all the fields of the same cells in both environments (F: 1.30 +- 0.29; PB: 3.17 +- 1.07; t86 = -12.47, $p < 0.0001$) (Fig. 11: C". Aspect ratio does not seem to be defined, how was aspect ratio calculated in the different mazes? To determine if the fields were oval in the 2D arena the authors would need to compare the aspect ratio values in the 2D arena to 1.0 (circular), with a one-sample t-test for example. The authors compared the 2D arena to the pegboard maze which does not answer the question. If fields in the 2D arena are not elongated, this inconsistency with the experimental data needs to be discussed in the manuscript.

6. Previous minor comment 19: " The trajectories for lattice mazes were inspired from the experimental studies but to keep the trajectories natural enough, the diagonal movements were added. Nevertheless, the probability of diagonal movement is comparably lower than the axial one."
    a. Table 1 lists the probability of diagonal movements as 24% of all movements, so these make up a significant fraction of movements in the mazes. Does the inclusion of these trajectories affect the output of the model?
    b. Why are diagonal movements in the AB plane included in the tilted lattice maze (Table 1)? This deviates from the experimental data where there were no differences between movements in the AB, BC & CA planes. Can the authors clarify this and explain how this impacts the results?
    c. I downloaded the data files provided by the authors and plotted the contents of "Trajectory_interpolated_tilted_lattice1.csv", this trajectory includes diagonal movements on the face of each maze cube not just in one plane, why does Table 1 only list diagonal movements in the AB plane if they were made in all 3 planes of the maze?
    d. The trajectory in "Trajectory_interpolated_aligned_lattice.csv" also has diagonal movements in all 3 planes, these can also be seen when zooming in on Figure 3, A1. This means that the pitch directions in which the agent can move in the aligned lattice are not

"0, 90 and 180 degrees" (page 30, line 8) because the agent will adopt intermediary pitch angles during these diagonal movements. Can the authors clarify this and explain how this impacts the results?

e. How are the upward/downward diagonal movements in the aligned lattice maze categorized in Table 1?

Minor comments

7. The introduction includes a nice paragraph on the different experiments assessing 3D head direction responses (Page 20, line 14). There should maybe also be some discussion of the dual-axis rule here (Page et al. 2018 https://journals.physiology.org/doi/pdf/10.1152/jn.00501.2017). Can the authors discuss why their 3D model does not include the dual-axis rule when encoding head direction, and if this would change their results?
8. Page 21, line 20: "In this study, we focus on the following set of questions in 3D navigation:" the lines after this seem to be a bit muddled, there seem to be two points presented, one of which is a bullet point, but neither of them are a question – the first is a proposal and the second is a result.
9. Please specify the statistical tests used throughout the manuscript, e.g. what kind of t-test was used to compare the minor axis lengths on page 34, line 21?
10. In Figure 11D the y-axis label should specify units of bits/spike.
11. In Figure 11E it is unclear what units are shown on the x-axis (maze layers?).
12. Page 5, line 25: What was the size of the 2D arena? What duration of exploration does the trajectory correspond to?
13. Page 30, line 8: "n2 = 3 units [45, 90 and 135 degrees for tilted lattice" why is a pitch of 90- degrees included for the tilted lattice? I can't see when a rat would be able to move at a pitch of 90 degrees (horizontally) in that maze.
14. The tilted lattices shown in Figure 3, B2 and Figure 4, B1 do not seem to have the correct orientation. The top and bottom vertices should be vertically collinear, but this is not the case (see Grieves et al. 2020 figure 1b). What was the orientation of the tilted lattice used in the model?
15. Figure 8.2: B4-B6, the legends should say A, B & C for these plots not X, Y & Z.
16. Page 8, Line 27: "These spatial cells are further tested on the 3D mazes and the results are reported." This sentence is repeated twice.
17. "To check whether the fields were elongated vertically, we also compared spatial information on the horizontal and vertical axis for both flat and pegboard" do the authors mean the X and Y axes for the flat arena?
18. Figure S2 shows 12 grid cells, but the main text states that only 10 neurons exhibited grid responses (page 34, line 13).
19. The type of error bars shown in plots (i.e. Figure 11B-E) should be specified in the figure legend.
20. Figure 8.2: could the graphs be plotted consistently? i.e. plot A3 is rotated 90 degrees around the z-axis relative to A2 and B3 and the elevation is also different to B3 which makes comparisons between plots quite difficult.