**Revisions Round 2**

Since there were some problems with the review process, I undertook a thorough reading of the manuscript myself to ensure that it undergoes a sufficiently thorough scrutiny prior to publication. I believe this is a valuable contribution to the literature on foraging deserving of publication in this special issue and found the final page of the discussion particularly interesting and well-written.

**Thank you very much for your comments, we are pleased you found the manuscript interesting and valuable.**

However, I do find that the manuscript overemphasises the successfulness of the model in predicting behaviour in the first paragraph of the discussion. The model does no more than numerically express a series of foraging biases, finding that it can predict quite well certain trials on which participants go for more proximal targets and from the top left corner (when working on a computer; presumably the bottom right bias on an ipad is because the participants were right handed and used a finger to select targets) but other trials less well.

**We have tried to reword this paragraph to avoid overstating our findings. We think it is interesting to evaluate the ‘trial-by-trial’ predictions of our model as it is not always the case that aggregate performance (as we evaluated in our previous paper) is a perfect match to  more detailed trial-by-trial exploration. The foraging literature has also focused strongly on target specific parameters that modulate performance (e.g. feature vs. conjunction search) so we think it is important to clearly make the argument (however obvious it may be!) that proximity is likely a more important factor underlying behaviour.**

The way in which parameters combine with each other is not presented in the main manuscript, where it should be (a reviewer also suggested that and I strongly suggest this should be done as it is important for understanding the manuscript, rather than tangential). You could have also evaluated different parameter contributions by fitting weights to evaluate how they combine.

As discussed at the end, the real value of this exploratory modelling exercise is in generating predictions for future more targeted experiments/modelling. In fact, a lot of very useful and thought-inspiring information is buried in the supplementary materials and should be moved to the main manuscript.

**We have now included the graphs containing how model accuracy varies with the strength of an individual’s bM, bP, pA and pS parameters in the main manuscript (this graph was originally only contained in the supplementary materials). We hope this addresses your concerns - we were not entirely sure we understood what reviewer 3 wanted, but we think this graph makes it clearer how different parameters contribute to the model accuracy.**

There seems to be a leap in conjunction search in fig.1 after ~20 trials, which I presume happens because one type of conjunction is exhausted and the participant moves to another target which can then be selected based on a single differential feature? While prior to this switch conjunction search is predicted worse than feature search, after trial 20 it is predicted equally well, presumably because proximity bias then becomes the most important parameter. This is worth discussing as it shows either a potential weakness in how the parameters for sticking with the same target and going for the most proximal target are implemented, or a small but significant effect on foraging behaviour outside the scope of the model’s parameters (potentially this could be evaluated by further modelling in which feature and conjunction search are allowed to have different weights adding up to 1 for how they balance the tendency to switch targets or go with proximity).

**Thank you for these comments, we have now discussed the ‘jump’ in behaviour in the conjunction trials more fully (though we would argue that the performance on the conjunction trials still lags behind the feature trials). Feature and conjunction search are indeed allowed to have different weights in the model (hence Figure 2 showing different values per participant for each parameter in the different conditions). We think that further modifications to the Clarke et al (2022) model are outside of the scope of this manuscript: we are working on a substantial update that will form a new manuscript, but our design decisions are being influenced by the weaknesses in the model that the current manuscript identifies.**

Looking at figure 3, what you call path-length optimisation could also be called extending the planning of behaviour towards the second of two targets when their proximity is very near and along the same path. This appears to be the case with at least half of the prediction errors in the well-modelled participants’ trials which switch between targets. Such motor behaviours that minimise energy expenditure are fairly common and could be added to the model.

**Our direction parameter should capture behaviour where participants choose the target that is near and along the same path (this would be a positive direction bias). Surprisingly, we find negative direction biases in these datasets, where participants are more likely to double back on themselves (this is something we discuss more in the original Clarke et al 2022 modelling paper).**

 In fact, it is obvious from figure 3 that in more poorly modelled participants’ trials, the failure to model them adequately is because they very strongly follow a single target bias (selecting the second target type only after finishing the first) as well as a single spatial bias (sweeping in a clockwise and then counterclockwise pattern; or in a left-right/down-up and then right-left/up-down i.e. reverse pattern). Again, this is a viable strategy that combines visual and motor biases in behaviour and could be discussed. In fact, this question is associated with the aforementioned problem of modelling conjunction searches being systematically poorer for the first 20 trials, as by this point one target conjunction 2 type is exhausted and the behaviour simply switches to shape-feature search between a single colour. As conjunction searches are more likely to have a mix of short and long paths due to preference for sticking with the same target type, this reveals the same inadequacy of the model as mentioned before, which seems to me to be more basic than a simple question of path-length.

**This is a very good point, thank you for drawing our attention to it (the poor/good trials are randomly selected - some of the poor trials do show evidence of more ‘disorganised’ behaviour, but you are absolutely correct that in these cases, the participants are showing a very structured strategy, just not one that the model captures well). We have now discussed this in more detail, both in the results and in the discussion: we hope this makes it clearer that we know our model is “wrong”, and that further work will hopefully come up with improved versions. The aim of our paper is really to show the limitations of our model, to motivate future work.**

There could have been an error with the upload or with the manuscript handling platform from MDPI Vision - the revised manuscript that I can see on the system does not seem to contain changes indicated in any discernible format (e.g. in a different colour, so the editor can see easily what has been changed/added). Unfortunately, I cannot see that the adjustments that the authors say they've made (e.g. discussing the motor system) have indeed been made. This is why I have selected 'reject and encourage resubmission', so that it can be made sure that the revisions the authors say they have done have indeed been done adequately (otherwise, this would be an 'accept after minor revision').

**Apologies, this was an oversight during submission (we were attempting to get the revisions done in the 7 days initially assigned to us by MDPI, which may have been overly ambitious!) We have now included a highlighted version that covers the changes made in both revisions.**

Minor comments:

Please add names when referring to studies on p.2, lines 80-83 – referring to studies without author names doesn’t work well with the numbered reference format used by the journal. Same applies to line 88 and all subsequent similar occurrences.

**We have now tried to make this clearer throughout the manuscript: see L80, 82, 89, 114 and 158.**

I appreciate that the authors didn’t want to repeat too much of their previous methods section, and as the other article is open source readers should be able to look it up easily. However methods should be comprehensive enough in how they describe the formulas so please add the specification that atan2(j, i) is the direction of travel from i to j and f(ϕ1, ϕ2) calculates the angular difference.

**This has now been included: see L101-102.**

The letters in Figure 1 are not very visible and need to be larger, as they cannot be read when viewing in a browser on a normal sized monitor. Please also use a more legible font. X and y axis have also not rendered correctly and are non existent in all figures, which makes it difficult to read the graphs.

**We have increased the font size, which hopefully helps with readability. The graphs are rendered in Tufte format, which aims to maximise data-ink and thus avoids the use of borders/axis lines/grids (according to the principles set out in ‘The Visual Display of Quantitative Information’).**

On y axis, when it doesn’t start from zero, please indicate with // in this and following figures.

**As far as we understand, this is the convention for bar graphs (or even better, start the graph at zero) - and we agree that in these cases, this is important as the relative area of the bars will be misleading if the y axis does not start at zero. However, this objection is not relevant for line or point graphs, and APA standard does not require it (e.g. see** [**https://apastyle.apa.org/style-grammar-guidelines/tables-figures/sample-figures#line**](https://apastyle.apa.org/style-grammar-guidelines/tables-figures/sample-figures#line)**). In the case of Figure 1, the important comparison is with the chance line, and so we have designed the figure to make this as clear as possible.**

in figure 2, x axis should say bp parameter value rather than just value.

**Thank you, this has been fixed.**

Is it worth including the trials in calculating average performance of the model after the departure from a roughly linear trend in the chance level in figure 1? One could remove those trials from the quantification (and grey them out in the model on Figure 1)?

**We don’t necessarily want to select a threshold for when the chance level becomes “non-linear”, as this seems like it could be rather arbitrary to us. However, we will make all code and data freely available such that interested readers can calculate their own preferred model accuracy metrics (as there are no doubt many different options that could be valid and informative!)**

In footnote on p.7 please spell out the abbreviation for ADFC prior to using it.

**We have reworded this footnote (the abbreviation was simply the initials of the lead author, but we appreciate this may not be clear).**

The writing style sometimes unnecessarily uses exclamation marks to drive certain points home. I suggest removing those, as they make for more unusual reading.

**We have removed these as suggested.**

Can you please unpack the last sentence of paragraph 2 of the discussion – it is too long and unclear. What are fatter trials?

**Apologies, we have attempted to make this clearer on L227-230.**

Please reformat your supplementary materials in a more reader-friendly format, including labels for supplementary figures.

**We have now included Figure numbering, which we hope makes the supplementary material clearer.**