We thank the Editor and the reviewers for their very helpful and constructive comments. Our responses can be found below in bold, and all line numbers in our responses below refer to the revised manuscript.

Associate Editor Comments to Author (Professor Chris Chambers):
Associate Editor: 1
Comments to the Author:
I have now obtained four constructive evaluations of your Stage 1 submission. As you will see, the general consensus is positive and I concur with the reviewers that your submission is a promising candidate for eventual IPA. Within the reviews you will find a range of helpful suggestions, chiefly clarifications (and strengthening) of the study rationale and hypotheses in certain areas, as well as points of methodological detail to add and confusion to resolve. Perhaps the most significant issue to address is whether the effect of scarcity can in fact be isolated as intended. In revising can I also ask that you include a study design table in the main text as outlined here: https://rr.peercommunityin.org/help/guide_for_authors#h_27513965735331613309625021 Although this is not a formal requirement for RSOS, we have found at PCI RR that it is a very useful aid for authors, reviewers and readers alike, and I think it would make the design even clearer.

We have now included a study design table, as requested.

For reviews this positive I would generally issue a minor revision, but in this case I'm making it a major revision to give you some extra time to prepare a response to the four reviews and a revised manuscript. I hope you find the evaluations helpful and look forward to receiving a revised submission in due course.

Comments to Author:
Reviewer: 1

Comments to the Author(s)
I thought this preregistered report was of a very high quality. The proposed study builds nicely on previous work from the group. More importantly, the current study has the potential to make at least two important contributions to the field. In terms of research question, and as identified by the authors, the work will fill a gap in our knowledge of human foraging, by exploring the intrinsic value of target prevalence, when this is the only factor manipulated. In terms of methodology/analysis, the study will help further popularize the use of generative foraging models that can provide another important analysis tool to replace/complement more standard analysis techniques.

The logic, rationale, and plausibility of the proposed hypotheses

I found all aspects of the study description to be clear, feasible and easily reproducible. The level of methodological detail and proposed analyses were all very clear and the overall logic of the

study made sense and gave rise to clear, testable predictions. The inclusion of additional data capture (eye tracking) – the inclusion of eye tracking –

As already noted, I have no major concerns with the proposed design. Just a couple of comments.

First, I wonder if the authors share my suspicion that blocking the conditions – as proposed – would be critical here. With this in mind, I think my tendency would be to include a few more repetitions per condition, than the suggested six trials. This is not a general concern with power – the inclusion of the pilot data was helpful – but if overall study duration is not a big issue, might be useful to give particular display characteristics time for their influence to be felt, and to overcome any influence from previous blocks. Just a hunch.

This is a very interesting point and one we had not fully considered. We have increased the number of trials per condition to 10 (L170), which we think will keep the experiment at a reasonable length while also allowing any blocking effects to become clear.

While fully supporting the use of the generative model and this general approach to hypothesis testing, it might also be useful to provide (as supplementary analyses??) standard aggregate descriptive stats to make comparisons with previous work in the field that much smoother.

We are happy to provide this information in Supplementary Materials as suggested. We have added a short section (L368 onwards) in '4.3 Planned Exploratory Analyses' to discuss this.

Minor text comments.

P2,L44. "By contrast, implicit valueS", singular.

Fixed (L47).

P2,L45 "to an to an". Repeated phrase.

Fixed (L49).

P4,L14. Sentence beginning "They also…" Could you further spell out the difference between "inherent bias" and "simply like to stay searching for the same target template". They both seem like they could be described in terms of inherent bias.

Apologies for the lack of clarity here. We have reworded to "they also do not allow us to distinguish between the case where a participant sticks with a particular target because of a preference for that specific target type" (L108-110) - the point is really that people might prefer target A over target B, or they may not have a preference for either but simply like sticking with

one target (so over a longer run of trials in this scenario, you might see that people have long runs, but interchangeably begin with target A or target B).

P5,L7. “Secondarily” This sort of jumped out at me.

We have reworded this to “as a secondary aim” (L145).

P5,L22. “collect DATA FROM 36 participants”?

Fixed (L153).

Reviewer: 2

Comments to the Author(s)
This Registered Report – Stage 1 mansucript proposes a study of the role of target scarcity in visual foraging. The authors identify contradictions in the prior literature as to how scarcity influences foraging behavior. They argue the practice of using behavior aggregated over many trials as the primary outcome measure has potentially obscured the structure of foraging behavior in these tasks. The authors have previously developed a generative Bayesian model of foraging behavior that can explicitly account for behavior at a much finer grain. This model was developed on data acquired for other purposes. Here they propose to collect data prospectively in order to analyse it with the model.

The proposed experiment will vary difficulty (conjunction vs feature targets) and target scarcity. Stimuli will be colored squares and circles. The primary hypothesis is that participants will show a preference for scare targets, where this preference is measured by model parameters. They also have four secondary hypotheses.

RSOS criteria for registered reports:

1. The scientific validity of the research question(s).

The research question, whether scarcity iumbues targets with implicit value, is scientifically valid and interesting.

2. The logic, rationale, and plausibility of the proposed hypotheses.

The proposed hypotheses, including the secondary hypotheses, are plausible and derived from the pre-existing literature on the topic.

3. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis where applicable).

The experimental methodology and the generative Bayesian model are sound and quite feasible. I commend the authors on using simulated data to justify their proposed sample size.

4. Whether the clarity and degree of methodological detail would be sufficient to replicate the proposed experimental procedures and analysis pipeline.

Since the authors have posted the software for both stimulus presentation and analysis on their GitHub repository, it should be very easy to replicate their proposed experiment with some precision. However, I would encourage the authors to include a more detailed description of the model, perhaps in supplementary materials (i.e., so the reader doesn't necessarily have to read the PLoS Computational Biology paper to understand the details of the model).

Thank you for this suggestion, we have now included extra details in the Supplementary Materials as suggested (see 1_check_sim_and_features.html).

5. Whether the authors provide a sufficiently clear and detailed description of the methods to prevent undisclosed flexibility in the experimental procedures or analysis pipeline.

The modeling procedure is clear, the proposed analyses and methods straightforward, and the authors have promised to document any changes to the analysis procedures.

6. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the results obtained are able to test the stated hypotheses.

Given this design, these issues should not be much of a problem. The authors have specified data exclusion procedures.

I have a couple of suggestions for improving clarity.

(1) In the second paragraph on p. 3, the authors introduce "target prevalence". Are "prevalence" and "scarcity" synonymous here?

Yes: we have clarified this on L71-72.

(2) At the bottom of p. 3, I think the point of the Sobel and Cave (2002) paper is not that rare distractors were higher value because they were scarce, but that it is faster to start search through a smaller group than through a larger group.

Thanks, we have clarified this on L100-102 - arguably, this could be another type of 'low level' explanation for a scarcity bias.

(3) the first paragraph on p. 4 describes the number of runs or the total number of targets found in the longest run as “aggregate statistics”, in contrast to the approach that the authors will take in the proposed research. However, in the last sentence (and later in the next paragraph) they switch to the phrase “summary statistics”. In addition to the confusion of inconsistent terminology, “summary statistics” may also be confused with “ensemble statistics” in the field of research, so I recommend sticking toi “aggregate statistics”. Furthermore, I find the final sentence in this paragraph unclear. Please rephrase.

Thank you, this is a good point and we have now renamed “summary statistics” to “aggregate statistics” throughout. We have also expanded upon the final sentence of the paragraph (L115-119) which hopefully makes our point clearer.

(4) In section 3.1, b_A is defined as $logit(p_A)$, but p_A is never defined.

Apologies, we have now clarified this throughout section 3.1.

(5) in equation (5), $scare_A$ should be $scarce_A$

Fixed (L328).

Todd S. Horowitz

Reviewer: 3

Comments to the Author(s)
This registered report aims to investigate the effect of target scarcity on visual foraging behavior. It involves an "exhaustive" visual foraging paradigm (i.e., the task is to select all available targets in the display), where target scarcity is manipulated across three levels: Target A is scarce, Target B is scarce, and an even ratio between the two target categories. Additionally, task difficulty, with two levels—feature foraging and conjunction foraging—is manipulated. The data will be analyzed using a Bayesian statistical model recently developed and published by the authors.

Overall, this is a good and interesting research project, with a robust theoretical background, methodology, and analysis pipeline. I have a few comments/questions below, that I have organized following the evaluation criteria of Royal Society Open Science:

1. The scientific validity of the research question(s)

The research question is clearly explained and justified in the literature review. The proposed research is interesting and grounded in the growing field of human visual foraging.

I have one comment/question regarding the literature review:

P.3, the literature review reveals contrasting results regarding the effects of target explicit value on visual foraging. Regarding target scarcity, the review demonstrates a preference for rarer elements in the display. However, there is another study by Wolfe et al. (2018) in which both target explicit value and prevalence were manipulated. In the condition where all targets had the same value but different prevalence, the authors observed a preference for the most common targets in the display. Although the procedure differs (non-exhaustive foraging task), I wondered whether this could also contribute to the discussion on the contrasting results in the literature regarding the effects of target value (explicit or implicit).
Reference: Wolfe, J. M., Cain, M. S., & Alaoui-Soce, A. (2018). Hybrid value foraging: How the value of targets shapes human foraging behavior. Attention, Perception, & Psychophysics, 80, 609-621.

Thank you very much for drawing our attention to this interesting paper: we agree that it is highly relevant, and have now discussed it at L81.

2. The logic, rationale, and plausibility of the proposed hypotheses

Overall, the rationale is straightforward, and the proposed hypotheses are clear and well-founded.

I have one comment/question regarding the hypotheses:
It is planned to contrast feature and conjunction foraging, but I'm not entirely sure that I understand the rationale behind this manipulation. Is there any hypothesis about a potential modulation of the effect of target scarcity between feature and conjunction foraging? In other words, does the main hypothesis H1 depend on the secondary hypothesis H2? I would guess that because conjunction foraging is more difficult and associated with longer runs, the effect of scarcity would be stronger in conjunction than in feature foraging. In feature foraging, the strong proximity bias may override the effect of scarcity. However, perhaps the authors have different expectations. In any case, my question is whether the authors could add an explicit justification in the manuscript regarding the feature/conjunction manipulation in relation to the main hypothesis on target scarcity.

We did not have a strong prediction about whether there is an interaction between feature/conjunction foraging and target scarcity (as discussed on L332), although we agree that your suggestion is plausible (and we will test whether there are any differences in the effect of scarcity dependent on condition; detailed in Section 4.1). However, the feature/conjunction manipulation is primarily included for testing secondary hypotheses H2-H4 i.e. testing the generality of previous results. We do not think this means H1 depends on these hypotheses - indeed, our plan is to marginalise over these conditions for testing H1. An advantage of testing scarcity in two different conditions is that it increases the applicability of our results and may stimulate further research.

3. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis where applicable)

All planned methodology and analysis seems feasible. The authors are experts in the research methods (visual foraging task) and analysis pipeline (power analysis and Bayesian statistical model developed by the authors) that are involved in the registered report, as demonstrated by their prior publications.

I have one question regarding the methodology:
In the planned experiment, there are 6 trials per condition with 20 target selections per trial. Although the number of trials per condition is small, the authors have provided a good justification for this choice. However, the number of target selections per trial is also small compared to previous studies in visual foraging (e.g., 40 target selections per trial in Kristjansson et al., 2014). Reducing the number of target selections per trial may alter task difficulty and potentially impact foraging strategy (e.g., fewer run-like behaviors in conjunction conditions). Why did the authors choose to have only 20 target selections per trial, and could the number of target selections per trial modify the expected effect of target scarcity on foraging behavior?

Our choice to reduce the number of target selections per trial is mostly pragmatic, to ensure that the experiment does not take too long. We note that our pilot data looks broadly similar to previous studies (e.g. we see similar stick probabilities and proximity tuning values), suggesting that this change is unlikely to strongly impact foraging strategy. In addition, when looking at the impact of scarcity, we expect the initial target selections to be most important for calculating bA (as if people have a preference for a particular target type, we expect them to be selected earlier in the trial) and therefore we thought it most helpful to maximise the number of trials in each condition, rather than the number of targets in each condition.

4. Whether the clarity and degree of methodological detail would be sufficient to replicate exactly the proposed experimental procedures and analysis pipeline

The methods section is clear and detailed, enabling easy replication. I however have a few questions/comments below:

P.5, L.48, it is said that the order of blocks will be randomized for each participant. I would rather suggest counterbalancing block order, so that half of the participants start with feature foraging and half with conjunction foraging.

Thanks for this suggestion, we have now implemented this (see L171-174). Half of the participants will complete the three feature foraging blocks (in a randomised order) first, while the other half will complete the three conjunction foraging blocks first (in a randomised order).

P.7, L.8: It is said that eye movements are recorded during the task. Does it mean that there is a chin rest? If yes, at which screen distance?

Yes, there is a chin rest at 60cm distance. This has now been added to the manuscript (L202-204).

P.9, L.38-42, about the initial selection bias: Does the initial selection bias depend on where the mouse cursor is at the beginning of the trial? And if so, in the current experiment where is it planned for the mouse cursor to be positioned at the beginning of each trial?

Thanks for this excellent question. We are really treating initial biases as a 'nuisance variable' to be controlled for in our analyses rather than an interesting question in its own right in this study, though we completely agree that if we really wanted to ask questions about what influences the initial bias we would want to control the mouse cursor position at the beginning of the trial. However, in the current study, we prefer to allow people to complete the task in the way that feels most natural to them, including allowing them to position the cursor wherever they prefer. However, there will always be a fixation cross in the centre of the display before the beginning of the trial.

5. Whether the authors provide a sufficiently clear and detailed description of the methods to prevent undisclosed flexibility in the experimental procedures or analysis pipeline

The "4.3 – Planned exploratory analysis" subsection shows that the authors are willing to test alternative analysis pipelines in case new models are published during data collection for the current experiment.

6. Whether the authors have considered sufficient outcome-neutral conditions (e.g. positive controls) for ensuring that the results obtained are able to test the stated hypotheses

I am not sure this item can be applied to the current research given that the rationale is very straightforward. I appreciate that the authors provided simulated data and pilot results. I have a few questions regarding this:

Simulated data (Fig. 2): Was the simulation run for a single condition (feature or conjunction foraging), or were the feature and conjunction conditions averaged? Could the authors present separate simulations for the feature and conjunction conditions?

We have now included a more sophisticated simulation that takes into account feature and conjunction conditions, as requested (see section 4.1).

Pilot results (Fig3): Do "easy" and "hard" conditions correspond to feature and conjunction foraging, respectively? This is not clearly stated in the figure legend.

Thanks for picking up on this - you are correct, but we have now clarified by renaming 'easy' to 'feature' and 'hard' to 'conjunction'.

(Very) minor comments:
- P.2, L.45: typo “to an to an”

Fixed (L49).

- P.6, L.28-29: Stimulus size is provided in pixels. It may be informative for the reader to also provide stimulus size in degrees of visual angle.

This is now provided on L180-181 and 194-196.

- P.9, L.45-52: In the equations 3 and 4, I’m not sure I understand what 15, 4 and 3 correspond to? Could the authors provide a bit more details for naïve readers unfamiliar with the planned analysis?

The values mentioned are the mean and standard deviations for the normal distributions that define the priors - in the case of equations 3 and 4, these are for the proximity and relative direction parameters respectively. We clarify this on L292-294.

Jerome Tagu

Reviewer: 4

Comments to the Author(s)
This is a preregistration for intended data collection. The aim is twofold, to investigate the effect of scarcity as well as collecting data for further testing of a recent Bayesian model that is aimed at accounting for foraging behaviour as a sampling without replacement process (Clarke et al. 2022). For the most part this plan sounds sensible, and I am in agreement with the authors that there is good reason to test this model further.

It was however not quite clear to me why this exact data which involves testing target scarcity is particularly useful for the model.

Regarding the question of checking the effect of scarcity, I am not entirely sure that this effect can be isolated. There is evidence in the literature that there is a push-pull relation between different factors in foraging (Wolfe, Cain, A. Alaoui-Soce, 2018). Wolfe et al. identified three factors (target value, spatial proximity and priming of the previous target features) that determine selection during foraging. Tagu and Kristjánsson then showed how there can be a "selection balance" between these factors – in other words they interact and tweaking one can led to differing effects of another etc. So any effects of manipulating one factor (such as scarcity) on its own may only reveal effects applicable to a single level of another factor.

We agree that the literature strongly suggests there is an interaction between scarcity and explicit value (although we also note that different studies have found different results). However, we would argue it is still of interest to look at scarcity on its own as a factor, particularly in this design where the targets deliberately don't have an explicit value. We also think that a well-powered registered report on scarcity effects would be a useful dataset for the field, even if it does not cover every possible interaction.

Also, I wonder with scarce displays, is there any likelihood that there could be any singleton effects? Could they complicate interpretation of the data?

By the nature of foraging tasks (where the targets disappear once selected) there are always going to be cases where a given target type is a singleton in the display, and we have not found this complicates interpretation previously (the model knows how many targets are left and is able to take this into account).

I was a little confused about the discussion of Nitaynanda and Chittka. Of course bees will choose the flower that gives the best pay-out – but the authors of the paper in question also talk about salience. Is there something missing from this description as this is described in the manuscript?

We have hopefully clarified this by extending our discussion of this paper on L47 onwards: we are really just using this as an example of how explicit and implicit value might be differentially manipulated.

With regard to the authors discussion of the effects of reward on page 2, I think that the authors need to consider that just finding a target can be rewarding (at least in some sense) – just as the completion of a task is rewarding. How easily can scarcity then be isolated from reward? How does this relate to Brock's ideas?

Brock's ideas specifically argue that the rewarding nature of finding a scarce target should be higher than that of more readily available targets, so we don't think this precludes the idea that just finding a target can be rewarding. However, we will of course discuss these types of ideas further in the overall discussion in stage 2: in particular, if we find an effect of scarcity, we recognise that our design will not necessarily tell us exactly how this effect is mediated e.g. whether it is via reward, or some other factor.

The authors planned data collection and analyses are all of very high quality, but I wonder whether, in light of my comments above, whether they should rethink their design a little.