Peer Review File

Manuscript Title: Innovative symbolic and technological behaviours 105,000 years ago in a wetter Kalahari

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

A. The authors argue that calcite crystals and ostrich eggshell pieces were collected by people over 100,000 years ago in the western Kalahari. Theses are argued to be evidence of an earlier appearance of symbolic behavior than proposed earlier. They also tie this behavior into evidence of a wetter period in this region, which today is quite arid.

B. This is an original and significant contribution, and ties together new discoveries with a review of evidence for similar technologies elsewhere in southern Africa.

C. The data and methods are appropriate for this kind of research. Some of the methods have not really been employed in such an analysis in the past,but are perfectly appropriate here. It is an excellent example of multi- and inter-disciplinary research.

D. The use of quantitative methods and statistics in appropriate.

E. The conclusions are appropriate. In the supplementary text, the authors try to evaluate their conclusions in the light of other possible explanations. This is an excellent way to approach this kind of data.

F. The references are appropriate and numerous. I note that the bibliographic entries are done differently for the main article as opposed to the supplementary text. I am not sure if this is the way Nature authors are supposed to create bibliographies, but, regardless, it is worthy of checking over. In addition, some journal titles are missing from some references in the main text, but also in the supplementary documents. Also, in SI Table S2 OES(=ostrich eggshell) is used as a plural noun in some places and as a singular noun in others. Pick one option, and make it the same throughout the table. There are a couple of other typos to be fixed, in addition to the absence of some journal titles, volumes, and/ or page numbers. In the same table, the first reference (Ames et al. 2020), the last authors initials are mixed in with the title of the journal. This appears as Collins, BJOQ. Collins should be NR, and the journal is Open Quaternary. In these bibliographies, some journal titles are italicized, while others are not. Please be consistent throughout. Pick one system and stick to it. Another typo - Steele et al. 2012. South Africa is misspelled as South Sfrica.

H. The text is quite clear and is a reasonable length. This paper should be published by Nature. It is a state of the art treatment of a critical issue - the origins and early evolution of symbolic behavior. Once again, Africa leads the way.

Referee #2 (Remarks to the Author):

The paper by Wilkins et al. describes a significant site in the southern African interior with precise

ages and with major implications behind it. The paper provides a significant contribution to our understanding of the southern African MSA and uses a paired dating approach that is robust. The main contribution includes evidence of modern human behaviors (collection of calcite crystals, ochre) in what has been an overlooked region of southern Africa for study of the MSA and the development of modern human behaviors. Their analysis of the climate during occupation of GHN is more qualitative than quantitative, given the quantitative analyses are largely from modern/historical records and extrapolation from distant cores. The presence of the dated tufa layers alone corroborates the interpretation that there was fresh flowing water at GHN and the dates confirm that the presence of water coincides with occupation (from the OSL ages). In future work, I hope efforts are made to provide local paleoclimate/paleoenvironmental records at GHN. The dating is an important aspect of the innovation of the work, as U-Th dating of tufas tends to be a serious challenge at many sites. These tufas were carefully studied to optimize the U-Th dating system to work for the objective to help date the site. So, it is extremely encouraging that the U-series and OSL ages agree within expectations guided by the stratigraphy, and this in itself is an exciting contribution to dating applications in archaeology.
The chronological coincidence with other records evidencing a wetter Kalahari is very encouraging, and I'd encourage the authors to elaborate on what this may mean for potential exploratory work in the region.

I suggest that the manuscript undergo one major revision. The evidence of modern human behaviors is clear from the combination of calcite crystals, other archaeology, and the antiquity of the site, but the emphasis on the OES being parts of storage containers, and not simply food waste, is not well supported. The calcite crystals are interesting and unique for their age, so the site has added significant value to our knowledge of the archaeological record, particularly in the Kalahari. I wonder why the authors did not go into detail about the significance of the ochre piece with parallel striations more, as this garners attention from those studying the evidence for modern human behaviors. Rather they spend time interpreting ostrich eggshell fragments with no clear support for the interpretation that they were once parts of storage containers. I suggest that this paper undergo revision to limit and temper the arguments hinging on the OES being a part of storage containers. I would encourage the authors to highlight the dating, since the combined effort of the two techniques provides a powerful and robust indication of people inhabiting this landscape and exhibiting behavioral modernity as early as the evidence suggests at many coastal sites. I would also suggest, along this line, that the OSL ages be supported by radial plots in the extended figures to show the distribution of removed v. included grains in those ages. The methods are described but it is useful to see the scatter of excluded and included grains to see how the included population is fit to determine De.

Below I provide some detailed suggestions on the manuscript for revision:

Line 88: please state that these are weighted mean ages of multiple samples, state the number of samples, and state what the uncertainty is reporting: if the uncertainties on ages are reporting 1 or 2 times standard deviation or some other error envelope. I see it is in the data tables, but this should be in the text for each dating system (OSL and U-Th) when it is first reported. Especially because, the OSL ages are 1 sigma and the U-Th are 2 sigma in the tables – if anything, it would be best to conform to one error reporting format so readers can directly compare the ages with the appropriate uncertainty. So ideally, please change the OSL errors to 2 sigma and report that in the text.

Line 125-126: Seems like speculation of spiritual use of the calcite crystals. This would be better placed in a discussion section, in the context of evidence to support the claim.

Line 131: you make a great case for the OES being brought in by humans. I don't think you make the case that the OES are from containers. For this section, I suggest modifying the point of it to highlight that yes, these are brought by humans (as your sub-title says), but remove the interpretation that they are indeed containers. The evidence is not stronger than that supporting the OES being food waste.

Line 133: these are extremely small fragments compared to many sites that host decorated eggshells (clear evidence of modern behavior). Do any of them reconjoin?

Line 137-138: there are other ways for OES to become colored red, for example, Fe oxide staining, or precipitate minerals adhering to surfaces (for example, goethite). Did you test any of these with XRD or XRF to verify that the discoloration was not from a stain or precipitate? You can also verify there was heating by analyzing the organic component of the OES.

140: the wording here is a bit inaccurate. Hyenas and other animals don't consume OES (ostrich eggshell), they consume ostrich eggs and leave behind the eggshell. I'd suggest re-wording for clarity.

Lines 145-148: Are the OES you see in the deposit shown to be a part of a container? Based on figure 2, and the description of the samples, the OES you see in the deposit are not shown to be a part of a container. they are confidently ID'ed as "human collected" but there is no evidence to support them as container fragments rather than food waste. If you want to introduce this idea, I suggest you place it in a context of speculation and not assertion that GHN provides this evidence. I suggest that either this is removed entirely, or it is kept, and the introduction of the GHN materials is modified to say something like: "while it is not clear that these are part of storage containers, the evidence for storage containers exists at this antiquity".

Line 150-151: remove "the appearance of this innovative technology at GHN is coeval with that at the coast". Stick with the fact that they are brought there by humans.

Line 162: transition to modern climate information is a bit abrupt. Please add some transition sentence or phrasing here to introduce the motive for telling the reader why we need to know about the current climate regime, since it is the basis for interpretations on rainfall in the past. This could even be a new paragraph.

Line 180: but the OES are not evidence of containers, based on the images of them in Figure 2 and your description of them in the text. See further comments below addressing the arguments made in the SI. I suggest removing the assertion of the OES being parts of containers here.

Line 190: "Pan-African" normally refers to more than one region of Africa – e.g., eastern Africa, northern Africa – so I suggest removing it here, since there is no review of sites outside of southern Africa in this paper.

Figure 2: Ysterfontein 1 has fully human-accumulated midden layers full of OES as is described in Avery et al. (2008) which you have cited in the text and SI. This is also described in Klein et al (2004, PNAS). I suggest changing the symbol on Figure 2 for Ysterfontein 1 to include orange coloring as well as purple, since there are layers that may have been mixed human-animal accumulations too. This also means that suggestions of human-collected OES should include Ysterfontein 1 in the main text and SI (lines 161-164 in SI, for example, and SI Table 2).

Extended Data Figure 6, line 697: please state the units of annual precipitation provided in the map figure b.

Extended Data Table 1: why are the OSL ages of the other layers not included in this table? Are they in the reference 22 (Wilkins et al., 2020, J. Paleolithic Arch.)? If they are reported in this paper for the first time, they should be included here. If they are in the other paper cited, please disregard this comment.

Extended Data Table 3: Were the detritus corrections made also on U isotopes, i.e. using

(232/238) and (234/238) ratios from detritus? The table caption only mentions (230/232).

SI comments:
Lines 183-188: these percentages should also be compared to the relative abundances of known sites in these different age ranges. The differences may be a result of sampling bias.

Line 186: it seems that this ("known importance of OES storage containers in southern Africa through much of the Late Pleistocene, Holocene, and into the historical period") is the main argument for calling the OES at GHN as early evidence of the technological innovation of storage containers. I don't think this is a reasonable argument. At GHN, there are heated OES that do not reconjoin (at least as shown in Figure 2). There are not OES fragments with evidence of them being a part of a flask mouth, nor are the OES etched or painted. The rebuttals to arguments made regarding these facts will be addressed in other comments below, but in summary, it is the burden of the authors to nullify other potential hypotheses that these fragments aren't just food waste. I don't think that burden has been met.

Arguments 1 and 2 make some good points (for example, that not all storage containers were necessarily decorated) – but then it is the burden of the authors to show some other evidence that would strongly favor a storage container interpretation over other simpler interpretations. Instead, the argument is based on the possibility that they could have been storage containers – which is highly speculative – but they could also be from food waste, a much more simple explanation, and by that same line of reasoning.

That does not mean the evidence for modern human behavior at GHN, and by proxy, in the Kalahari, is not there - the calcite crystals are exceptional, and the age of the site makes the point that there are behaviorally "modern" people inhabiting inland regions. There is plenty here to still make the point that coastal resources aren't the necessary pin in the machine to make humans "modern".

I would also emphasize that part of the reason that perhaps there are so few sites >100 ka with clearly human collected OES because there have been major challenges dating sites in south Africa that are older than the 14C limit (~50 ka). Luminescence techniques have helped profoundly to resolve this issue, and the U-Th dates in this paper are an elegant compliment to the OSL ages with their own innovative potential, so perhaps there is hope to continue populating the archaeological record of the southern African interior with better dates. This is a strength of your paper and I encourage you to highlight that in your discussion. If you wanted to highlight that, you could add the sites on your map (main text Figure 2) that we know exist but are not well dated, or they are not dated at all.

Referee #3 (Remarks to the Author):

The authors argue that non-utilitarian behavior in modern humans is as early inland in South Africa as it is in coastal areas, although they define coastal areas as anything less than 100 ka, which I think is a stretch. Nevertheless, I think the paper is an important contribution.

My main expertise is in OSL dating, so my comments on methodology mainly concern that. In that regard, I think the OSL dating is under-reported. I am not sure how the dates were actually derived. A more extended discussion should be included in the supplementary materials. Other pieces of evidence for their argument are discussed in more detail there.

Other comments I have:
Line 74 – 100 km seems like a fair distance for people travelling by foot.
Line 86 – What is fabric analysis and how does it relate to primary context

Line 88 – I think it should be DBGS
Line 106 – You should say what a minimal slope wash zone within a Benn diagram is and how that relates to natural depositional processes (or not).
Line 148 – Figure 2 shows that OES was collected at Border Cave as early as 200 ka, so the statement here does not seem accurate.
Figure 2: I am confused by the diagonal bars indicating dating uncertainty. Could be explained more? Plus how does the dating uncertainty in the histograms of part B for the older time periods jibe with the certainty at Border Cave. How is human use distinguished from carnivore use at Border Cave? This seems a critical issue given the argument of most human OES use after 105 ka. The supplementary materials says not much is known about the Border Cave OES. That does not strengthen the argument.
Lines 266-268 – This section needs expanding to clarify how the dose distributions were interpreted. There is no such thing as over-dispersion of FMM results. FMM requires over-dispersion as an input; it is not an output. How was the proper equivalent dose determined where over-dispersion (as determined by central age model) was high? How was microdosimetry dealt with?
In the extended Table 1, I am unsure what the FMM sigma is? If that is the over-dispersion input, why does it differ from sample to sample. I assume the k represents the number of components? I am suspicious of using the most abundant FMM component as the best estimate of equivalent dose, when its proportion is less than 70%. More argument is needed to justify this.
Lines 269-270 – Gamma dose rates can be determined from alpha-counting only by assuming secular equilibrium.
Line 273 – Was the cosmic dose rate adjusted according to the configuration of the rock shelter.

I think the references ae fine, and the abstract and paper as a whole are well-written.

Author Rebuttals to Initial Comments:

Referee #1 (Remarks to the Author):

A. The authors argue that calcite crystals and ostrich eggshell pieces were collected by people over 100,000 years ago in the western Kalahari. Theses are argued to be evidence of an earlier appearance of symbolic behavior than proposed earlier. They also tie this behavior into evidence of a wetter period in this region, which today is quite arid.

B. This is an original and significant contribution, and ties together new discoveries with a review of evidence for similar technologies elsewhere in southern Africa.

C. The data and methods are appropriate for this kind of research. Some of the methods have not really been employed in such an analysis in the past,but are perfectly appropriate here. It is an excellent example of multi- and inter-disciplinary research.

D. The use of quantitative methods and statistics in appropriate.

E. The conclusions are appropriate. In the supplementary text, the authors try to evaluate their conclusions in the light of other possible explanations. This is an excellent way to approach this kind of data.

F. The references are appropriate and numerous. I note that the bibliographic entries are done differently for the main article as opposed to the supplementary text. I am not sure if this is the way Nature authors are supposed to create bibliographies, but, regardless, it is worthy of checking over.

In addition, some journal titles are missing from some references in the main text, but also in the supplementary documents.

All references have been corrected.

Also, in SI Table S2 OES(=ostrich eggshell) is used as a plural noun in some places and as a singular noun in others. Pick one option, and make it the same throughout the table.

Done.

There are a couple of other typos to be fixed, in addition to the absence of some journal titles, volumes, and/ or page numbers. In the same table, the first reference (Ames et al. 2020), the last authors initials are mixed in with the title of the journal. This appears as Collins, BJOQ. Collins should be NR, and the journal is Open Quaternary. In these bibliographies, some journal titles are italicized, while others are not. Please be consistent throughout. Pick one system and stick to it. Another typo - Steele et al. 2012. South Africa is misspelled as South Sfrica.

All typos have been corrected.

H. The text is quite clear and is a reasonable length. This paper should be published by Nature. It is a state of the art treatment of a critical issue - the origins and early evolution of symbolic behavior. Once again, Africa leads the way.

Thank you!

Referee #2 (Remarks to the Author):

The paper by Wilkins et al. describes a significant site in the southern African interior with precise ages and with major implications behind it. The paper provides a significant contribution to our understanding of the southern African MSA and uses a paired dating approach that is robust. The main contribution includes evidence of modern human behaviors (collection of calcite crystals, ochre) in what has been an overlooked region of southern Africa for study of the MSA and the development of modern human behaviors. Their analysis of the climate during occupation of GHN is more qualitative than quantitative, given the quantitative analyses are largely from modern/historical records and extrapolation from distant cores. The presence of the dated tufa layers alone corroborates the interpretation that there was fresh flowing water at GHN and the dates confirm that the presence of water coincides with occupation (from the OSL ages). In future work, I hope efforts are made to provide local paleoclimate/paleoenvironmental records at GHN. The dating is an important aspect of the innovation of the work, as U-Th dating of tufas tends to be a serious challenge at many sites. These tufas were carefully studied to optimize the U-Th dating system to work for the objective to help date the site. So, it is extremely encouraging that the U-series and OSL ages agree within expectations guided by the stratigraphy, and this in itself is an exciting contribution to dating applications in archaeology.
The chronological coincidence with other records evidencing a wetter Kalahari is very

encouraging, and I'd encourage the authors to elaborate on what this may mean for potential exploratory work in the region.

We would like to thank Referee 2 for the positive outlook and constructive feedback. In response to this suggestion about future work, the following sentence has been added to the Discussion section of the manuscript: "*The archaeological, dating, and palaeoenvironmental methods employed here set the foundation for further investigations in the Kalahari and other under-studied regions*."

I suggest that the manuscript undergo one major revision. The evidence of modern human behaviors is clear from the combination of calcite crystals, other archaeology, and the antiquity of the site, but the emphasis on the OES being parts of storage containers, and not simply food waste, is not well supported. The calcite crystals are interesting and unique for their age, so the site has added significant value to our knowledge of the archaeological record, particularly in the Kalahari. I wonder why the authors did not go into detail about the significance of the ochre piece with parallel striations more, as this garners attention from those studying the evidence for modern human behaviors.

The ochre piece is significant in that it is worked – the striated and facetted surface attest to this. However, evidence for worked ochre is present at African archaeological sites going back beyond 300 ka (Watts, Chazan, and Wilkins 2016; Brooks et al. 2018), so from a behavioural perspective it is not particularly informative for the period of time and question addressed in this manuscript.

Rather they spend time interpreting ostrich eggshell fragments with no clear support for the interpretation that they were once parts of storage containers. I suggest that this paper undergo revision to limit and temper the arguments hinging on the OES being a part of storage containers.

Done – see details below about how the argument now focuses on human-collected OES, with more tempered interpretation of what this may signify about container technology.

I would encourage the authors to highlight the dating, since the combined effort of the two techniques provides a powerful and robust indication of people inhabiting this landscape and exhibiting behavioral modernity as early as the evidence suggests at many coastal sites.

Done – we added a highlight to the dating in the concluding section (see above).

I would also suggest, along this line, that the OSL ages be supported by radial plots in the extended figures to show the distribution of removed v. included grains in those ages. The methods are described but it is useful to see the scatter of excluded and included grains to see how the included population is fit to determine De.

We have followed the suggestion Referee 2 and added radial plots for each sample as Extended Data Figure 2. This shows the equivalent dose distributions with the components fitted using the FMM.

We also provide a table showing the number of accepted and rejected grains per sample, based on the individual rejection criteria, included in Supplementary Data. This provides the reader with complete and detailed information about the single-grain dataset and – in combination with the up-dated OSL methodology – how this dataset has been treated to generate optical ages.

We feel that including grains in the radial plot that failed these criteria (as suggested by the reviewer), i.e. are obviously suffering from intrinsic OSL behaviour problems will neither improve the quality of nor help with interpretation of the De distributions visualized in these radial plots. It would rather obscure the actual data and make the radial plots hard to read. It is also not common practice in single-grain OSL studies and we therefore refrain from doing so.

Below I provide some detailed suggestions on the manuscript for revision:

Line 88: please state that these are weighted mean ages of multiple samples, state the number of samples, and state what the uncertainty is reporting: if the uncertainties on ages are reporting 1 or 2 times standard deviation or some other error envelope. I see it is in the data tables, but this should be in the text for each dating system (OSL and U-Th) when it is first reported. Especially because, the OSL ages are 1 sigma and the U-Th are 2 sigma in the tables – if anything, it would be best to conform to one error reporting format so readers can directly compare the ages with the appropriate uncertainty. So ideally, please change the OSL errors to 2 sigma and report that in the text.

We have clarified that the original age estimate of the DBSR was a weighted mean based on 3 samples. The other age estimates were based on single samples.

OSL dating convention is to report errors at the 1σ level, whereas U-Th is reported at the 2σ level. This is just a pre-existing idiosyncrasy of geochronology. We think it is clearer to report the OSL ages from GHN at 1σ in the manuscript, so that they are more easily comparable to other OSL-based chronologies, both in the text and in future publications. For example, this allows for direct comparison of the Ga-Mohana OSL ages with the numerous OSL ages established at MSA sites from the South African coastal region (e.g., Blombos Cave, Pinnacle Point, etc.), and generates a coherent spatio-temporal framework that is central to our study. We have clarified that our OSL ages are reported with 1σ uncertainties at the first mention of the ages in the text.

That being said, in the Supplementary Discussion we now also present the 2σ age range in the context of directly comparing it to the 2σ age range of the U-Th ages. We state: "*The ages obtained from each sample are within error of one another, giving 2σ age ranges between 113.6 to 103.7 ka for sample 18-16, and 104.4 to 99.7 ka for sample 18-17. While there is some overlap between them, the two samples represent a time span of ~10 ka and are coeval with the occupation layers of the DBSR. The 2σ age range for the archaeological deposits based on OSL is 111.8 to 98.6 ka.*"

Line 125-126: Seems like speculation of spiritual use of the calcite crystals. This would be better placed in a discussion section, in the context of evidence to support the claim.

The way we have structured the paper is to summarize the result and its significance at the end of each section. We believe this makes the paper more readable. We then elaborate on the interpretation in the Discussion section of the main manuscript and in the Supplementary Discussion. We would prefer to the leave the organization as is, especially as paper organization was not a comment brought up by the other reviewers. However, if the Editor agrees that we should reorganize the paper, we will do so.

Line 131: you make a great case for the OES being brought in by humans. I don't think you make the case that the OES are from containers. For this section, I suggest modifying the point of it to highlight that yes, these are brought by humans (as your sub-title says), but remove the interpretation that they are indeed containers. The evidence is not stronger than that supporting the OES being food waste.

Done – see details below where more specific comments are made.

Line 133: these are extremely small fragments compared to many sites that host decorated eggshells (clear evidence of modern behavior). Do any of them reconjoin?

The GHN OES fragments are not smaller than those at Diepkloof Rockshelter. Texier et al. (2013) report the frequencies for three size classes as 67% (< 10 mm), 30% (10-20 mm), and 3% (> 20 mm). At GHN, the frequencies for these same size classes are 50% (<10 mm), 40% (10-20 mm), and 10% (> 20 mm). Based on these data, the GHN fragments are significantly *larger* than those at Diepkloof (Chi squared = 9.307, df = 2, p = 0.0095). There are only two other MSA sites in southern Africa > 50 ka with decorated eggshells. At Klipdrift Rockshelter (Henshilwood et al. 2014) and Apollo 11 (Vogelsang et al. 2010), details about the metrics of the OES fragments have not yet been reported.

We have not yet been able to reconjoin any fragments, but this may be possible as excavations continue and more OES fragments are recovered.

Line 137-138: there are other ways for OES to become colored red, for example, Fe oxide staining, or precipitate minerals adhering to surfaces (for example, goethite). Did you test any of these with XRD or XRF to verify that the discoloration was not from a stain or precipitate? You can also verify there was heating by analyzing the organic component of the OES.

We used a 20X hand lens to observe each OES piece in its entirety. If the colouration was due to precipitate mineral, we would have observed the material with the hand lens, as you can see it adhering to the surface. Additionally, adhering material on OES is not expected to be consistent across the surface. Adhering material was not observed on any of the OES pieces.

When OES is heated there are changes in the structure of the shell (Janssen, Mutch, and Hayward 2011; Craig et al. 2020), especially in the capillaries (the "vertical" structures that connect the interior to the exterior). These observable traits were noted consistently in the OES assemblage, which is how we determined the colouration resulted from heat exposure. We did not observe similar orange-red colouration in the faunal assemblage, which one would expect if staining were an issue. Burning is further supported by the presence of charcoal and burned faunal remains throughout the DBSR. These three lines of evidence, plus the colour changes consistent with OES being exposed to different temperatures (Collins and Steele 2017), are strongly indicative that the OES was exposed to heat.

To make this clearer in the manuscript, these additional details about the identification of burning have been added to the methods section.

140: the wording here is a bit inaccurate. Hyenas and other animals don't consume OES (ostrich eggshell), they consume ostrich eggs and leave behind the eggshell. I'd suggest re-wording for clarity.

Yes. Thank you. This is corrected now.

Lines 145-148: Are the OES you see in the deposit shown to be a part of a container? Based on figure 2, and the description of the samples, the OES you see in the deposit are not shown to be a part of a container. they are confidently ID'ed as "human collected" but there is no evidence to support them as container fragments rather than food waste. If you want to introduce this idea, I suggest you place it in a context of speculation and not assertion that GHN provides this evidence. I suggest that either this is removed entirely, or it is kept, and the introduction of the GHN materials is modified to say something like: "while it is not clear that these are part of storage containers, the evidence for storage containers exists at this antiquity".

We have now concluded this section with this statement: "*We show here that the appearance of human-collected OES at GHN is coeval with that at the coast"*. The reviewer is correct to point out that our empirical evidence supports the human-collected conclusion, and that the potential usage as container technology is an interpretation. We have modified word choice in the manuscript and Supplementary Discussion to make this clearer.

Line 150-151: remove "the appearance of this innovative technology at GHN is coeval with that at the coast". Stick with the fact that they are brought there by humans.

Done.

Line 162: transition to modern climate information is a bit abrupt. Please add some transition sentence or phrasing here to introduce the motive for telling the reader why we need to know about the current climate regime, since it is the basis for interpretations on rainfall in the past. This could even be a new paragraph.

Done.

Line 180: but the OES are not evidence of containers, based on the images of them in Figure 2 and your description of them in the text. See further comments below addressing the arguments made in the SI. I suggest removing the assertion of the OES being parts of containers here.

This sentence was edited to: *"The novel findings at GHN allow us to push back evidence for collecting non-utilitarian objects and OES in the Kalahari to ~105 ka, which is coincident with the evidence in coastal contexts."*

Similar edits were made in the abstract to accurately reflect this argument.

Line 190: "Pan-African" normally refers to more than one region of Africa – e.g., eastern Africa, northern Africa – so I suggest removing it here, since there is no review of sites outside of southern Africa in this paper.

Done.

Figure 2: Ysterfontein 1 has fully human-accumulated midden layers full of OES as is described in Avery et al. (2008) which you have cited in the text and SI. This is also described in Klein et al (2004, PNAS). I suggest changing the symbol on Figure 2 for Ysterfontein 1 to include orange coloring as well as purple, since there are layers that may have been mixed human-animal accumulations too. This also means that suggestions of human-collected OES should include Ysterfontein 1 in the main text and SI (lines 161-164 in SI, for example, and SI Table 2).

Based on what is published, the bone assemblage at Ysterfontein (YFT1) is probably not fully human-accumulated. Klein et al. (2004: 5710) originally stated that "the abundant artifacts and intertidal shells pinpoint MSA people as the principal bone collector", while also acknowledging the presence of hyena coprolites and tooth marks from small carnivores and rodents. However, more recently, Avery et al. (2008: 72) clarified that hyena could have contributed to the bone assemblage, based on the presence of coprolites, as well as the types of species present. In addition, they assert that some of the OES fragments have marks consistent with hyena tooth damage, concluding clearly that "hyaenas may account for much of the eggshell at YFTl".

We have further clarified these details to Supplementary Table 2.

Extended Data Figure 6, line 697: please state the units of annual precipitation provided in the map figure b.

In the caption, we have indicated that the units of the contours lines are mm, and added a reference for the Winter Seasonality Index calculation (Walsh and Lawler 1981).

Extended Data Table 1: why are the OSL ages of the other layers not included in this table? Are they in the reference 22 (Wilkins et al., 2020, J. Paleolithic Arch.)? If they are reported in this paper for the first time, they should be included here. If they are in the other paper cited, please disregard this comment.

They were reported previously by Wilkins et al., 2020.

Extended Data Table 3: Were the detritus corrections made also on U isotopes, i.e. using (232/238) and (234/238) ratios from detritus? The table caption only mentions (230/232).

In the caption, we explain that a correction for initial $^{230/232}$Th of 1.5 ± 1.5 was applied and the uncertainty was fully propagated through using Monte Carlo simulation. We did not make any corrections on the measured U isotope ratios (232/238) and (234/238) as in our experience this is not routinely done, as the main issue with detrital contamination is with the Th isotopes and not the U ones.

SI comments:
Lines 183-188: these percentages should also be compared to the relative abundances of known sites in these different age ranges. The differences may be a result of sampling bias.

Here and in Figure 2, we compared the relative abundances of all dated published sites *with fauna* to show that sampling bias is not a good explanation for the difference. The reason it makes sense to only include sites with fauna (bone and/or shell) is because of preservation bias against organic material. Therefore, the absence of OES at sites that do not preserve any organic material is not very informative. However, the absence of OES at sites that *do* preserve organic material may reflect behaviour. We show that sites greater than approximately 100 ka lack secure evidence for human-collected OES, even though there was potential for OES preservation. To better illustrate this point, the y-axis on section b of Figure 2 was updated to display relative frequencies instead of number of sites. A small amount of text elaborating on this point was also added to the relevant section in the Supplementary Discussion.

Line 186: it seems that this ("known importance of OES storage containers in southern Africa through much of the Late Pleistocene, Holocene, and into the historical period") is the main argument for calling the OES at GHN as early evidence of the technological innovation of storage containers. I don't think this is a reasonable argument. At GHN, there are heated OES that do not reconjoin (at least as shown in Figure 2). There are not OES fragments with evidence of them being a part of a flask mouth, nor are the OES etched or painted. The rebuttals to arguments made regarding these facts will be addressed in other comments below, but in summary, it is the burden of the authors to nullify other potential hypotheses that these fragments aren't just food waste. I don't think that burden has been met.

Arguments 1 and 2 make some good points (for example, that not all storage containers were necessarily decorated) – but then it is the burden of the authors to show some other evidence that would strongly favor a storage container interpretation over other simpler interpretations. Instead, the argument is based on the possibility that they could have been storage containers – which is highly speculative – but they could also be from food waste, a much more simple explanation, and by that same line of reasoning.

The reason we believe that container technology is a better interpretation than food waste, is because of the pattern of OES usage after ~100 ka across southern Africa. And, at coastal flagship sites and site complexes with long sequences (i.e., Klasies, Pinnacle Point, Diepkloof), OES is absent or not reported in lower levels > 100 ka.

A few sentences have been added to this section of the Supplementary Discussion to further clarify, as well as some other updates to Figure 2, which are detailed below.

That does not mean the evidence for modern human behavior at GHN, and by proxy, in the Kalahari, is not there - the calcite crystals are exceptional, and the age of the site makes the point that there are behaviorally "modern" people inhabiting inland regions. There is plenty here to still make the point that coastal resources aren't the necessary pin in the machine to make humans "modern".

I would also emphasize that part of the reason that perhaps there are so few sites >100 ka with clearly human collected OES because there have been major challenges dating sites in south Africa that are older than the 14C limit (~50 ka). Luminescence techniques have helped profoundly to resolve this issue, and the U-Th dates in this paper are an elegant compliment to the OSL ages with their own innovative potential, so perhaps there is hope to continue populating the archaeological record of the southern African interior with better dates. This is a strength of your paper and I encourage you to highlight that in your discussion. If you wanted to highlight that, you could add the sites on your map (main text Figure 2) that we know exist but are not well dated, or they are not dated at all.

Absolutely, challenges with dating have left big gaps in our knowledge about MSA behaviours and palaeoenvironments, especially in the interior, and we recognize this as an important contribution of our research. As per the suggestion of Referee 2, this sentence has been added in the Discussion section: *"The archaeological, dating, and palaeoenvironmental methods employed here set the foundation for further investigations in the Kalahari and other under-studied regions."*

However, we'd like to reiterate that dating challenges are not a good explanation for the rarity of human-collected OES more than 100 ka. There are 14 sites in southern Africa with deposits and preserved fauna dated to between 100 ka and 200 ka. More than half of them have OES, but in all cases where the agent of accumulation has been considered, it is not human. More dated sites (and more taphonomic analyses) will certainly improve our understanding and could change the overall pattern, but the current observations are not explainable by a dating bias toward younger sites, because we are considering relative frequencies, not absolute counts.

As previously noted, the y-axis on section b of Figure 2 was updated to display relative frequencies instead of number of sites and a small amount of text elaborating on this point was added to the relevant section in the Supplementary Discussion.

Referee #3 (Remarks to the Author):

The authors argue that non-utilitarian behavior in modern humans is as early inland in South Africa as it is in coastal areas, although they define coastal areas as anything less than 100 ka, which I think is a stretch. Nevertheless, I think the paper is an important contribution.

My main expertise is in OSL dating, so my comments on methodology mainly concern that. In that regard, I think the OSL dating is under-reported. I am not sure how the dates were actually derived. A more extended discussion should be included in the supplementary materials. Other pieces of evidence for their argument are discussed in more detail there.

Thank you to Referee 3 for the helpful feedback. In response, we have added additional details about the OSL methodology and results, which we outline below after each specific comment.

Regarding the definition of coastal areas, the distance of 100 km is based on the archaeological record of hunter-gatherers. Whallon (2006) identifies the radius of the maximal regional bands during the Mesolithic of Europe as ~123 km. Paleolithic data on raw material transfers indicate that 100 km is an important threshold for raw material movement – raw material transfers less than 100 km are thought to reflect regular seasonal foraging trips (Gamble 1999; Feblot-Augustins 1993). Even if foragers ~100 km from the coast were not regularly directly accessing marine resources, they would be immediately adjacent to groups of foragers who did, and through the exchange of ideas and materials would be expected to share technologies. Therefore, to understand innovation in interior environments, it is necessary to look at sites significantly distant from the coast (i.e. GHN, > 600 km from nearest coast).

Other comments I have:
Line 74 – 100 km seems like a fair distance for people travelling by foot.

See above.

Line 86 – What is fabric analysis and how does it relate to primary context

For added clarity, 'fabric analysis' was replaced here by "the slope and orientation'. The methods of this fabric analysis are fully reported by the cited article - Wilkins et al. 2020.

Line 88 – I think it should be DBGS

Yes - corrected. Thank you.

Line 106 – You should say what a minimal slope wash zone within a Benn diagram is and how that relates to natural depositional processes (or not).

The Benn diagram method and results for the DBSR are fully reported by the cited article - Wilkins et al. 2020. This is becoming a common approach in archaeological site formation studies and this summary result will be familiar to many readers. Because the results of this analysis were reported previously, and not in this manuscript, it doesn't seem appropriate to elaborate on the approach here.

Line 148 – Figure 2 shows that OES was collected at Border Cave as early as 200 ka, so the statement here does not seem accurate.

Yes. Thank you. This required clarification. We have improved the presentation of these data in Figure 2 to also indicate where there is uncertainty with respect to the agent of accumulation, coupled with further information provided in Supplementary Table 2. Because detailed taphonomic analyses have not been completed at Border Cave, and there are carnivore remains in the sequence, it is now categorized as 'unknown' for agent of accumulation. Based on the same criteria, a few other sites have also shifted to this category, including some younger ones. However, the improved representation now shows more clearly the evidence for human-collected OES starting around 105 ka, with little evidence before that.

Figure 2: I am confused by the diagonal bars indicating dating uncertainty. Could be explained more?

Some text has been added to the caption to explain this more.

Details about why the line is hashed for individual sites has been added to Supplementary Table 2 where it had not already been stated.

Plus how does the dating uncertainty in the histograms of part B for the older time periods jibe with the certainty at Border Cave. How is human use distinguished from carnivore use at Border Cave? This seems a critical issue given the argument of most human OES use after

105 ka. The supplementary materials says not much is known about the Border Cave OES. That does not strengthen the argument.

See comment above about edits made to Figure 2 and Supplementary Table 2. The agent of accumulation at Border Cave is not securely known, and this is now reflected more clearly in the manuscript.

Lines 266-268 – This section needs expanding to clarify how the dose distributions were interpreted. There is no such thing as over-dispersion of FMM results. FMM requires over-dispersion as an input; it is not an output. How was the proper equivalent dose determined where over-dispersion (as determined by central age model) was high?

The reviewer is thanked for their comment and pointing out that this section can be more clearly expressed. The text has been changed to clarify. It now reads:

"distributions were analyzed using the Central Age Model (Galbraith et al. 1999) (CAM) and the Finite Mixture Model (Roberts et al. 2000) (FMM). To determine the age model that best fit each De distribution and the De value that best represents the deposition of sediments, a systematic approach was employed that considered the stratigraphic position and environmental context of samples, the overdispersion of CAM results, the FMM σ and k inputs, and two statistical measures (the maximum log likelihood and the Bayes information criterion) (Jacobs et al. 2008; Roberts et al. 2000; Cohen et al. 2015; Gliganic, May, and Cohen 2015). For each k-value, σ was increased until either (i) the lowest BIC was obtained or (ii) parameter redundancy was observed. This process was repeated for increasing values of k until a fit could not be achieved with σ > 0.10. The optimized k-σ combination with the lowest BIC, significant $llik_k$-$llik_{k-1}$ value (i.e., $llik_k$-$llik_{k-1}$ > 2[62,63]) and stratigraphic consistency was considered the most appropriate fit."

The approach for using, as inputs, the stratigraphic position, CAM overdispersion, and FMM σ and k inputs, and the BIC and LLIK as outputs is described in Cohen et al (2015) and Gliganic et al (2015), and again here.

How was microdosimetry dealt with?

Beta dose rate heterogeneity is an extrinsic source of potential scatter in single-grain De values that cannot routinely be resolved.

The deposits are characterised by dolomite and chert gravels (~20 mm diameter). Deposits are clast supported with a silty-sand maxtrix, from which our measured grains are derived. In this context, it is likely that some beta dose rate heterogeneity has affected the De distributions. However, any of the small potential effects of beta dose rate heterogeneity were accounted for with the relatively high σ values used in the FMM (see other comments for FMM clarifications).

In the extended Table 1, I am unsure what the FMM sigma is? If that is the over-dispersion input, why does it differ from sample to sample. I assume the k represents the number of components?

σ is the FMM overdispersion input and k is the FMM number-of-components input, both of which are conventional nomenclature for the FMM.

The σ parameter is different for different samples because each De distribution yielded a different best fit σ and k combination (as assessed using the BIC and LLIK – see above and Cohen et al 2015 and Gliganic et al 2015). We strongly disagree that σ must be the same for all samples – the overdispersion of single grain CAM data is rarely consistent between samples, and has clearly been shown to include components that are dose and sensitivity dependent and unexplained in laboratory settings (see Thomsen et al., 2012). The best fit σ and k combinations have σ values that are within the range of overdispersion values shown in laboratory dosed grains(Thomsen, Murray, and Jain 2012).

I am suspicious of using the most abundant FMM component as the best estimate of equivalent dose, when its proportion is less than 70%. More argument is needed to justify this.

The most abundant FMM component is a commonly used approach for deciding on a FMM component (e.g., (Bateman et al. 2007; Jacobs et al. 2008; Rhodes 2011). Please note that all ages, based on the most abundant best fit FMM data, are in stratigraphic order and, where ages come from the same stratigraphic unit, are statistically consistent.

Lines 269-270 – Gamma dose rates can be determined from alpha-counting only by assuming secular equilibrium.

Noted. This has been added to the text

Line 273 – Was the cosmic dose rate adjusted according to the configuration of the rock shelter.

Yes it was. This has been added to the text.

I think the references ae fine, and the abstract and paper as a whole are well-written.

References

Avery, Graham, David Halkett, Jayson Orton, Teresa Steele, Madelon Tusenius, and Richard Klein. 2008. The Ysterfontein 1 Middle Stone Age rock shelter and the evolution of coastal foraging. *South African Archaeological Society Goodwin Series* 10:66-89.

Bateman, Mark D., Claire H. Boulter, Andrew S. Carr, Charles D. Frederick, Duane Peter, and Michael Wilder. 2007. Detecting post-depositional sediment disturbance in sandy deposits using optical luminescence. *Quaternary Geochronology* 2 (1):57-64.

Brooks, Alison S, John E Yellen, Richard Potts, Anna K Behrensmeyer, Alan L Deino, David E Leslie, Stanley H Ambrose, Jeffrey R Ferguson, Francesco d'Errico, and Andrew M Zipkin. 2018. Long-distance stone transport and pigment use in the earliest Middle Stone Age. *Science* 360 (6384):90-94.

Cohen, Tim J, John D Jansen, Luke A Gliganic, Joshua R Larsen, Gerald C Nanson, Jan-Hendrik May, Brian G Jones, and David M Price. 2015. Hydrological transformation coincided with megafaunal extinction in central Australia. *Geology* 43 (3):195-198.

Collins, Benjamin, and Teresa E. Steele. 2017. An often overlooked resource: Ostrich (Struthio spp.) eggshell in the archaeological record. *Journal of Archaeological Science: Reports* 13:121-131.

Craig, Caitlin, Benjamin Collins, April Nowell, and Christopher Ames. 2020. The effects of heating ostrich eggshell on bead manufacturing: An experimental approach. *Journal of Archaeological Science: Reports* 31:102287.

Feblot-Augustins, Jehanne. 1993. Mobility Strategies in the Late Middle Palaeolithic of Central Europe and Western Europe: Elements of Stability and Variability. *Journal Of Anthropological Archaeology* 12:211-265.

Galbraith, R. F., R. G. Roberts, G. M. Laslett, H. Yoshida, and J. M. Olley. 1999. Optical dating of single and multiple grains of quartz from Jinmium Rock Shelter, northern Australia: Part I, experimental design and statistical methods. *Archaeometry* 41:339-364.

Gamble, Clive. 1999. *The Palaeolithic Societies of Europe*. Cambridge: Cambridge University Press.

Gliganic, L. A., J. H. May, and T. J. Cohen. 2015. All mixed up: Using single-grain equivalent dose distributions to identify phases of pedogenic mixing on a dryland alluvial fan. *Quaternary International* 362:23-33.

Henshilwood, Christopher S., Karen L. van Niekerk, Sarah Wurz, Anne Delagnes, Simon J. Armitage, Riaan F. Rifkin, Katja Douze, Petro Keene, Magnus M. Haaland, Jerome Reynard, Emmanuel Discamps, and Samantha S. Mienies. 2014. Klipdrift Shelter, southern Cape, South Africa: preliminary report on the Howiesons Poort layers. *Journal of Archaeological Science* 45:284-303.

Jacobs, Z., A. G. Wintle, G. A. T. Duller, R. G. Roberts, and L. Wadley. 2008. New ages for the post-Howiesons Poort, late and final Middle Stone Age at Sibudu, South Africa. *Journal of Archaeological Science* 35:1790-1807.

Janssen, Jennifer D, G William Mutch, and James L Hayward. 2011. Taphonomic effects of high temperature on avian eggshell. *Palaios* 26 (10):658-664.

Klein, Richard G., Graham Avery, Kathryn Cruz-Uribe, David Halkett, John E. Parkington, Teresa E. Steele, Thomas P. Volman, and Royden Yates. 2004. The Ysterfontein 1 Middle Stone Age site, South Africa, and early human exploitation of coastal resources. *Proceedings of the National Academy of Sciences* 101 (16):5708-5715.

Rhodes, Edward J. 2011. Optically stimulated luminescence dating of sediments over the past 200,000 years. *Annual Review of Earth and Planetary Sciences* 39:461-488.

Roberts, R. G., R. Galbraith, G. M. Laslett, H. Yoshida, and J. M. Olley. 2000. Distinguishing dose populations in sediment mixtures: a test of single-grain optical dating procedures using mixtures of laboratory-dosed quartz. *Radiation Measurements* 32:459-465.

Texier, Pierre-Jean, Guillaume Porraz, John Parkington, Jean-Philippe Rigaud, Cedric Poggenpoel, and Chantal Tribolo. 2013. The context, form and significance of the MSA engraved ostrich eggshell collection from Diepkloof Rock Shelter, Western Cape, South Africa. *Journal of Archaeological Science* 40 (9):3412-3431.

Thomsen, Kristina J., Andrew Murray, and Mayank Jain. 2012. The dose dependency of the over-dispersion of quartz OSL single grain dose distributions. *Radiation Measurements* 47 (9):732-739.

Vogelsang, Ralf, Jürgen Richter, Zenobia Jacobs, Barbara Eichhorn, Veerle Linseele, and Richard G Roberts. 2010. New excavations of Middle Stone Age deposits at Apollo 11 Rockshelter, Namibia: Stratigraphy, archaeology, chronology and past environments. *Journal of African Archaeology* 8 (2):185-218.

Walsh, R. P. D., and D. M. Lawler. 1981. Rainfall seasonality: Description, spatial patterns and change through time. *Weather* 36 (7):201-208.

Watts, I., Michael Chazan, and Jayne Wilkins. 2016. Early evidence for brilliant ritualized display: specularite use in the Northern Cape (South Africa) between ~500 ka and ~300 ka. *Current Anthropology* 57:287-309.

Whallon, R. 2006. Social Networks and information: Non-"utilitarian" mobility among hunter-gatherers. *Journal of Anthropological Archaeology* 25:259-270.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

All the changes and corrected I asked for have been done. This is an excellent, innovative paper, and should be published without delay. It is part of a long tradition which shows the complexity of Stone Age behaviours in Africa, and their relation to the question of the origins of modern humans.

Referee #2 (Remarks to the Author):

The reviewer appreciates the efforts to make revisions to the manuscript and considers the effort mostly sufficient to go forward with publication, with one exception. The authors need to change the interpretation of OES at YF1 as being only hyena (or carnivore) -accumulated. This is because many layers in this sequence were accumulated mostly by humans, and these layers include abundant OES. Avery et al (2008) interpret that humans were the most likely accumulators of the middle and lower layers at YF1, which has abundant ostrich eggshell (kilograms worth of material – see Avery et al. 2008, Table 4). Even if YF1 is not fully human-accumulated, humans were clear agents of accumulation there, in layers with lots of OES. So, it is very likely that humans accumulated OES In the middle and lower layers as there is other abundant evidence of human agents of accumulation in these layers. The dating is not good enough to separate the middle and lower layers from the upper layers, but since you have a color option in your figure to have "agents uncertain", why not at least use this color (black)? I suggest this change be made in the figure and wherever it is mentioned in the text.
Likewise, in the supplement, the percentages reported for the agents of accumulation probably include YF1 in the category of "hyena" – I suggest these be adjusted to have YF1 be included in the category of "agents unknown".
This is a temperament of my original request to include YF1 as "human-accumulated". Essentially, the accumulation agents are up for interpretation, and the current published interpretation is that the accumulation agents of OES are mixed, depending on stratigraphic context. Thus, "agents unknown" is the most suitable category to place YF1 in the context of this manuscript, given there is not space (or relevance) to this manuscript to go into detail as to why the agents are mixed, and why different layers may have differing predominant accumulation agents.

Referee #3 (Remarks to the Author):

The new submission is improved, but I still have a few concerns.
1. In my original review I suggested the authors explain more what they mean by a minimal slope wash within a Benn diagram. The authors response was that Benn diagrams are commonly being used in archaeological site formation studies and therefore most readers will know what they are talking about. Well, I am one reader who does not know what you are talking about. I think scholarly articles should be self-contained as much as possible, and readers should not have to seek out references to understand what is being said. This sentence should be expanded to explain this more.
2. In applying the Finite Mixture Model, the authors say they adjust the α (over-dispersion) and k (# of components) parameters until they get the best fit. But all that tells you is the best fit to the model, not that the model itself is appropriate. I think it is a misuse of FMM. I do not think α should be determined by the model itself. This value is the over-dispersion that would be appropriate for a single-aged sample. What that value should be depends on other information. The over-dispersion from a dose-recovery, for example, is a place to start.
3. The authors claim that the most abundant component of FMM is commonly used to determine the depositional age. But it is does not always represent the depositional age. There are examples where the most abundant component does not correspond to the age of deposition (e.g., Araujo et

al. 2017, Geoarchaeology (https://doi.org/10.1002/gea.21628). What I see from extended figure 1 is some strongly mixed samples. You can use stratigraphic relationships to try to sort this out, but you cannot just assume the most abundant component is the right one.

Author Rebuttals to First Revision:

Referee #1:
All the changes and corrected I asked for have been done. This is an excellent, innovative paper, and should be published without delay. It is part of a long tradition which shows the complexity of Stone Age behaviours in Africa, and their relation to the question of the origins of modern humans.

Thank you!

Referee #2:
The reviewer appreciates the efforts to make revisions to the manuscript and considers the effort mostly sufficient to go forward with publication, with one exception. The authors need to change the interpretation of OES at YF1 as being only hyena (or carnivore) -accumulated. This is because many layers in this sequence were accumulated mostly by humans, and these layers include abundant OES. Avery et al (2008) interpret that humans were the most likely accumulators of the middle and lower layers at YF1, which has abundant ostrich eggshell (kilograms worth of material – see Avery et al. 2008, Table 4). Even if YF1 is not fully human-accumulated, humans were clear agents of accumulation there, in layers with lots of OES. So, it is very likely that humans accumulated OES In the middle and lower layers as there is other abundant evidence of human agents of accumulation in these layers. The dating is not good enough to separate the middle and lower layers from the upper layers, but since you have a color option in your figure to have "agents uncertain", why not at least use this color (black)? I suggest this change be made in the figure and wherever it is mentioned in the text.
Likewise, in the supplement, the percentages reported for the agents of accumulation probably include YF1 in the category of "hyena" – I suggest these be adjusted to have YF1 be included in the category of "agents unknown".
This is a temperament of my original request to include YF1 as "human-accumulated". Essentially, the accumulation agents are up for interpretation, and the current published interpretation is that the accumulation agents of OES are mixed, depending on stratigraphic context. Thus, "agents unknown" is the most suitable category to place YF1 in the context of this manuscript, given there is not space (or relevance) to this manuscript to go into detail as to why the agents are mixed, and why different layers may have differing predominant accumulation agents.

We can concur there is uncertainty here. As Referee 2 suggested, we have indicated as such in the Figure, and all relevant sections of the supplementary material. It will be interesting to find out more about the OES at YF1 in the future.

Referee #3:
The new submission is improved, but I still have a few concerns.

1. In my original review I suggested the authors explain more what they mean by a minimal slope wash within a Benn diagram. The authors response was that Benn diagrams are commonly being used in archaeological site formation studies and therefore most readers will know what they are talking about. Well, I am one reader who does not know what you are talking about. I think scholarly articles should be self-contained as much as possible, and readers should not have to seek out references to understand what is being said. This sentence should be expanded to explain this more.

We have added this detail with references: "...Benn diagram, which summarizes clast slope and orientation data for known disturbance processes on a ternary plot (Benn 1994; Lenoble and Bertran 2004)".

2. In applying the Finite Mixture Model, the authors say they adjust the α (over-dispersion) and k (# of components) parameters until they get the best fit. But all that tells you is the best fit to the model, not that the model itself is appropriate. I think it is a misuse of FMM. I do not think α should be determined by the model itself. This value is the over-dispersion that would be appropriate for a single-aged sample. What that value should be depends on other information. The over-dispersion from a dose-recovery, for example, is a place to start.

See below.

3. The authors claim that the most abundant component of FMM is commonly used to determine the depositional age. But it is does not always represent the depositional age. There are examples where the most abundant component does not correspond to the age of deposition (e.g., Araujo et al. 2017, Geoarchaeology (https://doi.org/10.1002/gea.21628). What I see from extended figure 1 is some strongly mixed samples. You can use stratigraphic relationships to try to sort this out, but you cannot just assume the most abundant component is the right one.

We thank Referee 3 for a constructive evaluation. While we respect their perspective here, we remain confident that our OSL methodology and chronology are robust. In response to the two comments above (#2 and #3), we now provide additional details about the methodology and OSL results in the main manuscript. We (i) elaborate on our FMM approach in more detail and provide a detailed reasoning why we use model determined σ and k components rather than fixed parameters and underline these decisions with (ii) additional literature and (iii) an up-dated version of the Extended Data Figure 1. The latter is underscoring the robustness of our OSL ages and FMM approach by putting the De values and FMM σ and k components into relation to each other. Furthermore, we have added a new section to the Supplementary Discussion about the OSL analyses. There, we elaborate on the issues raised here, put our decision for using the best model fit for estimating De components into a broader context, provide the relevant literature on which we base our point of view, and evaluate alternative interpretations.

The new text now states:

Main manuscript – Methods

"The component with the largest proportion of grains was taken to reflect burial. While some studies have observed that in specific contexts minor FMM components better reflect deposition[69], the large majority of grains that comprise the main components (59-94%) and the stratigraphic agreement between main component DBSR ages strongly indicate that the main FMM component reflects burial in the GHN samples (Supplementary Discussion). The component with the largest proportion of grains was taken to reflect burial.

There are alternative approaches to determining the best σ and k parameter combination for fitting a given single-grain dose distribution via the FMM. For GHN, model constrained parameters[64-68] are the most appropriate. Alternative approaches include (i) usage of overdispersion results for substituting σ from a similar-aged, well-bleached, unmixed sediment sample in the same depositional setting[70,71], (ii) obtaining an overdispersion (σ) value from dose recovery experiments[69,72,73], or (iii) applying an arbitrary σ value, often taken as 20%[74,75]. In this study, no appropriate natural well-bleached samples were available to calibrate the σ-value and the dose recovery overdispersion of 0% clearly underestimates the overdispersion expected in the natural environment. Additionally, studies have shown that the overdispersion of laboratory-dosed single-grain De datasets determined by dose recovery experiments are dose and history dependent[76] and that the overdispersion of well-bleached, unmixed natural samples can range up to 40%[77-79], undermining the assumption that 20% is a representative overdispersion value for all well-bleached, unmixed single grain De datasets. Our sample-specific optimized FMM k-σ combination approach based on statistical parameters (BIC and llik) allows a more accurate overall chronology to be obtained, which we attribute to sample-specific intrinsic variability and beta microdosimetry. The vertical and lateral stratigraphic consistency of the GHN OSL ages (see Wilkins et al.[22] for OSL ages of the younger GHN layers) indicates that our approach to determining the appropriate age model is sound (Supplementary Discussion). It should also be noted that for GHN samples, the choice of FMM σ value between 20% and 35% does not significantly change the main component De value or resultant age (Extended Data Figure 1, Supplementary Discussion), further supporting that the occupational chronology is robust."

Supplementary Discussion

"The age of the DBSR deposit – OSL methodological considerations

Finite Mixture Modelling and ways of determining the overdispersion parameter

Various ways have been put forward to determine the best σ (over-dispersion; OD) parameter for fitting a given single-grain dose distribution via the FMM (Finite Mixture Model). OD values for well-bleached sediment samples (e.g. well-bleached aeolian sediments) have classically been said to range from ~10 to 20%, and a value of ~20% has thus often been used in OSL studies as a fixed input (σ) parameter for fitting the FMM[70,74,163,164]. However, overdispersion is much too complicated to be boiled down to an arbitrary 20% input value. Firstly, many studies have shown that the overdispersion of single grain OSL De distributions from well-bleached, unmixed settings can range up to 40%[77-79]. Secondly, Thomsen et al.[76] show that the overdispersion value due to intrinsic factors is dose dependent and dependent on the thermal history of the grains, with overdispersion values up to 46.7% for unheated grains given 208 Gy by a uniform gamma source. These studies indicate that overdispersion is not something that should be boiled down to an arbitrary value.

Another approach for obtaining a best-estimate of the σ parameter is the use of the overdispersion value as determined from a well-bleached sediment sample in the same sedimentary setting from which the dating samples are derived[71,75]. This approach also suffers from the same general issues discussed above (dose and sample dependence of OD). Even so, the closest that we have to this at GHN is sample GHN5 from the DBGS (see Wilkins et al. for further details [22]) – this sample was run through the fitting procedure as described here, and a 1 component FMM (i.e., the CAM) was determined as the best fit. The CAM for this sample yielded an overdispersion of 28+/-2%. There is no sedimentological evidence that the DBGS sediments are any better bleached than the DBSR sediments.

It has also been suggested to use the OD value from a dose recovery test as an approximation for the σ parameter[69,72,73]. Dose recovery tests are typically conducted by bleaching aliquots (via natural or an artificial light source) and then administering a pseudo-natural dose via a beta source that is mounted onto the OSL reader (such as those on the commonly used Risø TL/OSL readers). There is strong evidence that the OD value from a dose recovery experiment determined in this way (i.e. with the pseudo-natural dose being delivered via an unidirectional in-situ beta source) is significantly smaller compared to the same dose recovery experiment, where the given dose is administered via an external and uniform gamma source [76,165]. Generally speaking, laboratory dose recovery experiments with a uniform gamma source are more akin to dosing in a natural radiation field compared to unidirectional beta dosing in an OSL reader. Importantly, the experiments of Thomsen et al.[76,165]. suggest that an OD value based on dose recovery test using an *in situ* beta source does not fully capture the dose variability caused by intrinsic factors (intrinsic factors being parameters such as grain sensitivity, thermal history and dose response curve shape), and thus might not be adequate to describe the intrinsic variability of the well-bleached part of the natural dose distribution. Consequently, dose recovery tests using the *in situ* beta source only provides an underestimate for the σ (over-dispersion) parameter for FMM modelling. In addition, OD in well-bleached natural samples is often assumed to be the product of both intrinsic and extrinsic factors (extrinsic factors including, for example. beta dose heterogeneity on the mm to sub-mm scale). Yet, OD arising from external factors goes undetected in laboratory dose recovery experiments, which is another reason why we refrain from using OD values from a dose recovery test as a reliable input parameter for our FMM approach.

We acknowledge, however, that in some settings and for specific samples, the laboratory OD is probably sufficient to explain the variability of the well-bleached part of the natural dose distribution and thus yielding reliable dose estimates[69,72,73].

In this study, we are dealing with archaeological sediments, which are typically more heterogeneous on the micro-level compared to, for example, aeolian sediments. Hence a certain degree of OD arising from beta dose heterogeneity must be assumed for these deposits, which should be taken into account for the FMM approach. Unfortunately, no well-bleached sediments (such as unequivocally aeolian layers within the archaeological succession) are available from GHN to derive a best estimate for the combined intrinsic and extrinsic variability of the well-bleached part of the natural dose distribution. Dose recovery experiments[22] conducted on the single-grain level using a Risø TL/OSL DA 20 reader yielded on OD value of 0%, which is not uncommon for single-grain dose recovery tests conducted via an *in situ* beta source[63,166,167]. A gamma source is not available to us for use at GHN, as they are only operated by very few specialized medical research or radiation physics centres. Hence, a representative estimate for the OD arising from intrinsic factors of our samples that is based on a uniform gamma radiation could not be directly determined.

Thus, we have chosen an approach that relies on constraining the σ parameter from the FMM itself and from stratigraphic observations, which has been successfully used by numerous OSL studies before[65,66,78,166,168-171].

We opted for using the modelled σ parameters that best describe our single-grain De distributions on a sample-by-sample basis following David et al.[68], Jacobs et al.[65], and Gliganic et al.[67,172]. Although we use the modelled σ parameter for solving the FMM, the FMM solutions are not unconstrained, nor are they viewed in a vacuum. First, the result that all four FMM-derived ages from samples taken within the same depositional unit yield consistent ages is a strong indication that our approach is yielding a robust dataset. Second, our approach yields a stratigraphically coherent dataset (see Wilkins et al.[22] for ages of younger units – OAS unit yields SG OSL age calculated using the FMM of 30.9+/-1.9 ka and DBGS unit yields SG OSL age calculated using the FMM of 14.8 ± 0.8.

Furthermore, as we show in Extended Data Figure 1, the De value of the main De component of each sample is basically unchanged regardless of whether a fixed OD value of 20, 25, 30 or 35% is chosen. As outlined above, an OD value of 20% (representing a typical well-bleached sample) is frequently chosen as a σ input parameter, and up to 40% has been reported for other well-bleached unmixed sediments. Thus, even if we follow this simplified and conservative approach of FMM modelling and use either an assumed OD of 20% (based on outdated literature) or 30% (based on the single sample from GHN that yielded a best fit CAM age) the OSL ages for all DBSR samples remain consistent. The FMM ages using either the best model fit or a fixed OD value of 20% or 30%, respectively, for each sample are:

GHN 1 = 106 ± 6.7 ka, 110 ± 6.5 ka, 106 ± 6.6 ka

GHN 3 = 103 ± 6.5 ka, 103 ± 6.4 ka, 99 ± 7.2 ka

GHN 2 = 110 ± 5.7 ka, 111 ± 5.6 ka, 109 ± 5.8 ka

GHN 6 = 99.2 ± 7.7 ka, 113 ± 7.9 ka, 100 ± 7.2 ka

This being said, allowing the OD parameter to be optimized for each sample based on the statistical parameters (BIC and lik), allows a more accurate overall chronology. We attribute the varying best fit FMM σ values to sample-by-sample intrinsic variability and extrinsic dose variability (namely beta dose heterogeneity).

Evaluating major versus minor De components for FMM age calculation

Although minor De components might under some (rather specific) sedimentary circumstances be more likely to reflect the correct depositional age, it is for the majority of cases much more likely that the main De component is associated with the correct sediment burial age. In our FMM data set the main De component comprises between 59% and 94% percent of the entire De distribution and all ages overlap with each other on the 1 sigma error range (Extended Data Table 1).

The main De component of sample GHN 1 is made up of 84% of all grains; while for sample GHN 2 94% of all grains are part of the main De component, suggesting that these samples are not heavily mixed at all. Indeed, the CAM OD for sample GHN 2 is relatively low for archaeological sediments – i.e. 37%. For sample GHN 1 the CAM OD is 74%. However, for this sample four low De outliers (forming the low De component centred on ca. 10 Gy) are contributing over-proportionally to this relatively high CAM OD value, which falls to 58% if these grains – that are maybe intrusive grains from younger strata – are removed. We also show that (i) the OSL ages of the three archaeological layers are all in stratigraphic order[22], and that (ii) the there is a climate-nexus between human exploitation of this (usually arid to semi-arid) part of the African continental interior and wet episodes (U-Th dated to 110-100 ka) that is in very good agreement with our OSL chronology.

Furthermore, we have no sedimentary evidence or mechanism to argue that any of the low De components better reflect burial. The age profiles for all of the best fit FMM components are shown in the Supplementary Data. The less populous components from all four samples from the same stratigraphic unit are not statistically consistent, whereas the most populous components are. Likewise, the less populous components produce a chronology that is not stratigraphically coherent – the oldest secondary component ages (component 2 of sample GHN 3 in Supplementary Data) are from those samples closest to the surface.

The fact that (i) the ages calculated using the main FMM component from all four samples from the same stratigraphic unit are in agreement on the one sigma level with each other, (ii) that we observe vertical and lateral stratigraphic coherency of our optical ages, (iii) that the chronology using minor FMM populations is NOT stratigraphically coherent, as well as (iv) a convincing climate nexus of independently dated sedimentary processes and deposits is in our opinion strong evidence that the main De component does indeed reflect the depositional age of the DBSR archaeological layer and that our OSL chronology is robust."

Table included in Supplementary Data:

Sample		**Component proportion**	**Component De (Gy)**			**Dose Rate (Gy/ka)**			**Component age (ka)**		
GHN 3	comp 1	0.01	25.0	±	3.3	2.18	±	0.10	11.5	±	1.6
GHN 3	comp 2	0.32	110.0	±	5.0	2.18	±	0.10	50.5	±	3.5
GHN 3	**comp 3**	**0.67**	**224.2**	**±**	**8.2**	**2.18**	**±**	**0.10**	**103.0**	**±**	**6.5**
GHN2	comp 1	0.06	95.2	±	8.7	2.40	±	0.10	39.7	±	4.1
GHN2	**comp 2**	**0.94**	**263.9**	**±**	**5.7**	**2.40**	**±**	**0.10**	**109.8**	**±**	**5.7**
GHN 6	comp 1	0.06	27.3	±	4.4	2.56	±	0.12	10.7	±	1.8
GHN 6	comp 2	0.36	82.8	±	6.6	2.56	±	0.12	32.4	±	3.1

GHN 6	**comp 3**	**0.59**	**253.7**	**±**	**14.4**	**2.56**	**±**	**0.12**	**99.2**	**±**	**7.7**
GHN1	comp 1	0.02	9.6	±	2.4	2.50	±	0.10	3.9	±	1.0
GHN1	comp 2	0.14	70.3	±	7.7	2.50	±	0.10	28.2	±	3.4
GHN1	**comp 3**	**0.84**	**263.7**	**±**	**9.9**	**2.50**	**±**	**0.10**	**105.6**	**±**	**6.7**

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #2 (Remarks to the Author):

The authors have satisfied all reviewer suggestions and I believe this is ready for publication. Congratulations on excellent work and placing a new exciting site on the timeline of human innovation!

Referee #3 (Remarks to the Author):

My expertise is in OSL and at this point, after two revisions, I am confining my comments to that subject. In this regard, the latest revision is a major improvement. I still disagree with the authors' contention that statistics can be used to decide what the proper over-dispersion value for the finite mixture model should be. I do not think it is a statistical question, but the authors make a spirited defense of this approach that itself is worth publishing. There are limitations of other approaches (dose recovery, well bleached sample), but these reflect limitations in our knowledge of factors that control the distribution of single-grain De values, rather than improper statistics. Fortunately, most times, this case included, the FMM has wide tolerance for variation in what the input over-dispersion should be. Showing that is more convincing to me than arguing for a statistical best fit. Their further argument that the largest component reflects the burial age because no other alternative provides the right statigraphic order is a good argument for that. That argument was lacking before, but now that discussion is much stronger.
Their overall evaluation of the distribution of De values is an important contribution in itself, so I now urge publication.

Author Rebuttals to Second Revision:

<u>Referee #2:</u>

The authors have satisfied all reviewer suggestions and I believe this is ready for publication. Congratulations on excellent work and placing a new exciting site on the timeline of human innovation!

Thank you!

<u>Referee #3:</u>

My expertise is in OSL and at this point, after two revisions, I am confining my comments to that subject. In this regard, the latest revision is a major improvement. I still disagree with the authors'

contention that statistics can be used to decide what the proper over-dispersion value for the finite mixture model should be. I do not think it is a statistical question, but the authors make a spirited defense of this approach that itself is worth publishing. There are limitations of other approaches (dose recovery, well bleached sample), but these reflect limitations in our knowledge of factors that control the distribution of single-grain De values, rather than improper statistics. Fortunately, most times, this case included, the FMM has wide tolerance for variation in what the input over-dispersion should be. Showing that is more convincing to me than arguing for a statistical best fit. Their further argument that the largest component reflects the burial age because no other alternative provides the right stratigraphic order is a good argument for that. That argument was lacking before, but now that discussion is much stronger.

Their overall evaluation of the distribution of De values is an important contribution in itself, so I now urge publication.

We sincerely thank all three Referees for critically reviewing the manuscript, which helped to greatly improved its quality and clarity.