

Dear Authors,

We have now received three reviews for your preprint entitled “A way to break bones? The weight of intuitiveness”. We are pleased to say that these were all positive and agree that this is an important paper, which makes a relevant contribution to the study of bone breakage for marrow extraction and, by inference, of complex behaviours e.g. in the Palaeolithic.

Furthermore, the reviewers concur that the methods are sound and the conclusions justified and, as such, they recommend publication, provided some minor issues are addressed: - The clarity of the manuscript could be improved by subjecting it to another round of editing.

Sometimes your arguments are not sufficiently clear or ambiguously phrased. For example, what is “quasi-systematic” marrow extraction or “quasi-systematic conservation of the epiphyses”?

- The presentation of the figures and their captions needs to be improved - Various edits and requests for clarification should be dealt with (see below) Reviewer #1 (Alan Outram) points out that more space could be devoted to discussing future avenues of research, e.g. the impact of cultural preferences on this practice (cooking methods, different implements, presence/absence of periosteum...). Furthermore, reviewers #2 (Terry O’Connor) and #3 have provided more detailed comments, which are listed below.

We would ask you to revise the manuscript within one month, according to the comments of the three reviewers, and to submit the revised preprint, along with a detailed point-by-point response. We shall be happy to recommend it, pending suitable minor revision.

Looking forward to receiving your revised manuscript.

Kinds regards

Beatrice Demarchi & Reuven Yeshurun

Dear Recommenders,

We would like to thank you for considering our manuscript and to give us the opportunity to submit a revised version of our article, as well as the three reviewers for all the constructive and helpful comments.

We are pleased to submit the revised version of the research article entitled: “A way to break bones? The weight of intuitiveness”.

We hope that the corrections we made respond to the comments and questions of the reviewers and the recommenders. Please find our corrections and responses, as well as the revised version of the manuscript (a “track and change modifications” and a final version 3, <https://www.biorxiv.org/content/10.1101/2020.03.31.011320v3>). Bellow, we would like to answer you point by point the argument. We remain at their disposal should they have any questions concerning this matter. We thank you again for your consideration of our revised manuscript.

Yours faithfully,

D. Vettese, T. Stavrova, A. Borel, J. Marin, M.-H. Moncel, M. Arzarello, C. Daujeard.

Reviews

Reviewed by Terry O'Connor, 2020-04-09 19:40

A way to break bones? The weight of intuitiveness – Review

This is an informative report of detailed and thorough actualistic research. The larger research context is topical and timely, given the current interest in complex behaviours inferred from Middle Palaeolithic assemblages. Such experiments are never simple to organise and can usually be criticised for having too little replication of experimenter, material and procedure. However, this team developed a good pragmatic compromise between replication and practicability: it would be foolish to require hundreds of volunteers smashing many thousands of bones even if that did increase statistical rigour. The working methods are described clearly, with enough detail to allow the work to be repeated but without making the text longer than necessary. Similarly, the results are set out well and are properly separated from the discussion of those results and inferences drawn from them. The conclusions are, to be honest, not surprising. If asked, I would have expected the morphology of the bone to be a strong predictor of the pattern of pits and fractures, and that different people would tackle the task in quite different and probably inconsistent ways. However, that would have been informed speculation, at best, whereas this paper presents firm supporting evidence. The conclusion regarding counter-blows is particularly important, and often neglected in studies of ancient material. The last paragraph could, perhaps, have said more about the experimental procedure and how it might be refined or modified in any future research? In all, I enjoyed reading this paper and learned from it. Below, I have a few comments on particular details.

Thank you very much for your comments and suggestions. We took in account each of them and provide answers below.

Abstract: This needs to be more concise, and I strongly recommend that the authors edit the Abstract with the support of a first-language English speaker to ensure that the text says exactly what they intend. As it stands, I am not sure that this is the case. NB this applies to the Abstract, not to the paper as a whole.

We revised the abstract and reduced the number of words.

133-4: why give the standard deviation of the volunteer's ages? Is there any reason to think that they would be approximately normally distributed? "Eight men aged xx to yy" would be enough?

We would like to show the age distribution of our sample and to highlight the homogeneity. That is why we specified it.

Methods statement: good. This is thorough and detailed without including tangentially relevant material. It would enable another researcher to repeat the process with confidence.

216-236: nice account of the stats analysis, and sensible to have used Fisher's exact test.

241-255: although this is interesting and perhaps unsurprising, all of the volunteers came from a culture with chairs and other forms of seating. It is questionable, therefore, whether these results can be used to propose or compare positional behaviour in any non-chair group such as Neanderthals.

Even with a culture with the chair, only two experimenters preferred the sitting position (but without chair). The main result was the variability of the posture selected along with the experiment by each experimenter. The use of actualism is quite complicated for Neanderthal individual, even more regarding behaviours. Taking into account the presence of this variability even with a particular culture could suggest posture plasticity regarding this specific task for novice individual. Besides, some experimenters developed some habits and their habits were different according to the individual. This result could suggest the development of some particular practice within isolated groups.

312-325: I realise that this would have been difficult to parameterise! There is a difference between a task that is 'hard' because it is difficult to undertake and one that is 'hard' because it requires a lot of physical effort. For example, lace-making is 'hard' because it is complex and requires intricate hand movements, but not 'hard' because it is tiring or requires strength. Chopping logs is the reverse. Does the scale of difficulty used here integrate those two rather different forms of 'easy/hard', or is it one more than the other?

Yes, it was in a way quite subjective. In this experiment context, it was the second one taking into account the tiring and the difficulty to break bones (required strength or number of blows). After the second or the third try, some experimenters changed their evaluation according to their previous tries. The idea was to evaluate if, for each novice, we could observe a pattern regarding the whole bone series according to the experimenter. Then, we compared the different patterns acquired.

340-346: that's really interesting.

383-6 and 418-23: does this indicate a significant degree of inter-observer variation in the recording? If so, this needs to be said more explicitly.

383-386: No, it is not an indication of inter-observer variation.

It seems to have an inter-observer variation regarding the number of pits and grooves. The observer 1 and 2 worked together verifying if they agreed on the type of percussion marks and which marks recorded. We need to test further and evaluate the degree of inter-observer variation. Besides, to assess the inter-individual variation, we use spearman's rank correlation analysis between the number of blow and percussion marks (all, pits, Cna). The results were scarcely significant (p-value <0.05) and seem not to relate to the observer (following table).

Individuals	Blow_PercussionMarks		Blow_Pits		Blow_Can	
	p-value	rho	p-value	rho	p-value	rho
1	0.1439	-0.506	0.7606	0.1108	0.0458	-0.6411
2	0.0969	0.5535	0.09918	0.5505	0.07245	0.5902
3	0.3902	0.3058	0.4088	0.2945	0.7029	0.1385
4	0.2513	0.4001	0.6929	0.1433	0.1573	0.4831
5	0.0518	-0.628	0.04346	-0.6463	0.6696	-0.1547
6	0.0623	0.6079	0.1662	0.4742	0.08065	0.5771
7	0.5367	0.2242	0.4765	0.2553	0.7621	0.11
8	0.0226	0.7056	0.1098	0.5366	0.01206	0.7524
9	0.1989	0.4438	0.171	0.4695	0.5724	0.2473
10	0.6383	0.1702	0.7635	0.1094	0.2473	0.4037
11	0.7507	-0.1155	0.8803	-0.0549	0.2351	-0.4134
12	0.2751	0.709	0.05582	0.62	0.04403	0.6607

418-423: The observer 1 recorded the first four series and the observer 2 the other element series. Both observers recorded at least a series of each element. The whole results show similar tendencies regarding the distribution of percussion marks.

Figure 3: that's good. Not only the PCA plot, but indication of which variables were loading which dimensions of the plot.

444-6: check the caption to Fig 5. The labelling does not match the figure.

We changed the Figure 5, following the third reviewer advice and modified the caption appropriately.

484-6: caption to Fig 9 is quite difficult to understand. Can this be improved by labelling specimens within the fig? Same for Fig 11.

We changed the illustrations (Figures 6 – 12) adding the number of individual.

538-9: how did experimenters 'clean' their hands? I would expect them to rub their hands with dust/dirt to congeal and remove fat plus sweat. Maybe that's just me.

Some experimenters try to rub their hands with dust, but most of them used their labcoat or tissues.

This paper is impressive in a number of respects. Firstly, the fracture experiment sample sizes are very considerable and this constitutes a very substantial volume of work and one of the larger datasets in bone fracture experimentation. Secondly, the GIS recording method is innovative and useful and likely to be adopted by others. The methodology used is very rigorous and the data are displayed well. The paper usefully recognises the value of investigating impact placement on bones as a way of understanding the practice and cultural/cognitive differences between different hominins and different groups of AMH. The paper establishes an expected pattern based upon the intuitive behaviour of modern human subjects. These experiments usefully establish that bone morphology has a greater effect upon placement of impact for some elements than others and therefore provides a useful comparative framework for whether behaviours observed might be pragmatically determined or the result of specific cultural learning or particular activities. Clearly the data need to be used as a point of reference rather than deterministically. The authors fully appreciate this.

It is clear that this body of work is of value and the methods may be applied by others. The paper is clear about the parameters used, but could potentially, in discussion, flag up further issues that might come into play in interpretations of patterns and areas where further experiments are needed. This set of experiments is aimed at marrow extraction on fresh bones, with periosteum present using a hammer stone. This practice could, indeed, be an expression of cultural preference, just in the same way that meat butchery patterns can represent a form of material culture. However, further discussion could be made (flags for future investigation) of issues such as various cooking methods (roasting or boiling rather than only fresh bones), different implements for breakage, effect of prior periosteum removal (noted in several ethnographies) and breakage function being directed at grease rendering, bone meal, craft activity or bone fuel etc. instead of marrow extraction alone.

In places, some idiomatic edits could be made to the use of English.

Thank you very much for your comments and suggestions, we made the changes and add sentences in the conclusion parts:

Line 750: 'Besides, to go further and specify our results, more experiment will be required with various cooking methods, like roasting or boiling bones. Furthermore, the effect of prior periosteum removal will be tested regarding the percussion marks formation or their number, and also, the breakage efficiency.'

Line 756: 'Moreover, it may be helpful to integrate our results regarding the whole 'chaîne opératoire' of the bone breakage. Indeed, breakage function could be directed at grease rendering, bone meal, bone fuel or craft activity instead of marrow recover alone.'

Reviewed by anonymous reviewer, 2020-05-01 09:48

In this important paper, Vettese and colleagues argue that there is an intuitive way to break a bone in order to obtain its marrow. They employ unexperienced people to break the bones. Then, by applying GIS method to identify percussion mark concentrations, they point to the existence of bone morphology constraints and thus to an intuitive way to break the bones. The

manuscript is very interesting and the fundamentals of the work – i.e. the experiment execution and the taphonomic and GIS analyses – are of high quality. The methods and data presentation are mostly detailed and the conclusions are justified by the data. I recommend publication if two main issues are taken care of:

1) The paper is not clearly written and thus, very difficult to follow. I strongly recommend subjecting the text to professional language editing.

We did it, we revised some mistake and the abstract was edited.

2) Every figure and table (including supplementary) must be independently understandable and the legend should include a title, clear explanation, abbreviations and reference to the source of the data.

We made the changes.

In addition, I have several suggestions and points to fix, detailed below.

Terminology – use of terms interchangeably (e.g., Individual - experimenter - volunteer). Please be consistent.

We made some changes to be consistent.

Figures 4–13 – n blows is needed at least for every bone if not in all elements.

We add this information in SI 14.

Figures 4–5 – it is impossible to differentiate between the shapes and colors.

As suggested, we changed the figures (4 and 5) with only the merged series of each element (Figure 4). The individual series are moved in SI (10, 11, 12, 13).

Figures 6–13 – 8 figures is way too much, at least half should removed to the supplementary.

We would like to keep all the figures of Hot Spot analysis results. We think each of them is useful to illustrate the patterns discussed in the text. Moreover, they also allow highlighting the intra- and inter-individual patterns.

Experimenter interviews – are missing, you could obtain a lot of supporting data (e.g. why did you hit/hold/use anvil etc.). Why didn't you ask those questions? If you did, please add the information where needed or add questionnaires to supplementary.

We did not perform specific interviews of the experimenters. However, we said to them that they can explain their choices or gesture during the video recording. Therefore, we did not have standardized answers. We agree that these data would be of interest to go deeper in the understanding of the relationship between the marks and the intentions of the experimenter. However, questioning to much the experimenters could bias this study which focuses on intuitiveness. Asking questions during the experiment could make the experimenters thinking too much theoretically about their action. Post-experiment interviews could be carried out in front of the videos for example but this would address a question that, we think, is out of the scope of this paper.

SI – references to SI are unclear (e.g. 286 "Supplementary Information 1, 2, 3 and 4") is it table or figure 1? And be more precise, you can't refer to 4 figures). Some figures referenced in the text are missing (358–359 Supplementary Information 6).

We made these changes clarifying the SI in the manuscript and in the SI documents.

Abstract

Too long (500 words)

We revised the abstract and reduced the number of words.

35–36 – "Indeed, a previous study was able to show the possible existence of intuitive patterns of distribution of these traces according to the elements" – then I expect an explanation what have you added on it.

We studied more series (two more for each element and therefore more remains). The first study showed that each individual developed a way of breaking bones. The comparison in this study, although relatively small with three individuals, showed that for three elements we have a convergence (a pattern of percussion mark distribution), so we have confirmed the first results of the article by Stavrova et al. 2019; except for the femur.

43 – what is innovative in the method?

The previous study tested several spatial analyses to propose the most accurate one. The paper of Stavrova and al. 2019 showed that the Hot Spot analysis is the best because it takes into account (in 2D) the four sides. This analysis was never used for this purpose before the paper.

Based on the results, we enlarged the number of bone remains studied with two more series for each element. We would like to confirm the hypothesis and verify the existence of inter-individual variation. Moreover, we would like to compare the experimenter behaviour with the percussion marks distribution.

Introduction 72–75 - wasn't noted before 2012?

Following the suggestion, we added some earlier references: Martin 1907, White 1992 and Dart 1971.

79 – after 1992 -) missing

We corrected it.

101–102 – "The definition of a butchery tradition is a systematic and counterintuitive pattern shared by a same group" – First, add reference or detailed explanation. Second, why must a tradition be counterintuitive? If a group applies systematic and intuitive patterns for generations, you will not define it as tradition? 106–107 – ditto.

We define the butchery traditions as a repeated counter-intuitive practice passed on within a group over several generations. If we specify counter-intuitive it is because we need to show that it is an acquired practice. Indeed, if it is not possible to distinguish intuitiveness in practice, it is also not possible to prove the transmission of this practice. In other words, we agree that traditions could consists in intuitive patterns but such traditions, by definition, could not be identify and/or demonstrated as they would not differ from an intuitive, not traditional pattern.

Material and Methods When and where were the experiment carried out?

We added line 118: 'We performed this experiment in a designated area in the property of the Museum national d'Histoire naturelle (Paris).'

Is it a different experiment then the one reported in Stavrova et al. 2019?

No, it is the same experimentation, we broke 400 elements, there are still bones to study and this experimentation will lead to many other studies.

145 – Shouldn't the protocol be a part of this paper? Why put it separately?

We have chosen to separate it, so as not to unnecessarily lengthen the already long paper. Moreover, it allowed us to detail the protocol so that it could be replicated by others. <http://dx.doi.org/10.17504/protocols.io.be7fjhjn>.

153 – size = height?

We made this change.

162 - of each experimenter or each bone?

We grouped the remains by bones in one bag.

(cf. Protocole: D. Vettese 2020. An archaeological experiment focused on the intuitive way of long bones breakage to extract marrow. protocols.io <https://dx.doi.org/10.17504/protocols.io.be7fjhjn>).

183–186 what is the difference from Marean and Spencer 1991? Why reinvent the wheel? If not it, why not following other known methods which compatible to other research? (e.g. Dobney and Rielly 1988; Lam et al 1999 ect.) and if using new one, why not following Saladié et al 2011 which have done the same based Pales?

The illustration could may have led to a misunderstanding, it is similar than used by Marean et al. 1991 and Saladié et al. 2011 and it is compatible with other researches. We choose to illustrate with a bone of bovid. Moreover, we followed the same nomenclature of Saldié et al. 2011 or Blasco et al. 2013 regarding the portion and the side. It is not a new one, it is a replication adapted on bovid bone.

Moreover, Castel and Grunwald are missing in the bibliography and anyhow referencing a private website is unacceptable.

The illustrations (for the Figure 1, 4, SI 10, 11, 12, 13) were previously published in the following protocol <https://doi.org/10.1371/journal.pone.0216733.s004> . We changed the caption.

241 – Figure with photos will be very helpful. Consider adding videos like Camaros et al 2013 for example. Camaros, E., Cueto, M., Teira, L.C., Tapia, J., Cubas, M., Blasco, R., Rosell, J., Rivals, F., 2013. Large carnivores as taphonomic agents of space modification: an experimental approach with archaeological implications. Journal of Archaeological Science 40, 1361-1368 <https://doi.org/10.1016/j.jas.2012.09.037>

We planned to make a documentary separately and we would like to edit the numerous videos we have. Following the suggestion, we added photos in SI 1, 2.

256–260 – may be partly belong to the discussion? clarify

We disagree because it is just results without interpretation. We note observations of which hand the individual choose to break the bone during the experiment.

276 – général - change to English

We made the changes.

280–281 - isn't it expected since it contains the lowest amount? Shouldn't it be the amount of the possible marrow to extract? See line 349 below too.

We answer at these questions below. (line 349).

286 – too general.

It is important to notice that, for experimenters the fact to evaluate the marrow quality could influence the way to break bones and recover marrow.

289 – what is the title of this table?

We made the changes.

300–301 – discussion?

They are results without interpretation.

316 – beginning In summary?

We modified by 'In conclusion,'.

320–322 – discussion.

We disagree and keep it in the results.

323 – how do you infer it out of table 5?

We changed, indeed, we inferred that from the SI 3, 4 and 5 (number of blow progression, marrow extracted).

341 – percentage or numerical? Consistency.

We added the percentages and numerical when it was necessary.

349 – You calculate efficiency by blows to marrow weight, but since each bone type (i.e. femur, radius) contains a different amount of marrow to begin with, your index should be normalized by bone marrow content.

We are aware of this difference between the elements tested. However, we were able to observe among the different elements proposed to the same individual, there was a difference in size, in the quantity of spongiosa in the bone that could lead to a difference in the quantity of marrow. It was very complex without prior treatment to know the exact marrow content of each bone. For these reasons, we consciously chose to test the empirical hypothesis of whether the fact of having intrinsically less accessible marrow (as for the radius, where the marrow is contained in the proximal diaphysis for example) leads to a multiplication of the number of blows, of remains produced. It is, of course, possible to look at the amount of marrow extracted on average for each type of element, and to normalize it by an average capacity. However, we can observe that there is on average more marrow contained in the femur, but it is not the bone from which the most marrow has been extracted. And we think that is an important issue to explore further. While recalling that the experimenters were novices, and some before their first bone, they did not really know where the marrow was and therefore where to place the blows.

357–358 "the Wilcoxon signed rank test did not show a significant difference between the first five and the last five tries or between the first three and the last three tries." There was no significant difference between the first five and the last five tries or between the first three and the last three tries (Wilcoxon signed rank test; $p < 0.05$).

We changed by p -value > 0.05 ; lines 357 -358. And we corrected Wilcoxon in these lines.

358 – SI table or figure? Anyhow it is missing.

We changed the names of SI to simplify and clarify them in the manuscript and in the SI document.

366 - add number of other elements

We made the changes “The NSP of the tibias (32%) is higher than for the other elements, which are almost identical (humerus: 25%, radio-ulnas: 23% and femurs: 21%)”.

370 – SI figure 4, should rotate the figure properly.

We rotated it.

375 - SI 7 missing

We changed the names of SI to simplify and clarify them in the manuscript and in the SI document.

Bone fragmentation – I miss fragment length which is a very common index in Paleolithic archology.

Yes, it is, we planned to present these results in a future paper.

434–436 why did you separate the different PM?

Because we wanted to differentiate between percussion marks that are created during fracturing and marks that are caused by the rebound effect and do not result in fracturing. We thought it would be wise to differentiate between them in the case of novice experimenters who do not fracture the bone on the first blow or who need a lot of blows before breakage. Moreover, these two types of marks have differential conservation, with pits and grooves (surface marks) being more prone to be affected by surface alterations, unlike notches, adhering-flakes and crushing marks.

453 – omit parentheses

We changed it in the manuscript.

455–456 you can make it easier for the reader and refer to figure 6c.

We made the changes according to the reviewer suggestion.

Discussion It is improper adding new information that was not mention in the results – e.g. 1 526–528; Some volunteers applied reduced force when they hit the bones, especially during the first attempts.

We made this change: ‘Some volunteers could applied reduced force when they hit the bones, especially during the first attempts.’

e.g. 2 536–539; it is important to note that sometimes the hammerstone slipped on the cortical surface or slipped out of the experimenter’s hand. This happened especially when the hammerstone and hands were covered in grease after several tries. When an individual broke several bones, he/she had to clean the tools or his/her hands to pursue marrow extraction.

It is additional observations not results. We did not make the change.

566 - it is the assumption of the observer or they actually said it?

We changed the sentence: ‘The use of both hands to grasp the pebble could be an expression of the difficulty felt by the individuals breaking radio-ulnas.’

References Consistency - choose your preferred bibliographic style and follow it e.g. Is it ordered later-earlier or the opposite? See Binford 1981, 1978 but Costamagno 1999, 2009, 2014. While Blasco 2014, 2012, 2013?!

Castel appear twice.

Stavrova et al 2019 is missing.

922 – Vettese – different font.

We corrected and updated the references.