

Reply to : ‘In the eye of the beholder: contextual issues for Bayesian modelling at the Middle-to-Upper Palaeolithic transition’, by Discamps, Gravina and Teyssandier (2015).

T.F.G. Higham¹ and Godfrey S. Heep^{2,†}

¹ ORAU, RLAHA, University of Oxford, Oxford OX13QY, United Kingdom

² 36 Flora Dora Parade, Lake Hawea, New Zealand

A cornerstone of science is replicating and corroborating results to ensure that they are real and can be trusted. For this reason it is always welcome to see that others test data that have been produced to examine whether the results stand up to scrutiny. One recent example, published in the journal *World Archaeology*, focuses on aspects of our previous work (Higham et al. 2014), which was concerned with the chronology of the last Mousterian and the disappearance of Neanderthals. The article, by Discamps et al. (2015), claims that several of the sites we analysed are sensitive to slight changes in the Bayesian priors and data included, and that the results cannot be replicated. Specifically, they suggest that some of the Bayesian models we built from the 40 sites we worked on across Europe, are prone to significant variations which mean that the results cannot be trusted and used to reliably date the Middle to Upper Palaeolithic transition. If these assertions are true it follows that this would seriously undermine our overall conclusions, and so naturally we were interested in investigating this further.

Discamps et al. (2015) examine some of the Bayesian modeling results from sites including Abric Romaní, Les Cottès and La Quina, challenging the idea that some Mousterian contexts post-date chronologically the start of transitional industries such as the Châtelperronian. Examination of these sites highlights a number of issues of importance when undertaking Bayesian modeling and testing the resulting models.

Les Cottès: the importance of testing a model by changing one parameter at a time

At Les Cottès, a site published by Talamo et al. (2012), Discamps et al. (2015) argue that there are significant differences in the modeled results once they remove non-humanly modified bones from the US8 context at the site. They say that once this is done the models “produce a considerably older modelled age for the end of the Mousterian”. We tested this by taking our original model (Higham et al., 2014), removing these unmodified bone results and rerunning the model. The results are shown in Figure 1 (the model boundary without the non-humanly modified bone is termed ‘Discamps model’):

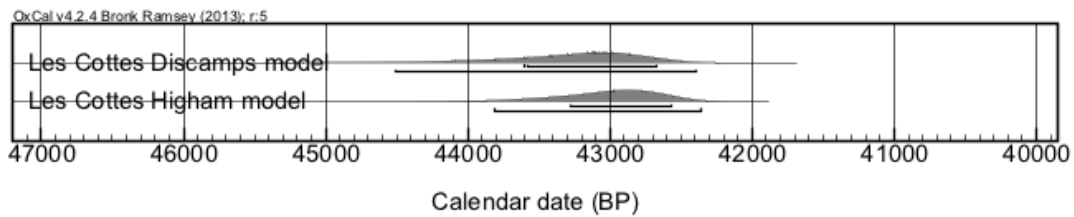


Figure 1: Boundaries for models from the Les Cottès site for the latest Mousterian upon removal of non-humanly modified samples of bone (Discamps model) and the inclusion of all determinations (Higham et al. 2014 model). All models were analysed using OxCal 4.2 (Ramsey, 2009) and the INTCAL13 dataset (Reimer et al., 2013).

The two PDFs show no statistical difference between them. On the basis of these results we conclude that it makes no significant difference whether you include or exclude the non-cutmarked bone results. We further conclude that there must be other variables that Discamps et al. (2015) must have changed in their model in addition to this in order to obtain the differences they observe. A perusal of their published data in their Figure 1 seems to confirm this because it shows that between the sequences of phases in the model there is no double boundary included. Discamps et al. (2015) attempt to justify this by suggesting that a single boundary is required; “on the assumption that the end of one layer and the beginning of next are strictly the same event”. This is not always true, which is why two boundaries are used to represent cases where there may have been a gap between the start and end of two phases, for example when two layers are separated by a sterile layer. By inserting a single boundary, the authors essentially constrain the Mousterian and Chatelperronian to be superposed one immediately atop the other. Talamo et al. (2012), however, state that; “The Châtelperronian (Unit 06) is separated from the preceding Mousterian by a 12-15 cm sterile unit”. This means that the model ought to have double boundaries between these phases if it is represent the archaeological sequence correctly. Leaving this out means that their model is different in more than one respect from ours. There are other differences of importance as well. While Discamps et al. (2015) state that they only included humanly-modified data they do not include in the Mousterian level 8 SEVA-13679 or SEVA-13680, both cut marked bone. Why is not explained, but once again it does raise some doubt about their conclusions and method, the reproducibility of the results and the validity of their comparisons with our own work.

To test the significance of a given parameter on a result within a Bayesian model, it is important that only one parameter is changed at a time. As shown above, when the model is kept the same and only cutmarked bones included, we find that the results are *not sensitive to their exclusion*. Discamps et al. (2015) appear to have changed several parameters (both priors and data) in order to reach a quite different conclusion. While we agree that the non-humanly modified data should be left out of the model (*contra* Talamo et al. 2012), in this case, leaving them in makes no significant difference to the overall result.

Abric Romani: statistical significance and outlier analysis

It is also important to appreciate the meaning of statistical significance, and how to test for it within the OxCal program. This is neatly demonstrated by figure 2, taken from Discamps et al. (2015), who use it to claim that the removal of two radiocarbon dates from the Mousterian of Abric Romaní, Spain results in a “significantly older boundary” for the end of the Mousterian.

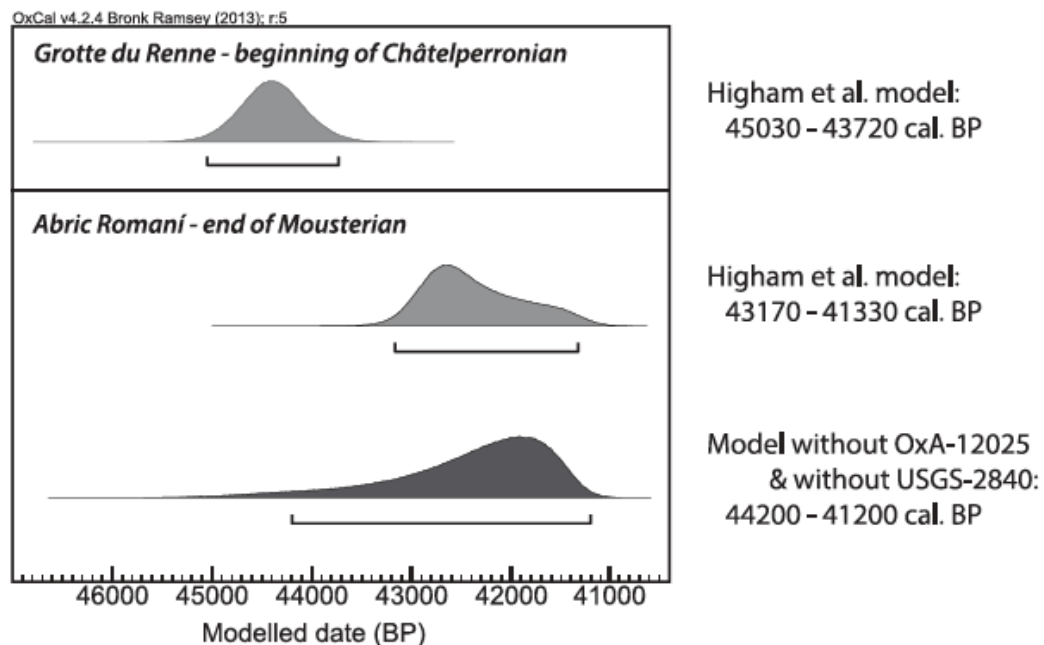


Figure 2: Comparison of the boundaries computed in the Discamps et al. (2015) paper for the end of the Mousterian at Abric Romaní, Spain. Discamps et al. (2015) claim that the second PDF, without two dates included, produces a range that is significantly older. It is not statistically significantly older, it simply has a longer tail than the Higham et al. boundary above it.

Are the two probability distributions shown above “significantly different”? Statistically different would imply that the two boundaries differ at 95.4% probability. Higham et al. (2014) calculated this by simply subtracting one PDF from another. The two are only considered different if the 95.4% probability range of the difference does not overlap with zero. In Figure 2, the lines underneath the two PDFs are the 95.4% probability ranges. A glance suggests a very obvious and wide degree of overlap that shows that the results cannot be different in a statistically meaningful manner. When we reran the Discamps et al. (2015) model, the difference overlaps with zero as expected. One cannot therefore say that one is significantly older than the other on statistical grounds.

Whether the longer tail makes a difference to our interpretation that the end of the Mousterian occurred after the start of the Châtelperronian is a slightly different question. We can repeat the process of finding the Difference between the Abric Romaní and Grotte du Renne Boundaries. In this case we find that once again the comparison does not result in a distribution that overlaps zero and therefore there is a significant difference between them.

An examination of all of the comparisons they undertake in their paper show that none are significantly different in any measureable way. There is no evidence for ‘drastically’ altered results. Discamps et al.’s (2015) arguments that some of our models are selectively, in their words, ‘dragging’ the results to the younger or older end of a spectrum and are significantly different cannot be demonstrated.

Aside from this key issue, the general approach followed by Discamps et al. (2015) is not one that can be recommended. At the Abric Romani site, for instance, they focus on two determinations in our previously published model (OxA-12025 and USGS-2840), which they think are responsible for ‘dragging’ the model boundary for the end of the Mousterian to make it too young. This seems to imply some kind of intent on our part. They suggest that OxA-12025 is ‘considerably younger’ than another date from this same layer (NZA-2312), implying that it might be this one that is problematic and hinting that we might be biasing our selection of samples. The reverse is the case in fact when the entire model is considered. The model suggests that it is the *older sample* from Layer B (NZA-2312) that is the outlier (96% outlying), based on its position relative to other dates; both U series and AMS dates, from above and below it. There is no reason to favour it over OxA-12025 in our view (which has a posterior outlier probability of 1%, suggesting it is not an outlier at all and should not require ‘removal’). Discamps et al. (2015) decide to remove these two results and see what the model result is like without them. This is contrary to conventional approaches, in which the data is modeled with all of the results included, and the outlier analysis or low agreement indices left to determine what is and what is not an outlier. In the original model, for example, USGS-2840 was included but because it had a 70% outlier probability it was only included in the model for 30% of the iterations. The new model, without these two supposedly problematic dates, shows no significant difference between them, as shown in Discamps et al.’s figure 7 (and our Figure 3 below). Their conclusion, however, is that “...it is sufficient to remove these two measurements (OxA-12025 and USGS-2840) to obtain **a significantly older boundary** for the end of the Mousterian at Abric Romani (44,200–41,200 cal. BP, Fig. 7). This example once again reveals the extreme sensitivity of Bayesian models to fixed priors when only a handful of dates are available per layer” (my emphasis). The probability that follows from this suggests that Arcy-sur-Cure’s Châtelperronian precedes Romaní falls from 0.99 under the Higham et al. (2014) model to 0.97 under theirs. Discamps et al. (2015) claim that this example shows extreme sensitivity. To us this shows the opposite, that the model is *insensitive* to the exclusion of the two results.

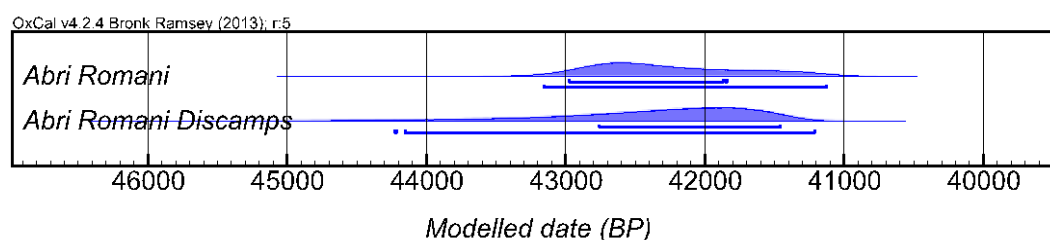


Figure 3: Boundaries for the end of the Mousterian at Abric Romani, the upper boundary is from Higham et al. (2014), the lower is the alternative without two AMS dates as outlined in Discamps et al. (2015). The distributions statistically overlap.

La Quina: Statistical significance and outlier analysis

We tested Discamps et al.'s (2015) modified La Quina model as well. Here, they claim that; *"the removal of a single age from Higham et al.'s Bayesian model radically changes the interpretation"*. They focus upon OxA-X-2362-22, for no apparent reason (it is not a significant outlier in the model of Higham et al. (2014)). Once removed, they state, the modeled age for the end of the Mousterian at La Quina no longer postdates the beginning of the Châtelperronian at Arcy-sur-Cure.

We reran the model following their removal of OxA-X-2362-22. The difference between the model boundaries for the end of the Mousterian under these two scenarios is shown below in Figure 4.

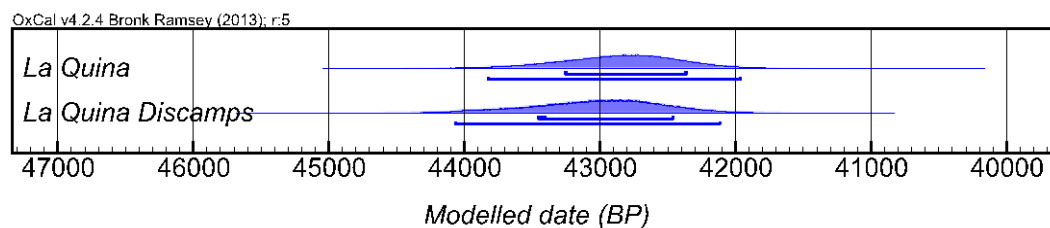


Figure 4: Boundaries for the end of the Mousterian at La Quina. La Quina is the boundary in Higham et al. (2014), La Quina Discamps is the boundary in the model in which is the difference is the removal of OxA-X-2362-22 from the model.

The distributions show, once again, there is no statistical difference between them. We compared the distributions with the PDF for the start of the start of the Châtelperronian at Arcy-sur-Cure and found that the probability for the La Quina boundary to be earlier is 0.0006 with OxA-X-2362-22 in and 0.0139 with it removed. The probability that Arcy is earlier than the end of the La Quina Mousterian is 0.9936 and 0.986 respectively for the two models. This demonstrates that the removal of OxA-X-2362-22 makes no significant difference (*contra* Discamps et al., 2015) and the likelihood that the La Quina Mousterian post-dates the Arcy Châtelperronian is high.

Geissenklösterle

Discamps et al. (2015) also critique the radiocarbon dating and modeling of the important Swabian site of Geissenklösterle (Germany). The excavator, Hahn (1988), initially identified seven archaeological layers in the Aurignacian levels of the site; IIIn, IIa, IIb, IIc, III, IIa and IIIb. Later, once refitting and other analyses

had been undertaken, Hahn realized that there were probably two main archaeological horizons, and these were termed AH II and AH III. These broad horizons were utilized by Higham et al. (2012; 2014) for the Bayesian modeling of the site. Discamps et al. (2015) suggest that there is “considerable post-depositional disturbance of the layers defined by Hahn”, based principally on refitting evidence that renders the modeling problematic and the boundaries in the model meaningless. The conclusion of extensive mixing is at odds with the work of Conard and Bolus (2008) who suggest the opposite; that there was *not* a significant degree of mixing. Their work was based on micromorphological studies undertaken by P. Goldberg, refitting studies and other taphonomic and contextual data. Discamps et al. (2015) acknowledge that there is spatial variation in the inferred degree to which material is not in its original context. In previous work, Teyssandier et al. (2006) presented an almost identical set of data to Discamps et al. (2015), including refitting data, and concluded that “our studies confirm the archeostratigraphic reconstruction by Hahn”. They concluded further that; “the scarcity of vertical mixing from AH II to AH III is confirmed by the stratigraphic position of characteristic organic objects such as split-based antler points or ivory figurines, which always lie within horizon II”. It seems confusing to us that one set of data can be interpreted in such diametrically opposite ways.

Higham et al. (2012) suggested that the transition to the Early Aurignacian AHIII spanned 43,060—41,480 BP (95.4%). Discamps et al. (2016) claim that this level cannot be as old as this and even if it were it must represent an “as yet unrecognized Protoaurignacian” industry (Discamps et al. 2015:14). Both Zilhão (2015) and Discamps et al. (2015) suggest that one should only look at the AMS dates of cutmarked bone and also restrict analysis to only prey animals, setting the other dates aside, in order to obtain the real chronological picture. This issue of the effect of these types of differences in prior and data upon the results of Bayesian modeling is essentially sensitivity testing. In this instance, we are asking how sensitive are the results to changes in the model?

We tested different models with various changes in priors and data to explore the effects on the posterior results compared with the original Higham et al. (2012, 2014) results. We ran three different versions to see what effects there would be on a key boundary in that model, the transition to AHIII. The first is that originally published by Higham et al. (2012; 2014). In this model it is important to note that dated non-cutmarked bones were given an outlier probability of 1.00 by Higham et al. (2012) to attempt to date the archaeological sequence more reliably, rather than include non-humanly modified results. The original rationale behind dating the non-modified samples from the site was simply to test whether or not the radiocarbon results published by Conard and Bolus (2003; 2008) were reproducible or not; they were not. Higham et al. (2012) concluded the likely reason was the lack of removal of contaminants in the bone collagen. The boundary for the Start Aurignacian for the Higham model is shown in Figure 5 (termed Start Aurig Higham Model). We also built a second model which was based on Zilhão (2015), in which only prey animals with cut-marks were included (this is shown in Figure 5 as Start Aurig Zilhão model). A final model, following Discamps et al.’s (2015) suggestions, consisted

of again only the prey animals with cut-marks, a total of 8 determinations according to them, which they say consist of 3 dates from IIIa and IIIb, 3 from IIb, IId and III and a date from layer 1c (so in reality a total of 7 determinations rather than the 8 they claim).

Both Zilhão's (2015) and Discamps et al. (2015) also criticized the phase ascriptions and stratigraphic sequence input into the model by Higham et al. (2012), so we also ran a model which simply made no stratigraphic phasing assumptions, but lumped all of the prey animal determinations into a single group of 'Aurignacian' dates. In this latter model the only constraints imposed are that the Upper Palaeolithic (Aurignacian) postdates the sterile layer IV and predates the start of the 1c level. All of the radiocarbon dates from AHIII and II are therefore considered to be unordered in this case. This is a very conservative model indeed. In Figure 5 this is termed 'model 4'.

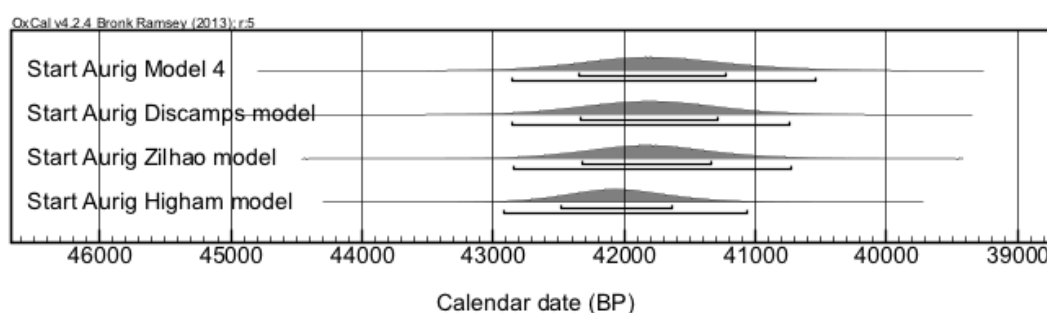


Figure 5: Comparisons for the Start Aurignacian boundary at Geissenklösterle when the priors and likelihood data is modified (see text and Supp. Methods for details).

The results presented in Figure 5 show that there is no significant difference in the estimated start boundary for the Aurignacian levels, no matter which of the four models is considered. This shows that the model is not sensitive to the choice of priors that were tested and reinforces the conclusion that the Aurignacian of Geissenklösterle begins at a very early date in the western European sequence. Discamps et al.'s (2015) suggest that; "although this initial occupation may in fact represent an Early Aurignacian without split-based points, it is nevertheless not as old as claimed by Higham et al. (2012)". Once again, an analysis of the results as shown above, shows the opposite is true.

Conclusion:

Discamps et al. (2015) suggest that the examples of Abric Romaní and La Quina show how the exclusion of dates can; "*drastically alter a Bayesian model*". Pettitt and Zilhão (2015) claim, in a review article, that it is obvious to see how a '*dragging*' effect occurs and how it produces '*serious errors*' in very sensitive models. They pronounce the Discamps et al. (2015) paper replete with '*potent*' examples showing this. As we have shown, the Discamps paper is only replete with examples showing the opposite. We have not able to reproduce their results

with actual data that have any meaningful statistical significance; the models do not produce different results. We conclude that their claims can be rejected.

Bayesian analysis is a continuing and ongoing iterative and statistical process. As further data is obtained or changes are made to the identification of lithic assemblages, models can be expanded or modified and the outcome assessed. In this vein, the Higham et al. (2014) paper represented the current best estimate for the timing of the various archaeological industries considered, and it also provided all of the information required to assess their interpretation. In some units, the number of dates was limited either due to problems with poor collagen preservation or the time constraints within a funded project, persistent challenges within archaeological practice today. It is therefore appropriate for the authors and those wishing to use their data to undertake sensitivity testing. However, when doing this it is important that the statistical significance of any changes and their implication for the overall interpretation is assessed. It is important that outlier analysis is carefully considered and undertaken statistically where reasons for identifying particular samples as outliers are based only on dates and not alternative data. Finally, to understand why differences may be obtained, only one change to the model should be made at any one time. If this is not undertaken appropriately, only confusion results.

Acknowledgements:

We are grateful to everyone working at the Oxford Radiocarbon Accelerator Unit (ORAU), University of Oxford. We are particularly grateful to Rachel Wood (ANU) and Katerina Douka (ORAU) for comments on the text.

All model codes are given in CQL in a Supplementary Online file.

References:

Bronk Ramsey, C. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* 51(1): 337-360.

Discamps, E., Gravina, B. and Teyssandier, N. 2015. In the eye of the beholder: contextual issues for Bayesian modelling at the Middle-to-Upper Palaeolithic transition. *World Archaeology* DOI: 10.1080/00438243.2015.1065759

Higham, T.F.G., Douka, K., Wood, R., Bronk Ramsey, C., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruíz, C., Bergman, C., Boitard, C., Boscato, P., Caparrós, M., Conard, N.J., Draily, C., Froment, A., Galván, B., Gambassini, P., Garcia-Moreno, A., Grimaldi, S., Haesaerts, P., Holt, B., Iriarte-Chiapusso, M-J., Jelinek, A., Jordá Pardo, J.F., Maíllo-Fernández, J-M., Marom, A., Maroto, J., Menéndez, M., Metz, L., Morin, E., Moroni, A., Negrino, F., Panagopoulou, E., Peresani, M., Pirson, S., de la Rasilla, M., Riel-Salvatore, J., Ronchitelli, A., Santamaria, D., Semal, P., Slimak, L., Soler, J., Soler, N., Villaluenga, A., Pinhasi, R. and Roger Jacobi. 2014. The timing and spatio-temporal patterning of Neanderthal disappearance. *Nature* 512, 306–309.

Millard, A.R. 2014. Conventions for reporting radiocarbon determinations. *Radiocarbon* 52: 555-559.

Pettitt, P.B and Zilhão, J. 2016. Problematizing Bayesian approaches to prehistoric chronology. *World Archaeology* 47(4): 525-42.

Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Grootes, P.M., Guilderson, T.P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T.J., Hoffmann, D.L., Hogg, A.G., Hughen, K.A., Kaiser, K.F., Kromer, B., Manning, S.W., Mu Niu, Reimer, R.W., Richards, D.A., Scott, E.M., Southon, J.R., Staff, R.A., Turney, C.S.M. and van der Plicht, J. 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon* 55(4): 1869-1887.