

## Response to the Referee

First we would like to thank the referee for their comments (and their quick turnaround in these busy times). They have helped us improve this manuscript, and made us really think about potential pitfalls. Major changes to the paper are shown in red. no changes have been made to the Supplementary Information.

First and foremost, we should highlight a major change to the presentation of our results. Following the referee’s comments, we now use  $Q_{\text{WMB}}$  to describe the mixture model weighting factor, where  $Q_{\text{WMB}} = 1 - P_s$ . This places the preference for weakened magnetic braking as the subject of our model analyses, making Figure 4 more readable, and hopefully our results more accessible. We have made sure to reverse any mentions of this parameter (e.g. ‘below  $P_s = 0.5$ ’ is now ‘above  $Q_{\text{WMB}} = 0.5$ ’).

### Major comments:

**1 - My main concern is that from the authors’ study one could only robustly conclude that Gyrochronology needs to be revisited or recalibrated, which we knew already. One could also conclude that WMB (or a mixture between the ‘standard’ and WMB models) offers a better description of the data than the ‘standard’ Gyrochronology model alone. However, comparing a single model to one that we know doesn’t work very well, is not enough to claim strong agreement of the data with WMB. From Fig 4, for example, it is clear that the cumulative probability  $P_s$  remains doesn’t change after  $P_s = 0.5$  for MS stars, which means that a mixture of ‘standard’ and WMB in equal proportions describes the data as well as the WMB alone. Surprisingly to me, the SG and Hot do show a steady increment all the way to  $P_s = 1$ , which doesn’t seem to support the idea that WMB describes better MS stars.**

The novelty of this work is the use of new asteroseismic rotation rates, which probe a parameter space for which the concept of weakened magnetic braking had not yet been tested. This conclusively shows that (some form of) weakened braking is a true physical phenomenon, not something that is just suggested by biases in previous measurement techniques. We have gone through the paper again to try and make the language used reflect this, and have changed our title from “strong agreement with” to “strong preference for”.

To address the referee’s other concerns in this paragraph: stars which have not undergone weakened magnetic braking yet will by design have equal preference for both models, as the models are identical until  $Ro = 1.97$ . This also explains why the Hot and SG stars show a steady increment in cumulative probability– they hold no diagnostic information in preference of one model over the other. We compare the stars to the models on an individual basis before multiplying the posteriors, and so stars that have not undergone braking just won’t have anything interesting to say. Conversely, any single star alone past the critical Rossby number will be able to differentiate between the standard or WMB models.

We would also like to clarify that in the mixture model, it will only be the cases with the most precise constraints on mass, age, rotation, temperature and metallicity that see an extremely strong result for  $Q_{\text{WMB}}$ . We are comparing two population distributions after all, which cover much of the same parameter space (with different densities). A star that prefers the WMB model will still be consistent with the standard model but with a much smaller probability. The mixture model is really showing which population the star is more likely to come from, not advocating for some model description that incorporates the physics of both.

**About the calibration of the rotation evolution model: the standard and WMB models are both calibrated on young clusters and to match the solar rotation at 4.6Gyr. The comparison of models is then one of “extrapolated solar braking” and “weaker than extrapolated solar braking”, and is not really as definitive as WMB vs standard. If**

there is some stochasticity in rotation evolution toward older, slower rotators, the Sun might present a biased calibration point. I would suggest the authors note that their study is strictly only supporting this conclusion that some stars have weaker than solar and extrapolated solar rotation braking (and only then if selection bias missing more active stars is not significant).

We have added the following to the first point in our conclusions: “Stars in our ensemble overall show a weaker rate of braking than one would expect from solar (and extrapolated solar) braking rates.”

We have no reason to suspect that seismic detection biases cause us to not measure active stars to the degree where it affects our results. Spot detection becomes extremely difficult for inactive stars, but asteroseismic detection biases are not the inverse of this. It is worth noting that over half our sample is active enough to have existing spot rotation measurements.

To respond to the risk of stochasticity in rotational evolution affecting our results: we don’t believe that there is enough support for this at late ages, especially in the space  $> 6$  Gyr where most of our sample resides. The rotation rates at 2.5 Gyr in Meibom et al. (2015) have an exceptionally low scatter, and there is no mechanism in the standard spin-down model to introduce scatter after that point.

**2 - Another big concern is that the critical Rossby number is fixed from previous studies in which the rotation periods came from star-spots observations. The authors admit those are biased and one of the motivations of this study is to test if that bias explains the observed deviation of post-MS stellar rotation periods from those predicted by Gyrochronology.**

The use of  $R_{\text{crit}} = 1.97$ , based on the van Saders et al. (2019) fit to the star-spot observations is chosen as the current best estimate of the critical Rossby number. We are not attempting to make any statements about the value of  $R_{\text{crit}}$  in this work. However with our method, if the true  $R_{\text{crit}}$  were to be much higher or lower than 1.97, this would water down our posterior probability for  $Q_{\text{WMB}}$  (either towards the standard model if higher, or towards no preference for either if lower). The fact that this  $R_{\text{crit}}$  is based on spot rotation rates, which are biased against inactive stars, is essentially unimportant for this comparison, as the stellar model populations we compare against are complete. However, that the WMB model is favoured when using independent asteroseismic rotation is important, of course.

**3 - I am also concerned about the selection bias:  
Methods, top p25**

*“More active, magnetic stars may also suppress their oscillation modes, making it more likely to detect seismic oscillations in older stars [45]. Through these two effects, asteroseismology is biased towards detecting older over younger stars. As it is older stars, undergoing magnetic braking, that are required to drive the distinction between stellar models, we do not expect this observing bias to effect our conclusions.”*

**This does not address the selection bias problem inherent in the study. Yes, the slower rotators are needed to find any WMB effects, but the synthetic populations do not include this selection bias and at face value will be lacking the more active stars that will favor the standard model, not WMB. The comparison is then inherently biased toward WMB.**

We think the referee is misinterpreting the asteroseismic selection effects. Spot rotation rates have a dearth of measurements for old, inactive stars. The seismic selection effect is not the opposite of this. Our sample includes both active and inactive stars (in fact, over half our sample has existing surface rotation measurements). The asteroseismic selection effect only excludes very young and very active stars ( $\lesssim 2$  Gyr, Salabert+16 and Mathur+19), which would not inform  $Q_{\text{WMB}}$  strongly.

For the purposes of this comparison, the asteroseismic selection is essentially complete. Page 25 is intended to inform the reader of asteroseismic selection biases for completeness, and we have now changed it to make sure it clarifies that this is not an issue for this study. The paragraph now reads as follows:

“Finally, we also consider bias in asteroseismic detections. The probability of an asteroseismic detection with *Kepler* scales with temperature and radius (i.e. scales inversely with  $\log g$ , and therefore main sequence age), and detections are unlikely below roughly 5200 K. Very young ( $\lesssim 2$  Gyr), very magnetically active stars may also suppress their oscillation modes. Though these two effects impose a bias on asteroseismic populations, they do so outside the parameter range on which our comparison takes place. As it is older stars, undergoing magnetic braking, that are required to drive the distinction between stellar models, we conclude that our seismic sample can be assumed to be complete for the purposes of this work.”

The referee also mentions that the synthetic populations are lacking more active stars, so we would like to reiterate that we have no reason to believe the stellar models are not complete. The standard and WMB model populations are built using the same synthetic stellar sample evolved under angular momentum evolution prescriptions which are identical until the critical Rossby number.

**The only proper way to compare the TRILEGAL synthetic populations is with stellar samples that are not detection biased in a different way. In this regard, it would also seem that a more complete sample could be constructed by including stars with spot-derived rotation period measurements to combine with the oscillations sample.**

The TRILEGAL sample is not detection biased, just calibrated to a particular critical Rossby number, which naturally includes the spot-derived rotation periods. The models describe what the Kepler field looks like in the cases with and without weakened braking at  $R_{\text{ocrit}} = 1.97$ . Our analysis simply tells us which of these two cases our stars are in better agreement with. There is no reason to believe these model populations are incomplete.

We should also note that any individual star will be able to tell the difference between the standard and WMB models (provided it is evolved past the point where the two models diverge). These individual star-by-star posteriors are combined to give us a single combined posterior that describes whether the ensemble as a whole prefers the WMB model, which drives our conclusions. We have tried to emphasise the fact that the model comparison is done star-by-star more in the text.

The unique selection biases that asteroseismology is subject to do not need to be reflected in the stellar models for the purposes of our analysis, for two reasons:

- On a star-by-star basis, the effects of a seismic selection effect are slowly changing (there are no sharp cut-offs except at the extremes of the parameter space, which are not relevant to our sample), so we would not expect a selection effect to alter our posterior probability.
- The selection effects would be applied to both models identically. Because we are comparing the two models in a mixture model, applying the selection effect would yield a net zero effect on the posterior probability, so there is no need to include it.

**This also raises the question of why not simply compute individual rotation models for all the parameters of the stars in the sample and compare directly synthetic to observed stars? At least as an additional test, it would seem the most obvious approach that should be much more amenable to simple statistical comparison.**

This is functionally what we have already done in our work. We have modeled rotation rates generated for each star in a simulation of the Kepler field. We then compare our observation to the distribution of these models in five parameter spaces simultaneously, and obtain a posterior distribution describing which angular momentum evolution prescription the observation more closely

agrees with.

## Minor comments:

**TRILEGAL rotation model simulations:** I would like to see a little bit more detail here on how the rotation models are computed. Eg. solid body rotation assumed? Definition of Rossby number in the code and how/where convective turnover time is computed? etc. While these might be described in detail in the primary references, the model ingredients are crucial for the study and it helps the reader to have a more complete summary so as not to have to hunt for this in another paper.

We’ve added an explanation of how the Rossby number and the convective turnover timescale are defined in the model, and mentioned our solid body rotation assumption.

**Page 2 - Asteroseismic main sequence targets:** I find this sentence not too clear: “A star’s rotation causes modes of oscillation to split into multiplets, enabling measurement of independent rotation rates by measuring the rate of splitting [22]” This will not be very helpful for anyone with no prior knowledge of asteroseismology, and will be obvious for anyone in that field.

We’ve changed this to: “Rotation causes certain modes of oscillation – which lie at the same frequency in a non-rotating case – to shift to higher or lower frequencies. The size of this ‘splitting’ provides a means of independently measuring the rotation of the star”. To accommodate the extra words in the introduction, we removed “This makes asteroseismic rotation rates a perfect test of weakened magnetic braking on the main sequence” from the following paragraph.

**Page 4 - Seismic versus surface rotation rates:** “How do the authors use the uncertainties to fit the line? “We fit a line of the form  $P_{\text{surf}} = m \times P_{\text{seis}}$ , using the larger of the asymmetrical uncertainties on asteroseismic rotation, and excluding stars above the 1.8:1 line to avoid likely multiples of the true rotation rate acting as outliers (transparent in the Figure).”

This was an oversight, and we’ve now included a more detailed explanation in Methods, in its own section, including an extra test of the  $P_{\text{surf}} = m \times P_{\text{seis}}$  fit using only stars that preferred the WMB model.

**Fig 3** would benefit from a more explicit explanation of the color code, which seems to be the number of stars predicted by each of the models. Also, the Caption overlaps with page number

Added the sentence “The z-axis colour scale indicates the density of stars in  $P$ - $T_{\text{eff}}$  space for the models.”, and reduced the image size.

$P_s$  is a mixing parameter and not a rotation period. I find the choice of name confusing in this context.

Good call. We’ve changed  $P_s$  to  $Q_{WMB}$  everywhere, which is more in line with other papers that use mixture models.  $Q_{WMB} = 1 - P_s$ , and we think this helps communicate the main conclusion more clearly (and improves the readability of Figure 4). While we’re at it, we made sure to use  $P$  instead of  $P_{\text{rot}}$  consistently for rotation.

**Page 6 - Results of Comparing Models: - “In order to assess the posterior probability for the full sample, we multiplied the individual posterior probabilities for Ps.” Do the authors mean “multiplied the individual posterior probabilities by Ps”?**

We think the referee may misunderstand what is being described here. We have probability distributions for  $Q_{\text{WMB}}$  ( $=1-\text{Ps}$ ) based on each individual star. By multiplying those distributions, we obtain a probability distribution that describes  $Q_{\text{WMB}}$  based on an ensemble of stars. We’ve added a sentence that clarifies this a little more here: “This joint posterior then describes the probability distribution of  $Q_{\text{WMB}}$  given multiple stars.”

**- “Both hot and sub-giant stars do not strongly prefer one model over the other (with both cumulative probabilities crossing the  $\text{Ps} = 0.5$  line at roughly 45%) [...] When performing this analysis on all 89 stars for which seismic rotation was measured (excluding the two stars with low metallicity), the total probability below  $\text{Ps} = 0.5$  is 96.6% I think it is not fair to mix populations with such different probabilities and then compute an overall probability and assume it represents anything.**

All stars provide meaningful information, and the multiplication of individual posterior probability distributions to create a joint posterior probability distribution is a valid way of studying the ensemble as a whole. We feel that, for completeness sake, it’s important to report the results given our full ensemble, even though subgiant and ‘hot’ stars in our sample don’t appear to have undergone weakened braking yet.

**Did the authors try a smaller Rossby number to see if the method works? If it also resulted in statistical agreement with a model of weakened magnetic braking at critical  $\text{Ro} = 1$  for example, one could argue there is an issue.**

We already know how different critical Rossby numbers affect the distribution of stars undergoing weakened magnetic braking (see van Saders et al. 2016). The further our models critical Rossby number deviates from the truth, the more watered down our distribution of  $Q_{\text{WMB}}$  will become, as stars will divert to favouring the standard model (in the case of a high  $\text{Rocrit}$ ) or the standard and WMB models will become essentially the same (low  $\text{Rocrit}$ ).

In this manuscript we are not making claims about statistical agreement between models and data, but relative statements about which models are preferred— we’re not trying to report on any value of  $\text{Ro}$  in particular.

**How do authors rule out that the issue is just a Gyrochronology calibration issue? Or two branches of rotators that, on average, favor a WMB model? That could be explored by separating the samples in faster vs slower rotators.**

We’re assuming that the referee is referring to the Barnes et al. model of gyrochronology that has a fast and slow rotating branch. Even if that model is assumed to be correct, it is applicable at young ages, and based on observations at 2.5 Gyr (Meibom et al. 2015) we do not expect this potential dispersion to hold any weight at the late ages ( $> 6$  Gyr) where most of our main sequence stars reside (i.e., any dispersion present would be far smaller than the uncertainties on rotation). Following the same line of thought, a fast and slow rotation branch at low ages would not average to form a distribution resembling the WMB model, as the WMB model has zero angular momentum loss past the critical Rossby number. Truncated braking may be a better name for the model, in retrospect.

We don’t think separating our sample by rotation rate is worthwhile, as we expect faster rotating stars to drive the distinction between the standard and WMB scenarios (rotating fast at late ages). For our distribution of  $Q_{\text{WMB}}$  to be due to two separate streams of rotational evolution the dispersion between these two streams would have to be very dramatic, which isn’t supported by

what we see in Meibom et al. (2015).

## Methods:

**Authors claim they use new measurements of rotation periods that do not suffer from the star-spots observational biased by which slower rotators will not be detected. However, they use a fixed critical Rossby number ( $Ro = 1.97$ ) that comes from van Saders et al. 1013, 1016, and 1019, and was derived from rotation periods measured from starspots, which do include that bias that is consistent with that Rossby number.**

This is a good observation, however we don't think these two things are mutually exclusive. The seismic observations are independent to surface rotation measurements, and are subject to entirely separate biases, as discussed in the text. The WMB model we compare to uses  $Ro = 1.97$  based on the van Saders et al. (2019) paper, as the reviewer mentions. The population generated from this model looks similar to the distribution of rotation rates seen in McQuillan et al. (2014), prompting the question whether it is a spot rotation bias causing this trend (as mentioned in the text and by the reviewer). By showing that our seismic rotation rates, which are not subject to the same biases as spot rotation, agree with the WMB model, we can conclude that this signature of weakened magnetic braking is not just a population bias effect.

Bear in mind also that we are comparing to a model, and not population measurements. The exact value of the Rossby number itself matters little, as a population drawn from any WMB model will look sufficiently different to the standard scenario past the critical Rossby number. We would also like to reiterate that we believe that both the model and observed populations are, for the purposes of this comparison, complete.

**What is the probability of  $P_s < 0.4$ , and  $P_s < 0.3$ ? This tells you more about WMB really being favored, rather than something in between.**

We've added the following: "The total posterior probability for all stars above 0.6 and 0.7 is 93.3% and 78.2% respectively. This does not necessarily mean that a mixture of the two models is preferred, as stars that have not yet undergone weakened braking, or which are near the transition period, will have a flat posterior distribution with no strong preference for either model."

**I am not sure how the authors reach the conclusion that their posterior estimates on rotational parameters are data and not prior-dominated. Just by looking at Fig 2 by eye?**

The analysis by-eye just reflects a comparison of the summary statistics of the posterior distribution vs the prior. We've changed the language so this is better reflected in the text:

"In the Figure, results with means (symbols) and 68% credible regions (error bars) that are close to those same values for the prior distribution (where the horizontal line is the mean, and the shaded area is the credible region) can be interpreted as prior-dominated (i.e. poorly informed by the data). Cases where the means differ or the credible regions are smaller than the prior distribution are data-dominated."