

# ‘On the red supergiant problem’: a rebuttal, and a consensus on the upper mass cut-off for II-P progenitors

Ben Davies<sup>1★</sup> and Emma R. Beasor<sup>2†</sup>

<sup>1</sup>*Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK*

<sup>2</sup>*NSF’s National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Ave., Tucson, AZ 85721, USA*

Accepted 2020 May 27. Received 2020 May 26; in original form 2020 April 2

## ABSTRACT

The ‘red supergiant (RSG) problem’ describes the claim that the brightest RSG progenitors to Type II-P supernovae (SNe) are significantly fainter than RSGs in the field. This mismatch has been interpreted by several authors as being a manifestation of the mass threshold for the production of black holes (BHs), such that stars with initial masses above a cut-off of  $M_{\text{hi}} = 17 M_{\odot}$  and below  $25 M_{\odot}$  will die as RSGs, but with no visible SN explosion as the BH is formed. However, we have previously cautioned that this cut-off is more likely to be higher and has large uncertainties ( $M_{\text{hi}} = 19^{+4}_{-2} M_{\odot}$ ), meaning that the statistical significance of the RSG problem is less than  $2\sigma$ . Recently, Kochanek has claimed that our work is statistically flawed, and with his analysis has argued that the upper mass cut-off is as low as  $M_{\text{hi}} = 15.7 \pm 0.8 M_{\odot}$ , giving the RSG problem a significance of  $>10\sigma$ . In this letter, we show that Kochanek’s low cut-off is caused by a statistical misinterpretation, and the associated fit to the progenitor mass spectrum can be ruled out at the 99.6 per cent confidence level. Once this problem is remedied, Kochanek’s best fit becomes  $M_{\text{hi}} = 19^{+4}_{-2} M_{\odot}$ , in excellent agreement with our work. Finally, we argue that, in the search for an RSG ‘vanishing’ as it collapses directly to a BH, any such survey would have to operate for decades before the absence of any such detection became statistically significant.

**Key words:** stars: evolution – stars: massive – supergiants; supernovae: general.

## 1 INTRODUCTION

Perhaps one of the greatest breakthroughs in massive star research of recent years has been the ability to directly associate supernovae (SNe) with their progenitor stars via archival pre-explosion imaging. Specifically, hydrogen-rich ‘plateau’ SNe (classified as II-P) have been unequivocally linked to red supergiant (RSG) progenitors (Smartt et al. 2004, 2009). Once the progenitor is identified, it is possible (by adopting a series of key assumptions) to estimate the star’s luminosity at death  $L_{\text{fin}}$ , and ultimately an initial mass  $M_{\text{init}}$ , providing a fundamental test of stellar evolutionary theory.

In the first attempt to analyse a sample of II-P progenitors, Smartt et al. (2009, hereafter S09) noted that the most luminous progenitor [SN1999ev with  $\log(L/L_{\odot}) = 5.1$ , but see introduction of Davies & Beasor 2020] was substantially fainter than the brightest RSGs in the field [a luminosity limit often referred to as the ‘Humphreys–Davidson (H–D) limit’, now established to be at  $\log(L/L_{\odot}) = 5.5$ ; Humphreys & Davidson 1979; Davies, Crowther & Beasor 2018; Davies & Beasor 2020]. The sample consisted of nine detections,

and 12 upper limits, and the authors quoted the significance of this discrepancy to be  $2.4\sigma$ , though fig. 6 in S09 suggests rather it is somewhat below  $2\sigma$  ( $<90$  per cent). This possible tension between the II-P luminosity distribution and that of field RSGs was termed the ‘RSG problem’.

Despite the statistical significance of the RSG problem being within  $3\sigma$ , several explanations for its existence have subsequently appeared in the literature. A popular hypothesis has been that the ‘missing’ RSGs [i.e. those that die with luminosities between  $5.1 < \log(L/L_{\odot}) < 5.5$ ] collapse to form black holes (BHs) with no observable SN, which resonates somewhat with the results of independent numerical work (O’Connor & Ott 2011; Horiuchi et al. 2014; Ertl et al. 2016; Müller et al. 2016; Sukhbold, Woosley & Heger 2018). Converting these terminal luminosities, as well as the luminosity of the H–D limit, into initial masses using e.g. the STARS evolution code adopted by S09, this suggests that stars with initial masses between  $17$  and  $25 M_{\odot}$  will still evolve to become RSGs, but rather than exploding as II-P SNe will simply vanish with no explosion. This has prompted searches for ‘disappearing’ RSGs in archival survey data of nearby galaxies (e.g. Kochanek et al. 2008), but as yet no convincing example has been found (Reynolds, Fraser & Gilmore 2015; Adams et al. 2017).

\* E-mail: B.Davies@ljmu.ac.uk

† Hubble Fellow

A more mundane explanation for the RSG problem is that, with its low statistical significance, it is possible that no problem exists at all. In Davies & Beasor (2018, hereafter DB18), we reanalysed the mass distribution of II-P progenitors, with improved measurements of foreground extinction from Maund (2017) and more realistic bolometric corrections. We adopted a Monte Carlo (MC) style analysis method to determine the posterior probabilities on the lower and upper mass limits ( $M_{lo}$ ,  $M_{hi}$ ) to the distribution when both were allowed to be free parameters. Despite a larger sample, due to more nearby II-P SNe in the intervening years since S09, we argued that the significance of the RSG problem was still less than  $2\sigma$ . In Davies & Beasor (2020, hereafter DB20), we studied the luminosity distribution rather than the (model-dependent) mass distribution, removed the assumption of a Salpeter initial mass function (IMF), and looked more closely at the expectation value (i.e. the observations of the H–D limit). Again, we found a significance below  $2\sigma$ . Nominally, DB20 found an upper mass limit  $M_{hi} = 18_{-2}^{+4} M_{\odot}$ , again assuming the  $M_{init}$ – $L_{fin}$  relation from the STARS models used by S09, and by comparing different evolutionary models estimated a further systematic error on  $M_{hi}$  of  $\pm 1.3 M_{\odot}$ .

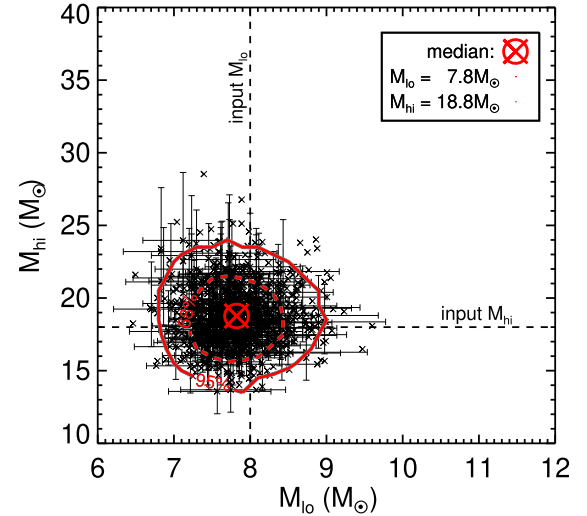
Recently, the analysis in DB18 has been challenged by Kochanek (2020, hereafter K20). In that paper, K20 adopt a Bayesian analysis method, initially finding results that are consistent with DB18 and DB20, specifically  $M_{hi} = 19_{-2}^{+4} M_{\odot}$  (STARS  $M_{init}$ – $L_{fin}$  relation). However, in performing a series of MC simulations with mock data, K20 found an apparent correlation between the upper error bar on  $M_{hi}$ , in this case  $+4 M_{\odot}$ , and the fitted value of  $M_{hi}$  (see fig. 5 in K20). This correlation was interpreted by K20 as being evidence that individual MC simulations, randomly scattered about an input value by the observational errors, may be corrected back to that input value. K20 proceeded to use this correlation to adjust their best-fitting value of  $M_{hi}$  obtained from analysis of the real data, arriving at  $M_{hi} = 15.8 \pm 0.8 M_{\odot}$ . This then gives a statistical significance to the RSG problem of  $> 10\sigma$ .

In Section 2, we will first rebut the conclusions of K20. We will show that the correlation observed between  $M_{hi}$  and its upper error bar is explained by error propagation when fitting a steep power law to data with non-zero errors. Furthermore, we will show that the low  $M_{hi}$  quoted by K20 is refuted by a comparison to the data used to derive it. Finally, in Section 3 we will assess the prospects for observing RSGs spontaneously collapse to BHs.

## 2 A REBUTTAL TO KOCHANEK (2020)

In section 3 of K20, an MC experiment is performed to demonstrate the accuracy and precision of the analysis method. The experiment involves randomly generating a sample of 24 masses from a power-law distribution characterized by a Salpeter slope ( $x = 1.35$ ), and input upper and lower mass limits  $M_{lo,in} = 8 M_{\odot}$  and  $M_{hi,in} = 18 M_{\odot}$ . These mock progenitor masses are randomly allocated to real progenitor sites to determine what their pre-explosion photometry would be, such that the posterior probability distributions on their inferred masses could be calculated. K20's Bayesian analysis is performed on each randomly generated sample to obtain a two-parameter fit ( $M_{hi}$  and  $M_{lo}$  are allowed to vary,  $x$  is fixed) to that trial's progenitor mass distribution. The results of 500 such trials are plotted in fig. 3 of K20, demonstrating a small systematic bias of the method (i.e. the offset of the cloud of points from the input values), as well as the random errors on  $M_{hi}$  and  $M_{lo}$  (i.e. the distribution of points about the median output values).

In Fig. 1, we plot the results of a similar analysis of our own. Following K20, we generate 1000 samples of 24 randomly



**Figure 1.** Results of our MC experiment, analogous to fig. 3 of K20, showing the posterior distributions of  $M_{lo}$  and  $M_{hi}$ . As in K20, error bars are shown for only 10 per cent of the points for the sake of clarity. The red cross shows the median output  $M_{lo}$  and  $M_{hi}$ , both of which have a small systematic bias as in K20. The dashed and solid lines show the 68 and 95 per cent confidence limits, respectively, analogous to  $1\sigma$  and  $2\sigma$  random experimental error bars.

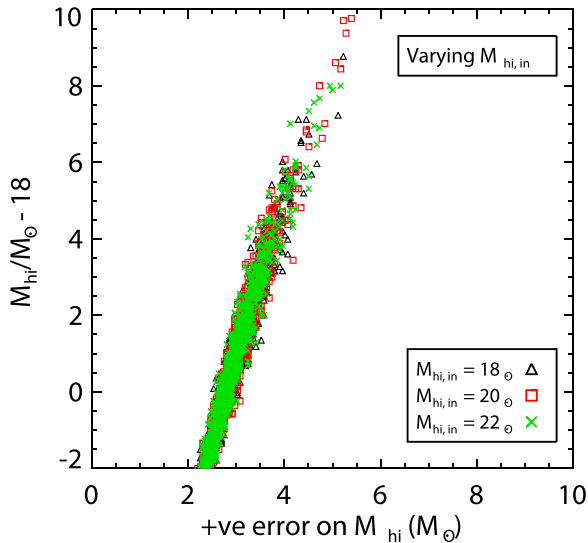
generated masses between  $M_{lo,in} = 8 M_{\odot}$  and  $M_{lo,in} = 18 M_{\odot}$  according to a power-law distribution with  $x = -1.35$ . To simulate experimental errors, we assume uniform fractional errors  $\sigma_M$  on all masses, initially fixed at 20 per cent, but the effect of varying which is studied later. For each mass  $M_i$ , we randomly sample from the normal distribution centred on  $\log(M_i) \pm \log(\sigma_M)$ . We then fit these masses with the function

$$M_i^{-x} = (M_{lo}^{-x} - f_i M_{lo}^{-x} + f_i M_{hi}^{-x}), \quad (1)$$

where  $x = 1.35$  is again the Salpeter slope of the IMF and  $f_i$  is the normalized ranking of the  $i$ th supernova out of the sample of 24, ordered in increasing mass. In each trial, we compare the simulated mass spectrum  $M(i)$  with those generated from equation (1) across a grid of  $M_{lo}$  and  $M_{hi}$ , determining the quantity  $\chi^2$  at each point in the grid. The best-fitting values of both  $M_{lo}$  and  $M_{hi}$  are determined from the location in the grid of the  $\chi^2$  minimum ( $\chi_{min}^2$ ), and the 68 per cent confidence limits on each parameter from the region of the parameter space defined by  $\chi^2 = \chi_{min}^2 + 2.3$  (following Avni 1976). We then repeat the experiment for each of the 1000 MC trials. As in K20, our results (see Fig. 1) show a similar cloud of points, centred close to the input values  $M_{lo,in}$  and  $M_{hi,in}$  but with a small systematic offset, and with points distributed about the median indicative of the random errors on  $M_{hi}$  and  $M_{lo}$ .

### 2.1 The misinterpretation of the mass-error correlation

The next step taken by K20 is to take the upper error bars on each MC trial's measured  $M_{hi}$  (i.e. the upper error bars on the data points in their fig. 3, or our Fig. 1), and plot them as a function of  $(M_{hi} - M_{hi,in})$  (shown in their fig. 5). K20 name this latter quantity 'overestimate of  $M_{hi}$ ', but which is equivalent to ' $M_{hi}/M_{\odot} - 18$ ', i.e. the difference between the output and input values of  $M_{hi}$ . K20 observed that these two quantities are correlated, and interpreted this correlation as each randomly scattered data point being aware of how far it is from the centroid of the input distribution, and



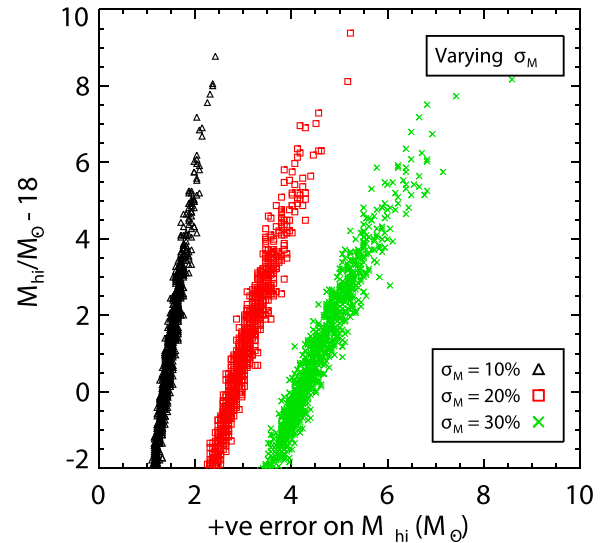
**Figure 2.** Our version of fig. 5 in K20, showing the quantity ( $M_{\text{hi}} - 18 M_{\odot}$ ) for each of our MC trials, as a function of the positive error bar on  $M_{\text{hi}}$  for that trial. In addition to showing the results for the same input value as K20 ( $M_{\text{hi,in}} = 18 M_{\odot}$ ), we also show the results for  $M_{\text{hi,in}} = 20$  and  $22 M_{\odot}$ . The exact same trend is seen irrespective of  $M_{\text{hi,in}}$ , demonstrating that the correlation seen in the plot does not depend on the input value of  $M_{\text{hi}}$ .

that this information is communicated back to the observer through the data point’s error bar. Continuing with this line of reasoning, K20 then claim that it is possible to transpose the results of each individual MC trial back to the input value of  $M_{\text{hi,in}}$  with a precision of  $\pm 0.8 M_{\odot}$ . Finally, K20 then treats the analysis of the real-world sample of 24 progenitor mass estimates as though it were a single MC trial, and concludes that the best fit of  $M_{\text{hi}} = 19_{-2}^{+4} M_{\odot}$  should be adjusted down to  $M_{\text{hi}} = 15.8 \pm 0.8 M_{\odot}$ . This reduction of  $M_{\text{hi}}$ , as well as its upper error bar, implies that the difference between  $M_{\text{hi}}$  and the mass associated with the H–D limit is now in excess of  $10\sigma$ , leading K20 to state that ‘the RSG problem remains’.

In Figs 2 and 3, we use the results of our MC experiment to make similar figures to fig. 5 in K20, again seeing a correlation. However, in Fig. 2, we make the same plot but for three different values of  $M_{\text{hi,in}}$ . Were the individual data points aware of the true input value of  $M_{\text{hi,in}}$ , the three sets of data would be offset in the y-direction, since a different correction factor would be required for each different value of  $M_{\text{hi,in}}$ . However, the exact same trend is seen *irrespective of*  $M_{\text{hi,in}}$ . Hence, we can say that the trend seen in fig. 5 in K20 has no dependence on  $M_{\text{hi,in}}$ .

Next, in Fig. 3, we again recreate fig. 5 of K20, but for MC experiments with fixed  $M_{\text{hi,in}} = 18 M_{\odot}$  and three different values of  $\sigma_M$  (10, 20, and 30 per cent), which are typical of those in the current sample of progenitor masses (see DB18). In each case, we see a correlation, but this time the slopes and offsets of these trends are dependent on  $\sigma_M$ . Our results show that the trend observed in fig. 5 of K20, rather than containing information on  $M_{\text{hi,in}}$ , is actually an illustration of how the random experimental errors on the progenitor masses propagate through to the error on  $M_{\text{hi}}$ .

We therefore conclude that K20 has misinterpreted the correlation seen in their fig. 5. The scatter of the data points in the MC trials about  $M_{\text{lo,in}}$  and  $M_{\text{hi,in}}$  (fig. 3 in K20 and our Fig. 1) illustrates the random experimental errors, and is driven by the uncertainties on the individual progenitor mass estimates. The errors on each progenitor mass are propagated from those on the host galaxy distance, the



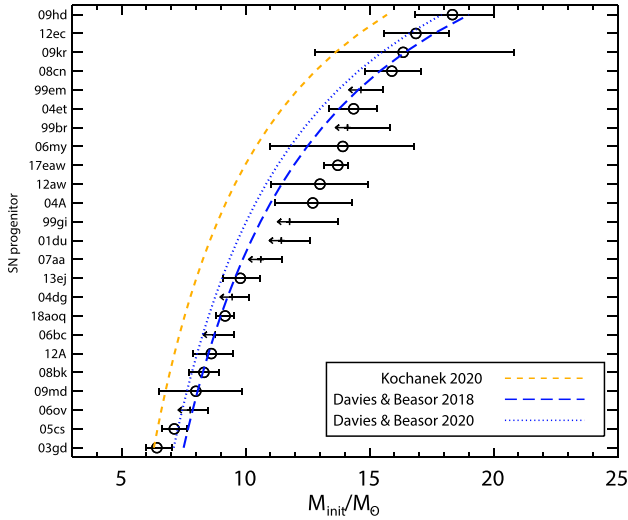
**Figure 3.** Same as our Fig. 2, but for fixed  $M_{\text{hi,in}} = 18 M_{\odot}$ , and with three different sizes of error bars on the input progenitor masses ( $\sigma_M = 10, 20,$  and  $30$  per cent). Different trends are seen for each value of  $\sigma_M$ , demonstrating that the correlation seen in fig. 5 of K20 is in fact caused by simple error propagation.

foreground reddening, and the bolometric correction, and represent the limits of our capability to measure these quantities. These cannot be corrected for. Instead, we assert that the trend seen in K20’s fig. 5, rather than being a means to correct for random errors, is in fact caused by the propagation of the observational errors into the uncertainty on the inferred value of  $M_{\text{hi}}$ . Using this trend to adjust the best-fitting  $M_{\text{hi}}$  would be erroneous, and would result in values of  $M_{\text{hi}}$  that were systematically low.

## 2.2 The cumulative mass distribution

Further evidence that the upper mass cut-off quoted by K20 is not supported by the data is found by a comparison to the numbers being fitted in that paper. In their analysis, K20 took the same photometric data, galaxy distances, reddenings, and bolometric corrections as in DB18, in order to prove that there were issues with the latter paper’s analysis. Hence, the masses of the individual progenitors being analysed are the same in both DB18 and K20. We can therefore assess the quality of the fits in both papers by overplotting the implied cumulative mass spectrum using the measured  $M_{\text{lo}}$  and  $M_{\text{hi}}$  and equation (1).

The fits of DB18, the more sophisticated analysis of DB20 (when converted from the luminosity plane to the mass plane), and that of K20 are overplotted on the progenitor mass distribution from DB18 in Fig. 4. It is clear from this figure that the K20 mass limits are inconsistent with the data. The predicted mass distribution is systematically offset to low masses for all progenitors, and in all but three progenitors this offset is greater than the quoted 68 per cent uncertainties. To quantify the quality of the K20 fit, we approximate the probability distributions on each detection as being asymmetric Gaussians, and determine the root mean square of the quantity  $z = (M_i - M_{\text{K20}})/\sigma_i$ , i.e. the average deviation from the data in units of  $\sigma_i$ . We then integrate a normal distribution between  $-\infty$  and  $z$ . We find that the K20 estimate of  $M_{\text{hi}}$  can be ruled out at the 99.6 per cent confidence level.



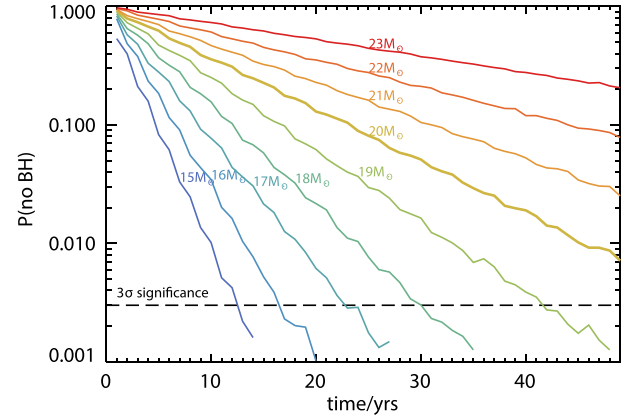
**Figure 4.** The cumulative progenitor mass distribution analysed by both **DB18** and **K20**. The blue dashed line shows the best fit of **DB18**; the orange short-dashed line shows the prediction if  $M_{\text{hi}} = 15.8 M_{\odot}$  by **K20** is adopted. For completeness, we show the fit of **DB20** converted to the mass plane.

### 3 THE RSG–BH PRODUCTION RATE

One of the most exciting prospects offered by a statistically significant RSG problem is the possibility that we could see an RSG disappear and form a BH in real time. Several research teams are currently monitoring nearby star-forming galaxies in the hope of catching such an event. To date, only two such candidates exist, neither conclusive (Reynolds et al. 2015; Adams et al. 2017).

As time goes on, and no vanishing RSGs are confirmed, it is worth asking the question at what point does the *lack* of a detection become interesting in its own right? And, can this lack of any detection be used to make an independent estimate of  $M_{\text{hi}}$ ? To answer these questions, we perform a series of simple MC experiments. First, we take the observed II–P rate in the nearby Universe<sup>1</sup> as being the number of events in the **DB20** sample divided by the time since the first event. This yields a II–P rate of  $1.3 \text{ yr}^{-1}$ . Next, we assume that stars may die as RSGs with initial masses up to  $25 M_{\odot}$  (see earlier), but that only stars with masses between  $7 M_{\odot}$  and  $M_{\text{hi}}$  will produce II–P SNe. Under the assumption of a Salpeter IMF, we can then estimate the RSG-to-BH rate as a function of  $M_{\text{hi}}$ . Next, randomly sampling from a Salpeter power law between  $M_{\text{hi}}$  and  $25 M_{\odot}$  at the appropriate BH formation rate, we run a series of MC experiments to determine the most likely number of RSG-to-BH events observed over a given time window. From the number of MC trials in which no RSG-to-BH event was observed, we can determine the probability that we would observe *no* such events as a function of observing time and of  $M_{\text{hi}}$ .

The results of this test are plotted in Fig. 5.<sup>2</sup> Specifically, we plot the probability of finding zero RSG-to-BH events within a given time period for several different values of  $M_{\text{hi}}$ . Also indicated in



**Figure 5.** The probability of observing no RSG collapse directly to a BH as a function of observation time, for a range of upper mass cut-offs  $M_{\text{hi}}$ . The 99.7 per cent confidence limit, analogous to  $3\sigma$ , is indicated by the dashed line.

the plot is the 0.3 per cent probability threshold, below which the significance of no observed RSG-to-BH event is greater than  $3\sigma$ . For a survey that had been running for  $\sim 12 \text{ yr}$  (e.g. Kochanek et al. 2008), we see that the lack of any detection argues against a value of  $M_{\text{hi}}$  below  $15 M_{\odot}$ . Within 5 yr, **K20**’s estimate of  $15.8 M_{\odot}$  could also be excluded. However, to provide a stringent independent test of the  $M_{\text{hi}}$  inferred from archival pre-explosion imaging on a time-scale shorter than approximately decades, the search volume would have to be dramatically increased, with e.g. the *James Webb Space Telescope*.

### 4 CONCLUSION – A CONSENSUS ON $M_{\text{HI}}$

In this letter, we have argued that the conclusion of Kochanek (2020) that ‘the RSG problem remains’ is invalidated by a statistical misinterpretation in that paper. Specifically, a correlation between the best-fitting upper mass cut-off  $M_{\text{hi}}$  and its error bar observed in MC tests was misconstrued as a means to correct for random experimental errors. Without this correction, Kochanek’s analysis of the II–P progenitor mass distribution finds  $M_{\text{hi}} = 19_{-2}^{+4} M_{\odot}$  (using the STARS  $M_{\text{init}}-L_{\text{fin}}$  relation, MLR) in excellent agreement with that of Davies & Beasor (2020) ( $M_{\text{hi}} = 18_{-2}^{+4} M_{\odot}$ , STARS MLR), not just in terms of the best-fitting value but also in terms of the probability distribution. With the same result being obtained seemingly independent of analysis strategy, we take this as evidence that the field is reaching consensus as to (i) the most likely value of  $M_{\text{hi}}$  and (ii) the precision to which this value can be quoted given the available data. Given this state of affairs, we can only reiterate our conclusions from Davies & Beasor (2020) that any mismatch between the luminosity distributions of field RSGs and those of II–P progenitors cannot be established beyond the  $3\sigma$  level without at least a doubling of the sample size of progenitors.

### ACKNOWLEDGEMENTS

We thank the two anonymous referees and the editor for comments and suggestions that helped us improve our paper. We also thank Nate Bastian and Phil James for comments on the draft, and Chris Kochanek for ongoing discussions. This work has made use of the IDL Astronomy Library, available at <https://idlastro.gsfc.nasa.gov>. EB was supported by NASA through Hubble Fellowship grant

<sup>1</sup>Here, ‘nearby Universe’ means close enough to be able to resolve and detect individual RSGs down to the mass threshold for BH production, which at the present time is  $\sim 30 \text{ Mpc}$ .

<sup>2</sup>The plot may be transformed from the initial mass ( $M_{\text{init}}$ ) plane to the terminal luminosity ( $L_{\text{fin}}$ ) plane by applying the calibration  $\log(L_{\text{fin}}/L_{\odot}) = A + B \log(M_{\text{init}}/M_{\odot})$ , where the constants ( $A, B$ ) are (2.67, 2.02) and (2.92, 1.82) for the STARS and rotating Geneva models, respectively.

*HST*-HF2-51428 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555.

## REFERENCES

- Adams S. M., Kochanek C. S., Gerke J. R., Stanek K. Z., Dai X., 2017, *MNRAS*, 468, 4968
- Avni Y., 1976, *ApJ*, 210, 642
- Davies B., Beasor E. R., 2018, *MNRAS*, 474, 2116 (DB18)
- Davies B., Beasor E. R., 2020, *MNRAS*, 493, 468 (DB20)
- Davies B., Crowther P. A., Beasor E. R., 2018, *MNRAS*, 478, 3138
- Ertl T., Janka H. T., Woosley S. E., Sukhbold T., Ugliano M., 2016, *ApJ*, 818, 124
- Horiuchi S., Nakamura K., Takiwaki T., Kotake K., Tanaka M., 2014, *MNRAS*, 445, L99
- Humphreys R. M., Davidson K., 1979, *ApJ*, 232, 409
- Kochanek C. S., 2020, *MNRAS*, 493, 4945(K20)
- Kochanek C. S., Beacom J. F., Kistler M. D., Prieto J. L., Stanek K. Z., Thompson T. A., Yüksel H., 2008, *ApJ*, 684, 1336
- Maund J. R., 2017, *MNRAS*, 469, 2202
- Müller B., Heger A., Liptai D., Cameron J. B., 2016, *MNRAS*, 460, 742
- O'Connor E., Ott C. D., 2011, *ApJ*, 730, 70
- Reynolds T. M., Fraser M., Gilmore G., 2015, *MNRAS*, 453, 2885
- Smartt S. J., Maund J. R., Hendry M. A., Tout C. A., Gilmore G. F., Mattila S., Benn C. R., 2004, *Science*, 303, 499
- Smartt S. J., Eldridge J. J., Crockett R. M., Maund J. R., 2009, *MNRAS*, 395, 1409 (S09)
- Sukhbold T., Woosley S. E., Heger A., 2018, *ApJ*, 860, 93

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.