We thank anonymous Referee #1 for his/her constructive criticism and valuable comments. In the following we address the points brought up, with referee comments in boldface and author responses in normal typeface.

The main comment I have is that whereas some limitations of lab experiments are acknowledged (age of cultures, lack of evolution), other major ones are not, and need to be. In particular, the experience with OA impacts on N2-fixation suggests that meta-analysis is not always the best way of elucidating the right answer. The large majority of experiments conducted have found that OA stimulates an increase in N2-fixation. However, they were all carried out at elevated iron concentrations compared to the open ocean. The study of Shi et al (2012), at more realistic iron, found the opposite effect. This should ring alarm bells for meta-analysis studies because even if an infinite number of lab studies were to be carried out then the overriding majority conclusion could still be completely wrong if they were mostly carried out under unrealistic environmental conditions, and if that difference alters the response obtained. This doesn't mean that the results of this study are not valuable (no single approach to studying OA is perfect), but it is helpful in the interpretation to consider the possibility of this sort of error. Another potential source of error is that lab experiments are carried out on monocultures isolated from the ecosystem with which they normally closely interact. However, this does not hold for the mesocosm experiments.
We agree with the referee. We will extent the paragraph about limitations of the single lab-experiments used in our meta-analysis and highlight the points mentioned above.
One should note, however, that the field of carbonate chemistry research has evolved over more than a decade and that many of the experiments used in our meta-analysis were conducted after standardized procedures, following the "Guide to Best Practices for Ocean Acidification Research and Data Reporting" (Riebesell et al. 2010).

14858/23: the other limitations should also be acknowledged.
The sentence will be changed to: "As the data sets used in this meta-analysis do not account for adaptive responses, ecological fitness and ecosystem interactions, the questions remains how these physiological responses play out in the natural environment."
In order to keep the abstract short and straightforward, we will address limitations of the single carbonate chemistry experiments in the discussion section.

14859/13: the impacts are not likely to be large on a centennial scale, see for instance the work of Christoph Heinze. The word "Thus" is any case not justified, because showing that there is an effect is very different from showing that the effect is significant.
The sentence will be rephrased.

14859-14860: it would be helpful if at this point it could also be explained how this paper differs from previous studies by Findlay et al and Ridgwell et al.
We will also refer to differences between meta-analyses by Findlay et al. and Ridgwell et al.

14860/15: how many studies?
The number of studies included in our meta-analysis is specified in the result section.

14862/15: how many experiments were excluded on this basis?
The number of experiments excluded on this basis will be added.

14863-14864: the precise equation/method for allocating weights should be provided, and the weights listed as an extra column in table 1.
We agree and will rewrite the paragraph as follows: "Using the variance v_i and the mean of the response ratio L_i for each experiment *i*, Cochran's Q (Cochran, 1954) was computed. With the help of Q an estimate of the between experiment variance (σ^2_λ) was obtained (Hedges et al., 1999). The weighted mean of the log response ratio $\overline{L^*}$ is given by:

$$\overline{L^*} = \frac{\sum_{i=1}^{k} w_i^* L_i}{\sum_{i=1}^{k} w_i^*} \qquad (2)$$

where *k* is the number of studies and $w_i = 1/(v_i + \sigma^2_\lambda)$.
Subsequently, the standard error of the weighted mean was estimated (see Eq. 7 in Hedges et al., 1999) and the confidence intervals were calculated.
Including weights in table 1 would seriously impair the clarity of the chart, as all the studies needed to be separated by experiment and CO_2 manipulation level. As the table is supposed to focus on the overall responses of coccolithophores within the studies, we would like to keep it as lucid as possible.

table 1: the criterion for distinguishing "some response" from "no response" should be described in the caption or the main text.
To clarify this, we will change the caption of the table to: "Summary of the available carbonate chemistry manipulation experiments and the responses of *Emiliania huxleyi* as reported by the authors of those studies. Symbols indicate: – no significant response, ／ increased response, ⌒ non-linear response and ＼ decreased response."

14870/2: "Another proposed explanation for the high difference in variance between..."
This will be changed as suggested.

14871/3: "2009), overall there is nevertheless a generally negative..." (these results do not say anything about how large the strain-specific variations are)
This sentence will be rephrased: "Although some strains of *E. huxleyi* appear to be less sensitive to ocean acidification (Langer et al., 2009), the species shows a negative response towards reduced pCO_2 levels in our meta-analysis, suggesting that strain-specific variations are small compared to the generally negative effect of ocean acidification on this species."

14871/11: "lead to a reduction in" rather than "minimize"
Will be changed to "decrease the confidence interval"

14871/21: although it should be noted that this effect is not observed in E. huxleyi, which has been most intensively studied.
We agree with the referee. For this reason, we already draw a distinction by saying: "[…] – at least for *Gephyrocapsa oceanica* – [...]"

14872/5: an increase in calcification rate at high CO2 does not necessarily mean that the species is benefitting from the high CO2 (resources can be reallocated, e.g. at the expense of reproduction rate).
This will be clarified.

14872/21-25: I think there are insufficient data (N=3) to conclude that the most prevalent species are the ones most affected.
We agree with this comment and will accommodate this in the revised version.

14874/17-19: but this would imply that they calcify to no purpose, which is hardly likely.
We will revise this section.

table 1: add extra colum for weightings. spelling mistakes in specifics column. "- no response" in caption.
The mistakes will be corrected. Concerning the weightings please see the reply above.

figs 1-3: remind readers in the captions that these are responses relative to 280 ppmv.
This will be changed.

References:
Riebesell U., Fabry V. J., Hansson L. & Gattuso J.-P. (Eds.), 2010. Guide to best practices for ocean acidification research and data reporting, 260 p. Luxembourg: Publications Office of the European Union.