Interactive comment on **"Carbon cycling and phytoplankton responses within highly-replicated shipboard carbonate chemistry manipulation experiments conducted around Northwest European Shelf Seas"** ***by*** **S. Richier et al.**

We are grateful for the comments of the two anonymous reviewers. In the following, we address each of these comments in turn, and outline the specific changes and clarifications that we have implemented. Reviewer's comments are in italics, while our responses are in normal typeface.

Generally, in relation to many of the points raised by both reviewers, we maintain that progress in investigating and understanding the sensitivities of organismal physiology to variability in carbonate chemistry over a wide range of timescales is of fundamental scientific value; ultimately it may also prove to be a productive route by which to address the wider questions related to changes (both natural and anthropogenic) which occur within the natural system. It is in this context that the study was conceived and presented. We further argue that no individual study, including ours, should be considered to represent a robust demonstration of the changes in marine microbial communities, which will occur as a consequence of anthropogenic ocean acidification (AOA).

Anonymous Referee #2

Received and published: 6 May 2014

This study examines the impact of ocean acidification on phytoplankton and net carbon production while taking spatial variability into account. In general, the manuscript is well written and focuses on an important topic. Until now ocean acidification research has often focused on single species or clones, thus, the here presented community approach is timely. However, the weak points in this work are mainly the short experimental duration and that the present experimental design does not allow disentangling if differences in response among the experiments are owing to environmental conditions, community composition or their interaction. It is also not possible to draw any definite conclusions about net production, since the carrying capacities of the different communities remain unknown. Moreover, the title claims that the authors conducted thorough carbon budgeting, which is apparently not the case. Some suggestions, questions, corrections, etc. are made in the present comment and might be useful for the authors to revise their manuscript.

We have considered the reviewer comments and below provide some justifications/clarification and will revise the manuscript accordingly.

General comments and suggestions:

1. *The authors traded off short experimental duration against several experiments at different geographical locations. They emphasise that a unique feature of their study is the inclusion of communities as a whole instead of just single species.*

We note that we did not state that working on natural communities was a unique feature of our study. Indeed we cited numerous prior examples of experimentation

over a range of timescales on natural communities (see Table 1 and numerous references in text).

However, as correctly stated by the authors, the only response that can be measured after the here presented maximum four days is physiological by nature. The short experimental duration does not allow responses to occur on the community (or genotype) level, i.e. in terms of species or genotypic sorting; yet this would exactly be the interesting point in a community approach.

We can only partly agree with the reviewer here and suggest that the distinct processes of physiological acclimation, within community selection and adaptation/evolution need to be clearly separated. We fully acknowledge that short-term incubation experiments cannot address any of the latter adaptive/evolutionary processes, which will only begin to be important over multiple-generational timescales.

As stated above, we acknowledge that such work is clearly important for understanding the potential long term influence of anthropogenic ocean acidification on marine systems (e.g. Lohbeck et al. 2012; Schaum et al. 2013), although arguably, such information in itself is also difficult to extrapolate to the ecosystem level where ecological interactions and numerous other environmental drivers will interact. However, this was not the aim of our study. Moreover, we note that our observations do in fact indicate clear, statistically robust and repeatable shifts in community structure with treatment effect and much of our discussion was based on interpretation of this effect. We fully agree that the responses involved in such short-term experiments will be driven by physiological (i.e. acclimative) effects, which will act at the individual/population level. However, irrespective of the underlying physiological cause, the differing magnitude of these effects between different phytoplankton populations (specifically as a function of cell size in the current case) must ultimately be responsible for the changes in community structure, which were clearly observed within our experiments.

We also note that equivalent shifts in natural community structure within longer timescale experiments (e.g. Riebesell et al., 2008, 2013) will be driven by within community selection rather than evolutionary adaptation, given the timescales likely to be involved in the latter (see Table 1, which we will amend in the revised version to include some indication of relevant biological processes operating under the observed timescales).

Riebesell U., Bellerby R. G. J., Grossart H. P., Thingstad F. (2008) Mesocosm CO_2 perturbation studies: from organism to community level, Biogeosciences, 5, 1157-1164.

Riebesell U, Gattuso J. P., Thingstad, F. and Middelburg, J. J (2013) Arctic ocean acidification: pelagic ecosystem and biogeochemical responses during a mesocosm study. Biogeosciences, 10, 5619-5626.

Furthermore, it is not possible to draw any definite conclusions about the net production from pure physiological responses. For this, it would be crucial to know the carrying capacities of the communities.

We do not understand the reviewer's point here. Net production is defined as the net increase in biomass/organic carbon. We directly observe such changes. We make no argument concerning the carrying capacity of the system, which may be set by, for example, total nutrient availability, if at all. Carrying capacity is not relevant to our study, which was designed to investigate the sensitivity of phytoplankton processes to short term carbonate chemistry changes.

The only general conclusion they can draw from this approach is the short-term response of total biomass and changes in particular phytoplankton groups.

We absolutely agree and note that within the experiments we performed, a clear treatment-related shift in community structure was observed.

This, however, is not thoroughly applied in the statistical analysis as currently presented (See specific comment 5 below for suggested amendments). Moreover, do the authors think that the investigated communities harbour any response diversity? More precisely, do they expect that the communities would reorganize in response to ocean acidification? Would this have any effect on net carbon production? Please elaborate.

Below we comment on our statistical treatment of the data in more details. In the original manuscript we discussed how observed variability in response between experiments might have resulted from differences in initial community structure (page 3511, lines 17-20). We also discuss at length the potential implications of our results for understanding whether communities will reorganise as a response to ocean acidification. However, as mentioned in the text and again above, it is difficult to extrapolate any such results to a much larger scale. Thus, in response to all these points we reiterate what we stated clearly in the original introduction (page 3494 lines 15-19) *"Although experimentation on natural communities can potentially account for compositional changes, which are highly likely due to both interspecific and intraspecific variations in the plasticity of response (Schaum et al., 2013), they will struggle to account for adaptation occurring through decades of evolutionary processes."*

We would argue that whether any hypothesized changes in community structure would have a subsequent effect on overall net carbon production, which appears to be the reviewer's focus (although was never stated as one of ours), will depend on many factors, including the scale being considered and interactive influences of a wide range of ecosystem and biogeochemical processes. This is why we consider such speculation to be well outside of the study undertaken and the manuscript presented.

2. Each experiment was carried out with the community present on the sampling day under in situ environmental conditions. As the authors state themselves it is highly likely that the communities were not all in the same state of growth when they were collected from the water column. Some might have been in a post-bloom phase and consequently depleted nutrients while others have not been. This makes it certainly difficult to compare the experiments. Please further justify this approach.

As the reviewer acknowledges, we fully agree with this perspective and stated so ourselves in the original manuscript. Within the original paper we discussed at length

how nutrient availability may have influenced our results and our specific inclusion of nutrient-pCO_2 manipulation experiments was specifically designed to test for this. Using these experiments, we clearly demonstrate that the pCO_2 sensitivity we observed was apparently irrespective of the nutrient enrichment level (see Figure 10), although we again note that this result should not be extrapolated without caution. More generally, differences between experiments certainly might be expected on the basis of initial community structure or a range of environmental factors other than nutrient availability. This is precisely why we chose an approach that enabled us to experimentally sample multiple sites and, as we discuss, this has allowed us to identify a strong and reproducible response that occurred in the majority of experiments (e.g. Figures 9 and 10). To put it another way, for the sampled region, the responses that were observed occurred in the majority of experiments and hence were largely irrespective of the sampled initial conditions.

With the present design it is also not possible to ascertain if differences in response among the experiments are owing to environmental conditions, community composition or their interaction. Only a full factorial design (pCO2 x community x environmental (nutrient) condition) would help to mechanistically explain the findings.

Agreed. However, we note that within three of our experiments we did in fact perform a factorial investigation of responses to both pCO_2 and nutrient conditions, effectively across three different starting communities. The pCO_2 effect in these three experiments was thus observed irrespective of imposed perturbation of nutrient availability (Figure 10). Thus, to the extent possible, we already include some factorial investigation of the starting community, pCO_2 and another easily manipulated characteristic of the environment (nutrient availability).

More generally, due to the tight coupling in marine planktonic systems between ecosystem processes and biogeochemical cycling (the cycling of nutrients and carbon), we would suggest that a natural oceanic microbial community cannot be removed from its environment without perturbing both the community and environment. Hence it is difficult to envisage any practical way of fully decoupling community from environment in an experimental oceanographic context. As stated above, we reiterate that our discussion and interpretation are largely based on the robust repeatable responses observed in the majority of the experiments, i.e. the sensitivity of the small cell size phytoplankton community to deliberate pCO_2 manipulation was observed to occur across 6 of our 8 experiments. We subsequently suggest a potential mechanistic explanation for this pCO_2 response, which was observed irrespective of initial starting nutrient concentration (Table 2) or deliberate nutrient amendment (Figure 10).

3. Throughout the manuscript the authors emphasise the high number of replicates featuring their study. However, it seems that they mix up true replicates with replicated experiments. As I understand, the number of replicates per treatment (pCO_2) is three, which is not exceptionally high. By comparison, Krause et al. had five replicates per treatment.

The paragraph related to the experimental set up in the Material and Methods was potentially a bit confusing and we will amend this within a revised version.

Depending on the variable measured there were either 3 or 9 biological replicates per treatment per time-point within every experiment. However, the 'highly replicated' in the title refers to the fact that we performed both types of experiments multiple times i.e. between 3 and 5 times each. Consequently 8 different initial conditions were sampled.

The specific feature of your study is that you consider spatial variability by running several experiments, each at a different geographical location. But it is statistically not sound to throw all the results of the different experiments in a single ANOVA, except when the variable "Experiment" is included as a factor into the analysis. Since a paragraph explaining the statistics performed in this study is completely missing, it is difficult to figure out exactly what the author did (See specific comments below). This needs to be fixed.

We acknowledge that a detailed description of the statistical analysis performed was lacking and will add a brief paragraph on statistical analysis in a revised version of the manuscript. The results of all the different experiments were not included within a single ANOVA, rather, individual time-points within individual treatments were all analysed separately, i.e. a separate ANOVA was performed on the between treatment variability for each individual time-point (i.e. 15 independent comparisons per variable were possible across the whole suite of experiments). It is this high degree of replicability which enabled us to both identify a strong, consistent response within many of the experiments (6 of 8), while further demonstrating that the observed responses are still not generic, i.e. they could not be observed in all experiments.

Specific comments:

1. *The title can be shorter and punchier; what is the main result/message of this study? Since the focus is not the entire 'carbon cycle' this term should be avoided. Also the phrase 'highly replicated' should be rephrased (see my comment above).*

As discussed above, we did not intend the title to suggest that we investigated the 'entire' carbon cycle, rather some aspects of it. However, we accept the comments of both reviewers here and subsequently we propose to change the title to: 'Phytoplankton responses and associated carbon cycling within highly-replicated shipboard carbonate chemistry manipulation experiments conducted around Northwest European Shelf Seas'

2. P 3497, L 14 Does the average natural irradiance vary among sites? What are the values?

The average natural irradiance experienced by the natural phytoplankton communities at each bioassay location will have varied as a function of meterological conditions (cloud cover), mixed layer depth and light penetration through the water column. A comparison between the integrated light dose within the experimental setup and an estimate for the same property in situ will be included in a revised version of the manuscript. Specifically, the integrated daily light dose in the experiments (6.5 mol photons m^{-2} d^{-1}) was specifically chosen to be broadly comparable to the estimated

integrated light dose experienced by the phytoplankton populations, which ranged from 3-17 mol photons m^{-2} d^{-1} (see Poulton et al. same issue)

3. Were grazers included in the experiment or were they removed before water was transferred into the experimental units? Please elaborate. In case grazers were present, what does it mean for the results; please discuss.

Water for the incubation experiments was not filtered in any way.
In the original manuscript we stated 'Water from within the surface mixed layer (<20 m) containing the intact natural community was collected from a unique CTD cast. Once on-deck, the total seawater collected within the 24×20 L CTD Rosette OTE (Ocean Test Equipment) bottles (480 L) was dispensed from randomly assigned OTE bottles through silicon tubing amongst 72 × 4.5 L (E1–E5) or 24 × 1.25 L (E2b, E4b and E5b) acid-washed and Milli-Q rinsed clean clear polycarbonate bottles (NalgeneTM).' However we will further clarify that no filtration step was undertaken in a revised version of the manuscript.

We note that for a natural marine microbial community it is never possible to remove all the grazers as many hetero- and mixo- tropic organisms will overlap in size (and in the latter case ecosystem function) with the autotrophs. Irrespectively, the results are analysed on the basis of the between treatment responses, so any ecosystem level effects (e.g. any hypothesized differential grazing under different pCO_2 levels) could certainly be influencing the results. However, in the absence of clear evidence or any mechanistic basis for such an interpretation we consider this to be outside of scope.

4. *At first reading, I found it difficult to understand the experimental design itself. It took a while to figure out the number of the true replicates within each of the replicated experiments. This needs to be improved.*

We will clarify the method section for this specific aspect.

5. *I miss a paragraph explaining the statistical analysis in the methods section.*

We acknowledge this and will add a section describing the performed statistical analysis.

The focus is on phytoplankton responses, thus it is important to address the inherent biological variability with a proper statistical design. The experiments per se are no true replicates since the authors neither control for environmental conditions nor community composition, and they do not take place at the same time (see general comment). I suggest that the authors run a separate ANOVA for each experiment and the response variables therein they are interested in.

As described above, such a statistical treatment is precisely what was performed. We acknowledge that this potentially wasn't made clear in the original manuscript and will explain in greater detail within an added section on statistical analysis within the revised version.

Then, in order to draw the general conclusion of CO_2 effects across experiments calculate the log response ratios, and finally run a metaanalysis across all the experiments. Only then it would be possible to draw any general conclusions.

As the reviewer points out and we discuss at length, we would not necessarily expect the response to be consistent across experiments. We thus argue that an aggregated analysis is not justified at the current stage. The conclusions we draw are on the basis of the proportion of experiments for which the statistically observable effect was apparent (6 out of 8), but we fully acknowledge that the response is not generic.

6. Particularly the results but also the discussion on the success of the carbonate chemistry manipulation seem to be excessive, notably because other differences in the set-up (see comments above) were completely neglected. I wonder if those parts can be cut.

We feel that the sections relating to the success of the carbonate chemistry manipulation are justified as they directly attest to the reproducibility of the primary manipulation factor which will be driving any observed between treatment effects.

7. The authors could further put their study in context with other recent studies investigating the effects of ocean acidification on phytoplankton community responses (e.g. Yoshimura 2013 J Oceanogr doi: 10.1007/s10872-013-0196-2, Eggers 2014 GCB doi: 10.1111/gcb.12421).

We acknowledge this point, which was suggested by both reviewers. We will hence add and discuss the suggested papers by Eggers et al. 2014 and Yoshimura et al. 2013 alongside some others in the introduction and discussion, both to illustrate the diversity of natural community responses *in situ* and acknowledge that multiple factors may be influencing community structure shifts in different experiments.

8. P 3493, L 7: change 'Egleton' to 'Egleston'

Noted, we will correct this in revision.

9. P 3502, L 14: correlation coefficient is given as 'r' not 'r squared'. Please change.

Noted, we will change this in revision.