Comments on: MS "***Fertilization success of an arctic sea urchin species*, Strongylocentrotus droebachiensis *under CO_2-induced ocean acidification***" Bögner et al. *Biogeosciences Discuss. 10, 8027-8064, 2013.*

General Comments:

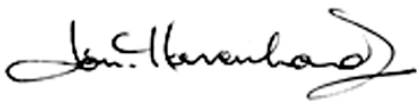
This paper deals with an important issue, namely the potential for ocean acidification to influence gamete performance before fertilization. The overall aims are good and the experiment reasonably well thought out, and reasonably well conducted. The overall structure is acceptable, and some of the findings, such as the observation that OA-induced developmental delay starts even at fertilization, and that the proportion of eggs with "Perfect fertilization envelopes" declines with increasing pCO_2, and increasing pre-incubation time, are interesting and justified. Several parts of the text are overly long (notably the Discussion).

Unfortunately, there are critically important flaws and other results are less reliable (or at least, it's not possible to identify the degree of reliability from the manuscript). For example, the "assessment" of polyspermy – a key component of the work – is highly questionable (see specific comments below). In addition there are many errors of logic and omission in the manuscript, which include lack of adequate referencing, and inadequate demonstration of understanding of some of the background literature (which is matched by inappropriately long consideration of other aspects). The Discussion engages in much unwarranted conjecture based on the (in my view unjustified) classification of "polyspermy".

Lastly, and importantly, the langauge is frequently so poor that it limits the reader's capacity to understand what the authors did, what their argument is, or what their conclusions are.

Unfortunately, I think these errors are collectively so bad as to be fatal, and I cannot recommend that the manuscript be accepted for publication. It may be possible to "rescue" something from this manuscript, focussing on the excellent data for perfectly fertilized zygotes (ignoring the polyspermy issue), highlighting the effects of pre-incubation time, and the developmental delay issues the authors note here. This would, I think, be a worthwhile contribution, however I do not think that's possible within the context of a revision of this manuscript.

I'm happy to explain my comments in more detail should the authors so wish.



Jon Havenhand

Specific Comments:

Title: This species isn't *only* an "arctic" species, it's found from well south of the arctic (at ~ 50°N) to 81°N. The title should be changed to reflect it's boreal distribution, or "arctic" should be dropped from the title.

p28 li 7: (and later) Why use "untreated filtered seawater" as a "control"? What are the authors controlling *for*? This makes no sense. They should either explain or drop any mention of this.

p28 li 24 ff: This introduction to the effects of elevated pCO_2 is somewhat dated and could be rewritten to reflect more recent projections & models.

p30 li 27 ff: If release of gametes in viscous fluids increases their longevity it may also moderate impacts of seawater pH. Therefore the authors' conclusion here that released gametes are exposed to lowered pH conditions that are typically used in fertilization assays is untenable.

p31 li 7 ff: The authors here reach a conclusion regarding the effects of pCO_2-induced pH change and base this on Hamaguchi et al's results using acetic acid. In the absence of any justification for why these two different methods of acidification are comparable, this conclusion is invalid.

p32 li 3 ff: The authors should state the pH / carbonate chemistry of the aquarium water during transport and holding at AWI. How can they be sure that their husbandry techniques didn't influence the pH of the water in which adults were held *prior* to experimentation?

p33 li 5: "controlled" is misuse of English – I assume that as the authors claim on the previous page that they manipulated pCO_2 without disturbing alkalinity values that the authors mean "checked". This should be changed.

p 34 li 1: Dilution to a factor of 10^4 is not a sperm "concentration", and tells us nothing about what the actual concentration was (other than that this was extremely high!). The authors should state clearly here what the concentration was.

p34 li 12: Why were different experimental conditions used for the with- and without-pre-incubation treatments?? The authors cannot separate the effect of treatment from the effect of experimental container (25 ml *vs* 100 ml)!! Whilst this is an unlikely explanation for the differences seen between these treatments, it's completely avoidable and has unnecessarily confused the experiment. This should be explained.

p35 li 15: Again, the authors provide inadequate information: they added "50µl of solution of *Rhodomonas salina*" to the hatched larvae each day, but what was the final resulting cell concentration? This should be given.

p36 li 13: What were the relevant factors and response variables used in the GLMM? Why was GLMM used? Which factors were fixed and which were random? This should be stated clearly here.

p36 li 16 ff: The sentence "In addition, we included as a random factor . . ." makes no sense.

p36 li 18: What ratiometric measurements? Ratios are notoriously non-normal, yet the authors used ANOVA here. Why? How did the authors ensure the data were normally distributed? Such information should be given here.

p36 li 25: The reference to "constant water parameters" is inappropriate: it's not possible to see from Table 1 whether the parameters varied over time, nor is it possible to see the parameters for each experiment (i.e. with, and without pre-incubation).

p37 li 13: The authors must justify their claim that the variables used were "independent from each other". Data?? Do they mean there were no statistically significant interactions between the 3 variables? If so they should cite the relevant statistics.

p37 li 14: Although I'm familiar with GLMM and factorial analyses in general, I have no idea what the authors mean here by "used the continuous pH variable in statistical analysis". Do they mean that pH was used as a covariate? Again, the language use obscures the authors' meaning.

p37 li 22: The authors provide no valid scientific justification for the classification of eggs/embryos they show in Table 2 and present in Figure 4. Assessing polyspermy is notoriously difficult (see e.g. Franke et al 2002 *Am. Nat.* – which is, astoundingly, not cited in this study!), and the authors show no indication that they understand this. In the absence of better justification I cannot accept these data as valuable and meaningful.

p38 li 15 ff: The authors provide no reference to their own data (Figs, Tables, statistics) to support the assertions made here. They should.

p40 li 7: This statement is incorrect: not only is this definition constrained to echinoderms that demonstrate a clear fertilization envelope, several authors, including myself, have made the argument in print that successful fertilization is much more than the presence of an elevated fertilization envelope: such a response can be obtained in polyspermic embryos and tells nothing of the rate of monospermic fertilization.

p43 li 15: The authors here perpetuate the fallacy that a non-significant result (the cause of the "robustness" claimed by Martin et al, Byrne et al, cited here), is biologically meaningful. In a null-hypothesis test, a non-significant result is inconclusive (Fisher, R.A. *The Design of Experiments* 1935). Consequently the absence of a statistically significant effect is not evidence that the treatment doesn't have an effect. This is a fundamentally important philosophical, and statistical, point.

p46 li 12ff: The fact that the BCECF-AM pH indicator system wasn't able to identify significant differences in pH_i between 180 and 980 ppm CO_2 or between 1400 and 3000 ppm CO_2 (Figure 6) indicates either that these cells only regulate pH_i at levels over 1000 ppm, and that regulation is the same at 1400 and 3000 ppm, or that the dye system was very insensitive. The authors show no calibration data, or positive controls, to indicate that this system was indeed working as they expected (others have used this same system with considerably more success). This brings the reliability of these results into question.

Fig 3: This Figure must be revised: it's statistically meaningless to show the "with-" and "without-" pre-incubation data together with "All-experiments". It is also unclear from the legend exactly why some letters are capitalised and some not. Lastly, it's completely meaningless, and misleading, to analyse (and cite statistics for) the "Unfertilized" fraction of the eggs: this is by definition 1 - the Fertilized fraction and therefore conveys no additional information, although it gives the impression of greater statistical certainty. The results of tests of Unfertilized (or Fertilized) should be removed from this figure.

What time point are the data in this Fig from? Is this 1h? 3h?? The authors should state this clearly.

Fig 4: Are these data for with- or without- pre-incubation? This should be stated clearly.