We are grateful for the opportunity to make the corrections based on the detailed reviews received. All comments were very valuable to us. Considering the number and importance of both reviewer's comments, we have made extensive changes and thus prepared a new version of our manuscript. We have not highlighted changes in the text but tried to respond to the reviewers' comments below.

Response to Reviewer #1 Inge van Dijk

Major comments:

I was surprised to read about the cleaning methods used in this study. When investigate chemical composition of foraminiferal shells, we use a much more intense cleaning method with oxidation and reduction steps (especially when specimens are collected in the field), to remove any organics as well as diagenetic coating. I wonder if e.g. only mechanical removal of any organisms present on the shell is enough to remove any (organic or chemical) trace completely, and if coatings are present on the shell after the described cleaning protocol. Are there studies comparing different cleaning techniques?

Response: Concerning any surface contamination, the difference between the foraminifera and mollusk shells is the specific surface area, which is low in case of relatively large shells. The argument against chemical cleaning protocols was based on the observed preferential leaching of Mg (and, therefore, potentially many trace elements) out of carbonate skeletons during chemical cleaning (Loxton et al., 2017; Mannella et al., 2020). We have added a paragraph in a discussion section (lines 383 – 401) to draw the attention to the fact that in this study the whole shells were analyzed, along with organic matter.

Furthermore, the amount of organic inside the $CaCO_3$ might have a huge effect on the elemental composition of the shell, e.g. on Na, and the contribution of organics might differ between species and $CaCO_3$ polymorphs, as also acknowledge by the authors (in l.54: 'species-specific organic matrix'). With the method described in this study, the organic matrix will also be analyzed. I would like to see this issue discussed in the revised version of the manuscript. Has there been any study on the chemical composition of the organics in different species of bivalves, albeit on the compounds of the matrix or by microscale-analysis of the shell with e.g. nanoSIMS on cross-sections? This should be at least mentioned in the discussion as a potential reason for the offset between species, if not discussed in full detail.

Response: Removal of organics without mobilisation of any trace elements associated with $CaCO_3$ is not a task that is easy to achieve. There are good studies on this subject, e.g., (Barker et al, 2003; Holcomb et al., 2015; Loxton et al., 2017; Mannella et al., 2020). However, we do not believe there is a single accepted protocol for bivalve shells, that is tested and validated for a large range of trace elements. We, therefore, have opted for the analysis of the bulk composition instead of trying to isolate and analyse the $CaCO_3$ phases.

In the revised version of the manuscript, we emphasise that the contribution of the organic matter to the elemental variation is likely to be relatively minor considering that bivalve and barnacle shells contain, in general, less than 5% organic matter (Bourget, 1987; Marin and Luquet, 2004; Rueda and Smaal, 2004)

The authors are comparing small and big adults and use the obtained data to look at ontogenetic effects. However, the authors also state this is a very variable environment, with big seasonal (and maybe yearly) changes. Overprinted on the size effect, there is a time effect: the bigger specimens have recorded events that the smaller specimens did not experience. I'm missing the longevity of the different species in the discussion of the results. For example: The size effect observed for A. improvisus (life span = 1 year) might simply be a seasonal signal in food supply (and thus maybe growth rate) or

physio- or chemical parameters like seawater temperature. If the samples are taken in end of summer (sample date not mentioned in the manuscript, should be added), it would explain why the Mg of the larger specimens is lower: lower temperatures lead to lower incorporation of Mg and these larger specimens likely experienced the winter period, while the smaller specimens maybe spawned in spring. As for the species with longer lifespans (of 10-12 years), could the decrease in element incorporation with size (thus, for older specimens) be due to increased heavy metal output of the Vistula river over the last years? Is there any (historical) data on this?

Response: In the revised version of the manuscript the discussion has been extended about the size effect on observed patterns of elemental levels with emphasis on environmental influence on it. As a result, patterns related with the biological effect of organisms have been more clearly justified. Unfortunately, we have been unable to present any data on the recent heavy metal output of the Vistula river, but we have added the paragraph 4.3, which contains environmental data describing the Gulf of Gdańsk in the context of elemental concentrations in shells.

The authors see some difference between size classes, they conclude that in general smaller specimens have increased trace metal incorporation. Can this also be due to absorption of elements or diagenetic precipitation on the outside of the shell, compared to the more pristine CaCO3 below the surface, leading to a surface/volume effect? E.g. larger specimens have thicker shells, and thus lower surface over area ratio?

Response: This potential issue has been mentioned in the discussion (lines 454 – 456).

I would also like to see habitat depth included in the discussion. The authors are assuming the shell chemistry of the different organisms tested are all reflecting either food or ambient seawater chemistry, but some organisms are living on hard substrates, while other live a few cm in the sediment, and the most extreme one (Mya arenaria) can live 20-30 cm in the sediment. The latter species would have a totally different "ambient seawater conditions", as it is exposed to interstitial water that is very likely to have totally different chemical signature then the overlying water. It would probably be in contact with e.g. much higher Mn concentrations. This is not reflected in the shell chemistry, so maybe this species does not take up elements from the seawater, but more from food intake? For your purpose it would be best to have also some kind of idea of the (evolution) of trace metal concentrations in seawater over time. Is there any chemical data on the seawater available from this area? E.g. about the metal concentrations close to the Vistula river (l. 478)? Is there data showing that station GN is indeed increased in heavy metals compared to the other stations? These other stations are located in a bay area, making it possible the residence time of the water is higher, and there might be an actual increase in the metal concentration here.

Response: The sediment type, feeding strategy and environmental sources of metals are important factors affecting the concentration of metals in shells and were discussed in the revised version of manuscript. We have added information about sediment type in the study area and have put more emphasis on discussing feeding strategy in the context of metal bioavailability.

In retrospect, for the main goal of the study, it maybe would have been better to not analyze full shells, but make small aliquots/subsamples by e.g. drilling the shell. Is there any data (in literature) on small-scale variation in the shells of (some of these) species?

Response: Our strategy of investigating the whole shells has been a deliberate choice in order to achieve better detection limits (comparing to e.g. LA-ICP-MS), and to analyze many individuals of different species instead of focusing on a thorough analysis of single individuals. We have presented data on the concentration of 12 metals in shells of mussels and barnacles from the Baltic populations (in total over 2,300 measurements), which have not been previously studied.

In my opinion, at the moment, you have a combination of too many variables: CaCO3 polymorph, different stations environmental variables (incl. unknown chemical compositions of the seawater), size effect and the vital effect (calcification pathway) of the organisms. It becomes very difficult to disentangle different drivers of shell chemistry, which means you have to be more careful in your conclusions, or at least convince readers which variables are minor/neglectable. I think some variables, like the different sampling stations, can be convinced as being minor, by showing chemical variability or the hydrological situations between the stations.

The authors often point out the strong seasonality in this region, section 2.1 and through the manuscript, e g. l. 481-483. Maybe it is possible to add a (supplementary) figure to section 2.1, if needed compiled from literature data, about the environmental variability in this area, to show differences in physio-chemical parameters. This way, readers, like myself, that are not familiar with the study area can have a good overview of the (yearly) environmental variability in this area.

Response: To make the manuscript more accessible for readers, the aims of this study have been clarified (lines 119 – 128) and the discussion has been divided into the three parts: $CaCO_3$ polymorph type and elemental concentrations; Size classes and potential biological impact on elemental concentrations; Environmental factors and elemental concentrations. As conclusions, we have distinguished patterns of inter-species and inter-individual variations in the concentration of elements in studied shells, which are associated with biological and environmental control over the biomineralization process.

Minor comments:

(Since I have a lot of major points for the discussion section, I give minimal textual changes, since I believe the manuscript, especially the discussion, will probably greatly change after revision.)

Throughout the manuscript:

-Change 'Mg/Ca ratio' to 'Mg/Ca'.

Response: Modified where sensible.

-Check manuscript for (double) bracketing issues, for instance in l.207: '(Darwin, 1854) (Arthopoda, Maxillopoda)' should be e.g. '(Arthopoda, Maxillopoda; Darwin, 1854). Also lines 217, 228, etc. For l.131 and l. 515: reference should not have brackets.

Response: Corrected where applicable.

Abstract:

-The abstract as it reads a bit stiff. Please consider rewriting this section. For example, l. 28-29 on sample location can be merged with the first sentences, while line 29-30 is an explanation of the method, which should be either removed, or shortened, in my opinion.

Response: The abstract has been rewritten.

-l. 26-27: 'The potential impact of environmental factors on the observed elemental concentrations in the studied shells is discussed': Is this really the case? Since there is no data on the environmental parameters presented, it is difficult to discuss the data in this framework.

Response: Although we have not measured the concentrations of elements in the environment, we believe the discussion about their impact on the composition of the shell is challenging, yet very valuable. In the newer version of the manuscript

more information based on the literature data about the environmental characteristics were added, and we based the discussion on it.

Introduction:

-l.64-65: 'crystal layers are precipitated successively at regular periodicities, 'is not true for all marine calcifyers, like Foraminifera. make it clear when you switch from all marine calcifyers to marine invertebrates.

Response: This has been improved.

-l.88-90: maybe add Stanley, 2008, it is a nice overview paper. Stanley, S. M. (2008). "Effects of global seawater chemistry on biomineralization: past, present, and future." Chemical reviews 108(11): 4483-4498.

Response: This has been included.

-l.144: remove . after shells

Response: This has been improved.

Method section:

-l. 263: When were the samples taken?

Response: The samples were collected in May 2013 and June 2014. This information has been presented in Table 1.

-l. 295 What was the Ca concentrations in the sample solutions? 100ppm or varying?

Response: Further details have been added to the revised manuscript.

-l.297: What were the accuracy and precision of the measured elements?

Response: We have added such details to the method section (lines 256 – 260).

Discussion section:

-I would advise to divide the discussion session in smaller paragraphs to increase readability.

Response: The discussion has been divided into the three parts to make reading easier.

-I would like to see variables as life span, habitat depth and organic material in the shell (see above) included in the discussion.

Response: This has been added and existing parts improved.

-l. 372: You obtained specimens with the same polymorph from two contrasting temperatures: i.e. aragonite Cerastoderma glaucum (16.9◦C), Limecola balthica (4.6◦C), Mya arenaria (16.9◦C), I would like to see a discussion on the (absence of) temperture effect Sr incorporation, which is currently lacking in the manuscript, while it is being discussed for Na and salinity.

Response: The environmental variables have been discussed in more details in the revised version of the manuscript.

-l. 478: ref? Are there any studies on this?

Response: The literature data regarding the concentration of some studied metals around the study area (such as Rainbow at al. 2000; Rainbow at al., 2004) and sediment type has been included into the manuscript.

-l. 483: suggest to change 'animal' into 'organism'.

Response: This has been changed as suggested.

-Fig 2 and 4: indicate which polymorph of CaCO3 is used, like Fig. 3.

Response: Figures and tables have been improved to make them easier to read.

-Fig. 3: where possible, please use the same scaling for the y-axis for comparability, e.g. y axis of Mg for aragonitic species.

Response: Figures and tables have been improved to make them easier to read.

Response to Reviewer #2 Anonymous Referee

Fundamental information regarding the measurements and concentration calculations is missing or unclear, and the analytical uncertainties and repeatability based on appropriate reference materials are also lacking. In addition, in my opinion, the sample preparation and pre-cleaning procedures are questionable. The authors must prove the validity and explain their analytical procedures, and take into account analytical uncertainties when presenting or interpreting their data.

Response: The analytical details have been described in more detail in the revised version of manuscript.

Furthermore, as detailed below, I have problems understanding how whole shell bulk measurements may be used to assess the role of ontogeny or even environmental variations. By using entire shells, the authors 'average' the composition of the growth lines precipitated during earlier and later stages of life, as well as the composition of growth lines built during different seasons or under different environmental conditions. Thus, I am not sure that by comparing bulk values from smaller vs. bigger (younger vs. older) individuals it is possible to determine whether environmental or ontogenetic controls drive the composition of the shell. Simply, the differences between the mean bulk values would depend on the elemental variability encompassed in the shell, which would depend on individuals' growth and environmental conditions experienced. The mean bulk values from older individuals integrate large intra-shell variabilities, while in younger individuals smaller intra-shell variabilities, but I think that with this design it is difficult to disentangle the underlying controls on the elemental composition of the carbonates.

Moreover, the problem of using bulk also limits the interpretation of the data when it comes to the different polymorphs, and particularly this is the case for the bimineralic bivalve. Here, the authors may only conclude whether the composition of the mixture is different to other species building pure calcite vs. aragonite. However, it does not answer the question whether the composition of the calcitic part or aragonitic part within fundamentally differs and how much, which I think is the relevant question here. When discussing the composition of the bimeralic bivalves the authors could, at least, attempt to estimate the contribution from each polymorph to the mixture, and discuss the implications.

Moreover, I am wondering why the entire shells were crushed? Surely, >100 mg is not required for the analyses, as concentration measurements are typically done on <mg level. Why did the authors decide to measure the entire shell instead of e.g. a profile across the shell or different growth bands? Such approach, I believe, would be much better for defining an ontogenetic trend, and could also provide some insights into the intra-shell variability. The intra-shell variability, in particular, would be very meaningful to assess before any mean bulk values are used for interpretation of ontogenetic or environmental signals – i.e. how heterogenous are the shells, what is the driver, is it random or not, how

big is the variation and what it reflects? I think it is really a shame this was not considered beforehand as a great amount of information from the shells is lost when measuring the whole shells rather than specific parts. Furthermore, I am not convinced that comparison of bulk large vs. small individuals, in this case, answers the question whether ontogenetic trend drives the elemental variability. For numerous calcifiers group, the partioning of elements between seawater and the carbonate is within a certain range band "baseline" which is principally determined by their calcification mechanisms and mineralogy, and then this variability of the "baseline" may be driven by environmental factors. In such case, simply by a probability, larger individuals would have lived longer vs. smaller individuals and thus likely witnessed during their life time more environmental fluctuations (e.g. temperature, nutrients, pH, O2, etc.). Thus, when using an average of an entire shell, it is reasonable to assume that the mean of the shell integrates larger intra-shell and therefore elemental variations in the older individuals in contrast to the younger, simply because they experienced more changes over their life. I believe that this is also quite apparent in Fig. 3. How may one, therefore, discriminate between ontogeny vs. environmental variability?

Response: Our strategy of investigating the whole shells has been a deliberate choice in order to achieve better detection limits (e.g. LA-ICP-MS), and to analyze many individuals of different species instead of focusing on a thorough analysis of single individuals. We have presented data on the concentration of 12 metals in shells of mussels and barnacles from the Baltic populations (in total over 2,300 measurements), which have not been previously studied. The manuscript has been significantly modified in accordance with the reviewers' suggestions and now provides a better discussion of the major points raised above.. Furthermore, the discussion now includes more detailed environmental dataset (based on literature data) to draw more robust conclusions about the biological effect.

As mentioned in the responses to Reviewer #1: Removal of organics without mobilisation of any trace elements associated with $CaCO_3$ is not a task that is easy to achieve. There are good studies on this subject, e.g., Barker et al (2003), Holcomb et al (2015), Loxton et al (2017). However, we do not believe there is a single accepted protocol for bivalve shells that is tested and validated for a large range of trace elements. We, therefore, have opted for the analysis of the bulk composition instead of trying to isolate and analyse the $CaCO_3$ phases. In the revised version of the manuscript we have emphasised that the variation observed could also be due to the presence of organic material within the carbonate structure.

In my opinion, the authors need take into account these problems, before any interpretations can be made. While a great deal of information is unfortunately lost by using average values, and I think the authors really have to reconsider the interpretations that can be made from this data and discuss their limits, I do acknowledge the authors' efforts for measuring numerous individuals, which I do not think is often done, and perhaps a point to that could be better taken advantage of. Just as a suggestion, maybe, this could be of use for defining the "typical range" for each element for each species in the Gulf of Gdansk, which could be then compared to literature values from same / similar species in other parts of the world with very different settings. If possible, I think it would be interesting to see how the general elemental concentrations and variability compares between regions or not, and could be of use when constraining environmental influences on the biomineral composition. In addition, I would also like to see a comparison between the different sampling sites within the Gulf of Gdansk. While on one hand it could be perhaps assumed that the differences between the sites are negligible, this is a very dynamic environment, and it might be that spatio-temporal variations account, at least partially, for some of the observed variabilities.

Response: These issues have been discussed more clearly in the revised version of manuscript.

One thing that has surprised me the most about this study is that, despite the careful organism sampling strategy, the authors did not consider collecting and measuring water samples. In my opinion, this should come first in this kind of

studies, and something I was expecting to see, and thus a real shame it was not done, especially since the authors had the opportunity to do so (and elemental analyses on water samples are relatively more straightforward that on carbonates). Data on seawater chemistry is critical for the calculation of partitioning coefficients, which could ease the interpretation of the results from different sites (in the case that the chemistry at the different sites strongly varies). While it may be a tall task to ask for the measurements at this stage, the authors should, at least, compile the available information on local concentrations of elements in seawater (including additional physico-chemical characteristics), and estimate the partitioning coefficients for each element for the different species.

Response: It is hard to not agree with the reviewer on above aspect. And because we have not measured the concentration of elements in the environment, the discussion about its impact on the composition of the shell is challenging, yet very valuable. In the newer version of the manuscript we have put more emphasis on the environmental characteristics based on the literature data. This also include sediment type, feeding strategy and environmental sources of metals which are all important factors affecting the concentration of metals in shells. All these issues were discussed in more details in the new manuscript.

Specific comments:

Line 1-2: I would suggest to reconsider the title language ' –composition 'and 'com- posed', as well as 'different 'twice in the same sentence, this is not orderly.

Response: Based on the reviews received and the modification that we have introduced in the manuscript, the title has been changed to: The patterns of elemental concentration (Ca, Na, Sr, Mg, Mn, Ba, Cu, Pb, V, Y, U and Cd) in shells of invertebrates representing different $CaCO_3$ polymorphs: a case study from the brackish Gulf of Gdańsk (the Baltic Sea).

Line 32: 'Mg > Sr > Na 'this needs a written definition first.

Response: This was improved.

Line 195-197: Here, it would be particularly useful to provide concrete numbers on the local carbonate chemistry (other than Ω). Ideally, this should have been measured upon the collection of the specimens from in situ water samples, however, if this is not available the authors could at least summarize the information from the literature. An overview table with the physico-chemical characteristics of the local waters (including temperature and salinity trends etc., carbonate chemistry as well as the elemental composition), would be particularly useful.

Response: As mentioned above, available literature data about environmental factors has been included in the revised manuscript.

Line 267: Why no water samples were collected?

Response: We agree that environmental research would be very useful in the discussion, yet unfortunately we do not have them. Therefore, we have used the data available in the literature for the discussion of our results.

Line 282-286: I have difficulties following this protocol and serious doubts on its effectivity and validity. Previously, the authors state that the periostracum was first physically removed. This is good and indeed important as it constitutes a large amount of organic material, which is difficult to treat chemically without having an impact on the carbonate. However, organic rests might still be present on the inside of the shell for example from the mantle, and foremostly in the pore spaces. Thus, physical cleaning is insufficient, and at least at a powder stage it is a generally established routine to apply a cleaning protocol step, consisting of oxidation of organics by buffered hydrogen peroxide (Barker et al., 2003 G3

4, 8407). As far as I am aware, this protocol or close adaptations are commonly applied to a wide range of calcifiers from forams to corals, bivalves and even brachiopods. In this sentence the authors indeed mention the use of H2O2, but only after the dissolution of the sample, which logic I cannot follow. All in all, I do not think that this is the correct way to treat carbonates samples, and would strongly recommend to first demonstrate the validity of this protocol (if the authors insist on using it, or follow a more broadly used protocol such as that of Barker et al., 2003).

Response: We have discussed these issues in more detail in the revised manuscript. However, we do suggest, as Inoue et al (2004) did that "plausible pre-treatment method [for the removal of organics] is yet to be established". With full appreciation of the importance of the protocols discussed and tested by Barker et al (2004) we believe that the benefits of any chemical treatment still remain controversial, see, for example, discussion in Holcomb et al (2015) and Loxton et al (2017). Mannella et al. (2020) showed that the suitability of chemical pre-treatments for organic matter removal from carbonate matrices should be evaluated on a case-by-case basis and, in case of relatively low organic content, should be avoided. We have discussed the potential contribution of organic matter in the revised version of manuscript.

Also, when it comes to ontogenetic trends, let's take for example bivalves and specifically Mytilus, as far as I am aware, broadly speaking their shell growth follows von Bertalanffy growth curve (see e.g. ig. 3; Steffani & Branch, 2003; Mar Ecol Prog Ser 246, 197-209), which is common for many calcifiers. This means that during the very early shell formation the carbonate precipitation is relatively faster, which for the incorporation of numerous elements translates into kinetic effects. It is thus the geochemical composition of the umbo and the first growth lines vs. the latter growth lines (the ones at the growth 'plateau') that form the greater part of the valve that is commonly at- tributed to being driven by ontogeny. Potentially, in the case of the very small and thus very young individuals, their geochemical composition may reflect one environmental condition e.g. certain season and one ontogenetic stage i.e. the one dominated by kinetic factors, but I am not sure this can be directly compared to older individuals which mean elemental composition then reflects different ontogenetic stages (with potentially different contribution of each to the bulk), and broad range of seasons. Or am I missing something?

Response: As mentioned earlier, this part of the discussion has been elaborated based on a more detailed environmental background. This, in our opinion, strengthens the inference of potential biological control on the elemental concentration in shells. Despite the influence of many factors, we have observed statistically significant patterns of metals concentration in shells.

Line 294: What type of solutions? What do you mean by matrix-matched – one solution for each carbonate polymorph? Please provide more details.

Response: The method has been described in more detail in the revised manuscript.

Line 300: Why were the standards not treated the same way as samples? First, I do not think it is acceptable that the authors do not process the standards and the samples in the same way, and second, I do not think that the standards are representative and should be compared to these samples. The authors need to provide the measured absolute values (as well the relative standard deviation over the analysis period at least) of comparable biogenic standards such as JCp-1 or JCt-1, or similar internationally accepted alternatives.

Response: The authors appreciate that using biogenic standards would potentially be preferable; however, those standards were not available to the authors at the time of analysis. On the other hand, complete digestion eliminates potential uncertainty that might originate from potentially incomplete conversion of organic matter typically performed when digesting by HNO_3 only (Inoue et al 2004). The use of reference limestone and dolomite for the control of the analysis

(without considering the digestion step) is fully justified for the digestion protocol used ($HF+HNO_3$ with evaporation and matrix modification to HNO_3 solution of the same concentration). The digestion step includes $HNO_3+H_2O_2$ mixture, which is perceived to be suitable for digesting $CaCO_3$-based materials with low amount of non-refractory organic material. For comparison, a well-cited paper on the composition of JCp-1 or JCt-1 standards (Inoue et al 2004) employed a milder treatment of HNO_3 only at room temperature. We agree that a full method validation employing homogenized samples of clams containing high level of organic matter (in addition to biogenic reference materials, which potentially do not cover the natural range in terms of organic matter content / reactivity) would be desirable, but this must be a subject of a separate study.

Also, regarding the methodology, I am wondering how were the obtained counts converted into concentrations; e.g. did the authors use a calibration line for this or standard-bracketing?
Response: Calibration has been performed typically using 5 points covering the range of concentrations.

Did you normalise all measurements to a stable concentration of a selected element, e.g. Ca?
Response: We have not normalized measurements to Ca, but we have discussed the variability of Ca concentrations.

What was the precision of the individual analyses, and the long-term reproducibility? How many times was each sample measured? Line 305 'most trace elements – 'which elements were measured in He mode and which not? The authors must provide these details with rigour. Line 307 'periodic analyses 'do you mean the standards were not measured along with the samples in a sequence? I have serious doubts on these analytical protocols, and especially do not consider it a good practice to not include standards along with samples in a run.
Response: We have updated the Material and Methods section, in the revised manuscript, to include the following details: ICP-OES: The accuracy and reproducibility of the analyses were checked using two calcium carbonate-rich certified reference materials (CRMs): JLs-1 Limestone and JDo-1 Dolomite (both from the Geological Survey of Japan) prepared by total digestion method (using hydrofluoric acid). The reference materials were diluted to match the concentrations of Ca in sample solutions. Ca, Mg and Sr concentrations were found to be within the uncertainty (1 standard deviation) of the reported values (Imai et al. 1996). Limits of quantification (LOQ) in solution for ICPMS were generally determined as a concentration corresponding to ten times standard deviation of the signal obtained by analysing 5% HNO3 solution (6–7 times) in each individual run. ICP-MS was run in helium (He) mode (5 ml min−1 He, 99.9995% purity) for lighter trace elements (V, Mn, Cu, Y and Cd) to minimize the molecular interferences from plasma and solution components and Ca from samples. The accuracy and reproducibility were checked by analyses of JLs-1 and JDo-1 before and after every batch of samples. The results obtained for all elements (Table 1) were within the uncertainty (2.5 SD) of the recommended values (Imai et al. 1996). Accuracy of Pb determination cannot be checked using these CRMs because of the large spread of reference values probably due to insufficient homogeneity of Pb distribution in these samples. Based on the analyses of CRMs and matrix-matched solutions, the maximum analytical error for the typical range of concentrations in the shells can be estimated (in relative percentage) as 1.5% for Ca, Mg and Sr; 3% for Ba; 20% for Cu and U; and 4–10%for all other elements.

Line 311: I would really welcome some visual representation for this – i.e. pictures of the different species, maybe with the different ontogenetic stages for each. It is really shame this is not provided; the authors study various interesting species, which offers an opportunity to include visually appealing picture figures, which is not used. Perhaps this is too

much to ask, but given that the species build very different carbonate types and I assume microstructures, scanning electron microscope images could also be very relevant and interesting here.

Response: Unfortunately, we have not had the opportunity to make scanning electron microscope images, but we include macroscopic images presenting the studied species (Figure 1).

Line 321: Throughout the Results section the figures are referred to very sporadically only, and there are several instances that a value is given and a statement is made, however the figure is not referred to afterwards. Foremostly, all individual panels of the figures need sub-categories (e.g. a, b, c, etc. please check the Biogeosciences format style), and need to be mentioned where the individuals results are being discussed.

Response: This has been improved in the revised version of manuscript.

Line 322: I am not sure what the authors mean here, please rephrase.

Response: This has been improved in the revised version of manuscript.

Line 327: I would say it is more appropriate to use µg/g rather than mg/kg.

Response: This has been improved in the revised version of manuscript.

Line 328: When concluding that some elements were 'generally present at higher concentration' or lower please also provide the concrete numbers in the text, here, but also in further parts of this section it is missing.

Response: This has been improved in the revised version of manuscript.

Line 334: What do you mean by 'lack of ontogenetic trend'?

Response: This has been improved in the revised version of manuscript to be clearer.

Line 371: The entire Discussion section needs major revision, and foremostly substantial reorganisation in order to make it more suitable to the readers and a wider audience. I am aware that dealing with many different variables like several elements, size classes, species and carbonate polymorphs is not easy, but the authors really need to find a better way for presenting their findings and extracting their 'main mes- sage points' to the audience. At the moment I find the Discussion very broad and, to me, it does not provide clear answers to the research questions. I am afraid that often problems are addressed that cannot be resolved by the present dataset. I would say that it is better if one or two key points are discussed in-depth rather than touching on the surface many (these may still be mentioned, but the in a more concise form, with focus on the key points). The structuring is also relevant for the other parts of the manuscript and especially the Results section. I would start with ensuring that were possible, the geochemical data is presented in a more systematic manner. The Discussion could benefit from being divided into different subsections, where different aspects are being discussed. The data quality and limitations need discussing, as well each of the different factors controlling the incorporation of the elements into the carbonate (preferably in different subsections), a comparison to other studies, and the implications of the presented findings (for e.g. biomineralisation, application as recorders of environmental conditions). At this stage, it is difficult for me to make a concrete suggestion on how to subdivide this, the authors need to see what works best when structuring the Discussion and the message they would like to convey. I would also suggest to separate the Results section, perhaps by species could work well for this part.

Response: The manuscript has been reorganized to be more accessible for readers. The aims of this study have been clarified (lines 119 – 128). The discussion has been focused on finding patterns of inter-species and inter-individual

variations in the concentration of elements in studied shells, and was divided into the three parts: $CaCO_3$ polymorph type and elemental concentrations; Size classes and potential biological impact on elemental concentrations; Environmental factors and elemental concentrations.

Line 372: There are numerous studies on Mg and Sr in carbonate, which uses and incorporation mechanisms, potential proxy-applications etc. need a better summary. Same for all other elements, the discussion of each element should be opened by the factors that control its incorporation into the carbonate. Also, as these are often not similar for calcite and aragonite, and especially since this study is focused on the incorporation of elements into different polymorphs, these two should be treated separately.

Response: This has been improved in the revised version of the manuscript. We put more attention on the factors that control the incorporation of elements.

Line 375: The statistics should be provided in brackets. Also, please be specific, how much?

Response: This has been improved in the revised version.

Line 378: 'Mg was the dominant impurity', please rephrase, what do you mean?

Response: This has been improved in the revised version.

Line 383: Please be specific, what species?

Response: This has been improved in the revised version.

Line 397: What is the origin of the high Sr in barnacles?

Response: This issue has been improved and discussed in the revised version of the manuscript.

Line 405: The concentrations are sometimes given in mg/kg and sometimes in wt%, which is confusing. Please be consistent throughout the manuscript in figures, and this should be preferably µg/g.

Response: This has been improved in the revised version.

Line 414: Please explain, what do you mean?

Response: This has been improved in the revised version.

Line 478: I wonder how would the data look if the metal concentrations are plotted as a function of the distance to the Vistula River mouth? Can you conclude that it is the contamination that controls the trace metal composition? A comparison to the species from non-contaminated water might help.

Response: In the revised version of the manuscript we have included more information about the environmental background. Unfortunately, we do not have the data that could be used to create a mentioned figure, but we have given more attention to this in the discussion. We also have included the data in Table 4, that show the concentration of elements in the shells of the studied organisms from regions other than the Gulf of Gdańsk.

Line 481: Yes, and it is really necessary to add that the whole shells were measured. Therefore, the mean values integrate these variations.

Response: This has been improved in the revised version of the manuscript.

Line 486-489: Please rephrase. Also, of course, they varied but it is difficult to determine why.

Response: In the revised version of the manuscript, the variability of elemental concentrations has been discussed with the new details, and the discussion has a new structure. The enrichment of environmental data facilitates the inference of the biological contribution to elemental concentration in shells. We were not able to define exactly which factor is responsible for the variability, but we present suggestions whether the incorporation of a given metal in the shell is more influenced by biological or environmental factors.

Line 497: 'chemical profiles' please rephrase, as far as I am aware no chemical profiles were made.

Response: This has been improved in the revised version of the manuscript.

Line 479-509: These sections contains many redundant parts, and the discussion could be sharpened.

Response: This has been improved in the revised version of the manuscript.

Line 510: In addition to relative increase or decrease in concentrations, also the variability in the elemental concentration for a size class should be considered (although I am not sure if the differences between size classes will be significant).

Response: This was improved in the revised version of the manuscript.

Line 520: Please be specific, which trace elements (please provide in brackets; similar cases can also be found in other parts of manuscript).

Response: This has been improved in the revised version of the manuscript.

Line 527: Yes, but as mentioned I doubt this has anything to do with the size/age.

Response: This has been improved in the revised version of the manuscript.

Figure 1: Please provide the full site names in the figure caption to abbreviations. What are the grey lines in the big panel (bathymetry?), please specify in caption as well.

Response: This has been improved in the revised version of the manuscript. Yet in case of the sampling site names, they are given symbol by us therefore they have no full names.

Figure 2: This figure needs error bars. The analytical uncertainty should be shown here, as well as the variation of the mean i.e. the 2SD of the mean for each group and the respective n should be provided too. Also, what size classes were used for this? Is this the mean of a certain size class or the mean of all individuals, this needs definition in the caption. It may be more appropriate, too, instead of the mean of all individuals to depict the mean and the variation of each size class. I would also include information on the different polymorphs of each species. In general, I have no problems with the figures being black-and white only, but personally, I would try to improve the visual representation. In this case, maybe increasing the figure size to double and placing the legend within the top right corner could help separate a bit more out the different elements. Also, this is a detail, but to make it more intuitive, the grey filled symbols could be the aragonitic species, empty symbols the calcitic and half-filled for example bimineralic.

Response: These comments have been taken into account and the figures have been improved.

Figure 3: What is the x-axis? Please make the y-axis similar where possible, this is really difficult to read for me. Also, the information on the differences between size classes should be removed as at the moment there is too much information in this figure. The individual panels are missing sub-headings that should be also referred to in the manuscript text.
Response: This figure has been simplified to be easier to read.

Figure 4: Please appropriately label all panels as 'a,b,c, etc.' What do you mean by 'raw data as black dots'? (I see blue dots.) Please include polymorphs, analytical uncertainty, indicate the sizes for each category. Maybe better to put each species in a separate row. Why some size classes have values in between the size class number categories?
Response: These comments have been taken into account and the figures have been improved.

Figure 5: I find this figure difficult to follow, maybe there is a better way to illustrate the message? Should be 'dashed line' instead of 'broken line'. Why are some panels darker? Please specify in the caption.
Response: These comments have been taken into account and the figures have been improved.

References used in the above responses to the reviewers' comments:

Barker, S., Greaves, M., and Elderfield, H.: A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry, Geochem. Geophys. Geosyst., 4, 8407, https://doi.org/10.1029/2003GC000559, 2003.

Bourget, E.: Barnacle shells: composition, structure and growth, in: Barnacle biology, edited by: Southward, A.J., A.A.Balkema, Rotterdam, The Netherlands, 267–285, https://doi.org/10.1201/9781315138053-14 1987.

Holcomb, M., DeCarlo, T.M., Schoepf, V., Dissard, D., Tanaka, K., and McCulloch, M.: Cleaning and pre-treatment procedures for biogenic and synthetic calcium carbonate powders for determination of elemental and boron isotopic compositions, Chem. Geol., 398, 11–21, https://doi.org/10.1016/j.chemgeo.2015.01.019, 2015.

Inoue, M., Nohara, M., Okai, T., Suzuki, A., and Kawahata, H.: Concentrations of trace elements in carbonate reference materials Coral JCp-1 and Giant Clam JCt-1 by inductively coupled plasma-mass spectrometry. Geostand. Geoanal. Res. 28, 411-416, https://doi.org/10.1111/j.1751-908X.2004.tb00759.x, 2004.

Loxton, J., Najorka, J., Humphreys-Williams, E., Kuklinski, P., Smith, A. M., Porter, J. S., Spencer Jones, M.: The forgotten variable: Impact of cleaning on the skeletal composition of a marine invertebrate, Chem. Geol., 474, 45–57, https://doi.org/10.1016/j.chemgeo.2017.10.022, 2017.

Mannella, G., Zanchetta, G., Regattieri, E., Perchiazzi, N., Drysdale, R. N., Giaccio, B., Leng, M. J., and Wagneret, B.: Effects of organic removal techniques prior to carbonate stable isotope analysis of lacustrine marls: A case study from palaeo-lake Fucino (central Italy). Rapid Commun. Mass Spectrom., 34, e8623, https://doi.org/10.1002/rcm.8623, 2020.

Marin, F. and Luquet, G.: Molluscan shell proteins, Compt. Rend. Palevol, 3, 469–492, https://doi.org/10.1016/j.crpv.2004.07.009, 2004.

Rainbow, P. S., Wolowicz, M., Fialkowski, W., Smith, B. D., and Sokolowski, A.: Biomonitoring of trace metals in the Gulf of Gdansk , using mussels (*Mytilus trossulus*) and barnacles (*Balanus improvisus*), Water Res., 34, 1823–1829, https://doi.org/10.1016/S0043-1354(99)00345-0, 2000.

Rainbow, P. S., Fialkowski, W., Sokolowski, A., Smith, B. D., and Wolowicz, M.: Geographical and seasonal variation of trace metal bioavailabilities in the Gulf of Gdansk, Baltic Sea using mussels (*Mytilus trossulus*) and barnacles (*Balanus improvisus*) as biomonitors, Mar. Biol., 144, 271–286, https://doi.org/10.1007/s00227-003-1197-2, 2004.

Rueda, J. L. and Smaal, A. C.: Variation of the physiological energetics of the bivalve *Spisula subtruncata* (da Costa, 1778) within an annual cycle, J. Exp. Mar. Bio. Ecol., 301, 141–157, https://doi.org/10.1016/j.jembe.2003.09.018, 2004.