INTERNATIONAL COURT OF JUSTICE

WHALING IN THE ANTARCTIC
(AUSTRALIA v. JAPAN: NEW ZEALAND INTERVENING)

Response to

"Scientific review of issues raised by the Memorial of Australia including its two Appendices" by Professor Lars Walløe [9 April 2013]

Marc Mangel

University of California

Santa Cruz

31 May 2013

1. INTRODUCTION

1.1. In preparing this analysis responding to the statement of Professor Lars Walløe dated 9 April 2013, with limited exceptions I will not repeat material that is in my Original Expert Opinion or Supplementary Expert Opinion.

1.2. I do not respond to all the views expressed by Professor Walløe and the absence of a comment from me should not be taken as agreement. I concentrate on the following:

- Professor Walløe's assertions that: i) the criteria in the Original Expert Opinion are not applicable in the Southern Ocean; and ii) that general or vague hypotheses are sufficient;
- The two specific examples (the work of Gregor Mendel and the effects of acid rain in lakes and streams in Norway) that Professor Walløe offers as identifying research undertaken without hypotheses;
- Professor Walløe's views regarding data mining and exploratory data analysis in the context of science; and
- Professor Walløe's views regarding determination of sample size.

1.3. When I quote from Professor Walløe's Report I do so by reference to the 'LW' page number and to the relevant paragraph on that page (my numbering of paragraphs includes part paragraphs at the beginning of a page).

2. SCIENCE REQUIRES HYPOTHESES

2.1. Professor Walløe states that the description of science in my Original Expert Opinion is "too restrictive" (LW pg 5, para 2). He suggests that the description is "perhaps … adequate" for "research in a fairly advanced biological field in which there are generally accepted hypotheses about the main functional connections in the system under investigation" (LW pg 5, para 3). However, in his view, the description is not appropriate where "[e]xisting knowledge … is very limited" (LW pg 5, para 4) such as for work in the Southern Ocean. In such a situation, according to Professor Walløe, scientific research can be carried out on the basis of hypotheses that are general and often vague (LW pg 5, para 5). However, Professor Walløe provides no authorities for these propositions.

2.2. I disagree with Professor Walløe that the criteria outlined in my Original Expert Opinion are inapplicable if there is a limited existing knowledge of the biological field being studied. In such fields, it is perhaps even more important for scientific research to proceed on the basis of clear and testable hypotheses, which have been carefully articulated and can be evaluated. Professor Walløe's assertion that general or vague hypotheses will suffice is not consistent with accepted scientific method.

2.3. Merely collecting data is not research for scientific purposes. As I noted in my Original Expert Opinion, "[d]escription is not tantamount to understanding: descriptive data cannot by themselves furnish an explanation of the mechanisms behind the observations, nor can they easily identify the processes that brought about the situation described. Complicated descriptions can become goals in themselves and may delude us into thinking progress has been made" (Valiela 2001, pg 11). The notion that we can simply go out and collect data or 'observe' is not scientific; doing so is meaningless from a scientific perspective since it is impossible to observe without having first reflected or thought about the purpose for which the observation is required. Science is not a bucket of data.

2.4. Professor Walløe offers no alternative definition of science or the scientific process to the one I have put forward, and does not explain which of the components referred to by me he would drop. In a program for 'purposes of scientific research', a conceptual framework, the correct empirical and statistical tools to answer a question and peer-review – the very foundation of the modern and consensual nature of science – are all required.

2.5. The conceptual framework brought to a particular problem depends upon the current understanding of the biological system being studied. Simply studying something because we do not know about it with the hope that some insight or understanding might emerge is not sufficient to characterize the activity as science. I cannot imagine a journal or funding panel would publish a paper or recommend a project for which there was no conceptual framework.

2.6. Professor Walløe does acknowledge that "there are always general hypotheses behind any collection of primary data … [h]owever, these underlying hypotheses are often vague and not easy to formulate in scientific language" (LW page 5, para 5). Professor Walløe then proceeds to state that "[t]he research carried out under the JARPA and JARPA II programmes includes both data collection to test specific hypotheses and collection of data to provide background primary data … which may be valuable in the future." However, Professor Walløe does not indicate what some of the supposed general or specific hypotheses in JARPA or JARPA II are – either in scientific language or even vaguely formulated in non-scientific language. Nor does he provide any assistance as to what data are considered to be of hypothetical future value; we are left to speculate on all these matters.

2.7. The collection of data without a guiding hypothesis or conceptual framework has at least three problems, each of which reveals why it does not amount to scientific research. First, even data that are collected speculatively need to be justified as to why some particular data items were to be included, and others excluded. Without something to guide the decisions on what data to collect, the outcome will be arbitrary. Second, using previously collected data for a new purpose often proves to be problematic. The experience of many scientists when trying to use collections of existing data is that some piece of information critical for the evaluation of a *post hoc* hypothesis is missing. Third, *post hoc* hypotheses do not represent new knowledge without corroboration by new observations and new data. This is because a *post hoc* hypothesis is itself formed based on the data set already collected and therefore by definition is supported by that data. It therefore cannot be tested properly by the data from which it was formed, but rather must be tested against new observations and new data.

3. THE EXAMPLES OF MENDEL AND ACID RAIN

3.1. Professor Walløe offers two examples that he claims support the proposition that collecting data absent a conceptual framework can be considered science. However, closer reading of each example offered by Professor Walløe shows that they are both, in fact, clear examples of data being collected within a conceptual framework.

Mendel

3.2. Gregor Mendel is considered to be a scientific genius and the founder of genetics. It is true that Mendel collected a considerable amount of data but he did so within a clear conceptual framework and with a specific objective in mind. A variety of theories of inheritance (i.e. hypotheses) were prevalent in the late 19th century and Mendel set out to test these hypotheses (Allen 2004, pg 65ff; Deichmann 2010, pg 98; Gliboff 1999, pg 225).

3.3. Orel (1996) noted that "Mendel may have found inspiration in the physics textbooks of his teachers at Vienna University … The model of discrete pairs of traits was his *initial theoretical framework*" (pg 162, italics in the original). That is, Mendel began with a conceptual framework, which he modified as he collected data (as a program for 'purposes of scientific research' does): he did not begin with process of data collection absent any conceptual framework, and did not proceed to formulate hypotheses only after the collection of data. Mendel developed up to nine hypotheses, with experiments (not random data collection) to test each of them as he worked (Orel, Figure 5.13, pg 162).

3.4. Mendel's approach is the very antithesis of the approach of "data collected without any specific hypotheses in mind" (LW pg 6, carryover para) that Professor Walløe ascribes to Mendel's work.

Acid Rain

3.5. Mason (1990a) provides an excellent overview of the Surface Waters Acidification Programme (SWAP) to which Professor Walløe refers, and the papers in the

Mason (1990a) publication offer detailed insights of the program. To further understand the effects of acid precipitation on forests and fish, SWAP was undertaken through a collaboration of three national academies of science (whose members comprised a steering committee), for a five-year duration, with completely independent publication of results, and based on a set of four focused questions (Mason, 1990b).

3.6. Professor Walløe indicates that SWAP consisted of random collection of data until the culprits (in particular aluminum) that caused the death of fish in streams were discovered; and that scientists "were searching for a possible unknown factor which could explain the death of the fish" (LW pg 6, para 2). However, Morris and Reader (1990) in their contribution to Mason (1990a) noted that the lethal effects of aluminum on fish had been known for at least a decade and the measurement of inorganic aluminum concentrations was included in the SWAP integrated research program from its inception (Mason 1990). Thus scientists had a clear hypothesis – that inorganic aluminum might be having a lethal effect on the fish; what they did not know was the precise mode of action by which aluminum had its effects. This is far from mere data collection.

3.7. At the end of the program, when the mode of action was understood, Muniz and Walløe (1990, pg 337) stated "[a]s far as pH and inorganic aluminium are concerned, the results are not surprising and corroborate earlier results both from the field and laboratory". This is entirely different from the random search of data that Professor Walløe described (LW, pg 6-7, last line, carryover para).

3.8. With this example, Professor Walløe has described an excellent model for environmental research. This model includes hypotheses that are clearly stated, comprehensive and focused research programs of finite duration, and involving several disciplines and many different institutions. It is indeed the antithesis of JARPA and JARPA II.

Summary

3.9. In summary, a closer look at the work of Gregor Mendel fully refutes Professor Walløe's suggestion that Mendel worked in the absence of a conceptual framework. Similarly, in the example of acid rain, Professor Walløe described a program of research that differs from JARPA II in almost every important characteristic. It is true that in science we sometimes collect large amounts of data to investigate hypotheses. However, this is properly undertaken within a conceptual framework and it does not mean that collecting large amounts of data in itself – that is, without the conceptual framework – is science.

4. DATA MINING IS NOT SCIENCE

4.1. Professor Walløe writes that “[t]oday powerful computer programs exist that can be used for such ‘exploratory data analysis’, or ‘data mining’, as it is sometimes called” (LW pg 7, para 1) and implies that this turns mere data collection into a program for ‘purposes of scientific research’.

4.2. Data mining has developed in recent years because of the advances in computing technology and uses computer programs to seek patterns and relationships in large volumes of data (Clifton 2010). The basic idea is that the computer programs will scan large volumes of data, and thereby discover relationships within the data.

4.3. However ‘data mining’ can quickly turn into ‘data dredging’, in which the computer programs ‘discover’ misleading relationships in the data. The error occurs because researchers do not form a hypothesis beforehand and thus search for combinations of variables that might show some relationship. When many such combinations are tested by statistical methods, some combinations will turn out to show a relationship or trend purely by chance and researchers are mislead into believing they have discovered a relevant hypothesis *post hoc* when none in fact exists (Davey Smith and Ebrahim 2002). That is, a pattern might appear from the data that does not actually reflect any real phenomenon. Exploratory data analyses are often called ‘fishing trips’ – i.e. one is fishing through the data hoping to find something interesting.

4.4. Davey Smith and Ebrahim (2002, pg 1438) discuss data mining in human epidemiology and note that after the fact it is generally easy to find a plausible explanation for the observed relationship or trend in the data, even if it is not real. Further, they note that standard statistical techniques are not very good at correcting errors arrived at in this way. This shows the inherent risks in data mining and why it is no way to run a program for ‘purposes of scientific research’.

4.5. The foundational goals of statistics have not changed due to modern computation. Rather, our ability to implement them has. Most exploratory data analyses do not lead anywhere meaningful, and do not contribute to scientific knowledge or understanding. Since there is a tradition in science not to publish non-results, it is difficult to estimate the frequency with which exploratory data analyses are successful. In my own experience, very few (if any) exploratory analyses have yielded important insights. If scientists do not know how the data will be analyzed, they are not ready to collect it.

4.6. Simon et al (1987, pg 47) in a volume on scientific discovery put it simply: “[s]cientific discoveries seldom, if ever, emerge from random, trial-and-error search”. In the case of JARPA and JARPA II one may also ask: at what point should the exploration component of exploratory data analysis stop? JARPA II is indefinite in duration; its exploratory data collection could go on for many more decades. I am not aware of any scientific research bodies that would support the approach of exploration lacking a conceptual framework going on for decades.

5. SETTING SAMPLE SIZE

Statistical Basis

5.1. With respect to setting sample size, Professor Walløe writes that I am asking for "an exact answer to the wrong question" (LW pg 8, para 2). However, he has not identified either the right question or how to obtain an answer to it – whether it be approximate or exact. Increasing sample sizes in medical clinical trials, in which the objective is to save lives, is fundamentally different than increasing sample size for what Professor Walløe describes as "precautionary" reasons in JARPA or JARPA II.

5.2. Professor Walløe appears to suggest that since criteria for setting sample size are difficult to apply in practice, one can simply forego using them. I disagree – a program for 'purposes of scientific research' requires transparency and clarity in setting sample sizes. In this respect, Professor Walløe appears to agree when he writes "it must be admitted that the Japanese scientists have not always given completely transparent and clear explanations of how sample sizes were calculated or determined" (LW pg 10, para 2). I concur.

5.3. Professor Walløe and I also agree about the selection of the ultimate sample size when analysis suggests a variety of possible choices – "[t]he final decision about sample size would then have to be the largest of the different sample sizes determined for each hypothesis" (LW pg 9, para 2; Supplementary Expert Opinion paras 3.15-3.18).

Funding

5.4. Professor Walløe writes that "Japan has chosen to cover part of the costs of its whale research programmes by selling whale products on the commercial market. To obtain sufficient income in this way, the yearly catch has to be of a certain magnitude" (LW pg 9-10, carryover para). Professor Walløe further states "on reading the research proposals for JARPA and JARPA II submitted to the IWC Scientific Committee, I often had the impression that sample sizes were also influenced by funding considerations" (LW pg 10, para 2).

5.5. He in effect confirms that the setting of sample sizes in JARPA and JARPA II is driven by non-scientific considerations. Whether there is sufficient funding for a research program is not a scientific question but a matter of national priorities for the country engaged in the activity.

5.6. To my knowledge, almost all of the other large-scale marine research programs in the Southern Ocean are conducted without any income derived from the research. These generally involve one ship (as in the US Antarctic program in which I am involved) and on occasions two to three ships (e.g. IDCR/SOWER). There are no programs that I am aware of that operate annually with as many vessels as JARPA or JARPA II. The major reason for the scale of this fleet appears to be that it is a lethal program and requires a factory ship and a major re-fuelling vessel. A non-lethal program could operate at a significantly smaller scale. Thus, Professor Walløe's assertion that it would be impossible to carry out a major research program in the Southern Ocean without income derived from killing animals is contradicted by other research programs undertaken there.

6. CONCLUSION

6.1. Professor Walløe concludes: "As long as an activity is genuinely motivated by an intent to conduct scientific research, other additional motivations, e.g. obtaining some of the funding by selling products, may even be regarded as an advantage and not as a counterargument" (LW pg 10, para 4). However, to follow Professor Walløe's own logic, one must reason that if a program lacks a conceptual framework, clarity in how sample sizes are collected, and *bona fide* peer-review, it is difficult to conclude that it "is genuinely motivated by an intent to conduct scientific research". Once more, the conclusion reached in both of my earlier reports remains unchanged – although JARPA II is a program of data collection, it is not for 'purposes of scientific research'.

7. LITERATURE CITED

Allen, G.E. 2003. 'Mendel and Modern Genetics: The Legacy for Today'. *Endeavour* 27:63-68.

Clifton, C. 2010. *Encyclopædia Britannica: Definition of Data Mining* (http://www.britannica.com/EBchecked/topic/1056150/data-mining).

Davey Smith, D., and Ebrahim. S. 2002. 'Data Dredging, Bias, Or Confounding'. *British Medical Journal* 325:1437-1438.

Deichmann, U. 2010. 'Gemmules and Elements: On Darwin's and Mendel's Concepts and Methods in Heredity'. *Journal of General Philosophy of Science* 41:85-112.

Gliboff, S. 1999. 'Gregor Mendel and the Laws of Evolution'. *History of Science* 37:217-235.

Mason, J. (ed) 1990a. *Surface Water Acidification Programme*. Cambridge University Press, Cambridge, UK.

Mason, J. 1990b. The rationale, design and management of the Surface Waters Acidification Programme. In Mason (1990a) pp1-8.

Morris, R., and Reader, J. P. 1990. The effects of controlled chemical episodes on the survival, sodium balance and respiration of brown trout. In Mason (1990a) pp357-68.

Muniz, I. P., and Walløe , L. 1990. The influence of water quality and catchment characteristics on the survival of fish populations. In Mason (1990a) pp327-39.

Orel, V. 1996. *Gregor Mendel. The First Geneticist*. Oxford University Press, Oxford and New York.

Simon, H.A., Bradshaw, G.L., and J.M. Zytkow. 1987. *Scientific Discovery*. The MIT Press, Cambridge, MA.

Valiela, I. 2001. *Doing Science. Design, analysis, and communication of scientific research*. Oxford University Press, New York.