DRAFT
Do not cite or quote.

HOW DO SCIENTISTS LEARN TO CONDUCT SCIENCES FOR SUSTAINABILITY? AN ANALYTICAL FRAMEWORK AND THE IMPORTANCE OF CONTEXT

Atsushi Ishii*
Ayako Okubo[a]

1. Introduction

There is no doubt that scientific knowledge is playing an important role in environmental diplomacy and ultimately in the pursuit of sustainability. It can guide us by pointing out the cause-and-effect relationship, the environmental impacts, the policy options, and the effect of those options to environmental problems in concern. In order to obtain such scientific advice, negotiators usually institutionalize scientists through scientific assessments. However, this is hardly an easy institutionalization; advisory scientists may face confusion especially when first entering the hybrid domain of science and politics because they have been trained to be residents of the 'Republic of Science' (Polanyi, 1964) and the Republic's norms, rules and structure are different from those of the hybrid domain. Moreover, the domains of science and politics have inherently tense relationship because their norms, rules, and interests are often incompatible and their boundaries are essentially contested (Jasanoff, 1987). To accommodate this tension, advisory scientists may engage in learning endeavor and become able to adapt to different norms, rules, and structure of the hybrid domain. Besides the need to improve assessment procedures and thereby enhance their effectiveness in policy advice through learning, this also constitutes the importance of learning in scientific assessments and provides one of the bases of effective scientific assessment.

This paper basically shares the research questions of Siebenhuner (2002a,b): "How did the assessments at hand learn over the years and in the different phases of the assessment process? … How could the learning process be characterized …? Which … factors did influence learning in assessments? Which conclusions can be drawn to improve learning processes in assessments in general?" To answer these questions, we will first discuss the conceptual framework developed by Siebenhuhner (2002a). Then, we develop an alternative framework on the basis of that discussion. In doing so, we draw particularly on literature on organizational learning (as same as Siebenhuhner (2002a,b)) and policy learning. Thirdly, we apply it to the case of the international whaling regime to demonstrate the usefulness of the alternative framework and to answer some of the research

* E-mail: ishii@cneas.tohoku.ac.jp; Telephone: +81-(0)22-795-6076; Postal Address:
Center for Northeast Asian Studies, Tohoku University
41, Kawauchi, Aoba, Sendai, Miyagi, JAPAN 980-8576.
[a] Postal Address: Kaiyō Senpaku Building, 1-15-16, Toranomon, Minato-ku, Tokyo, JAPAN 105-0001.

questions raised above.

Needless to say, we can only provide answers reflecting the context of the international management of whaling. Yet the whaling case is suitable for our purpose because the scientific assessment conducted in the case has a long history providing “opportunity for learning from past experience and from other assessments” (Siebenhuner, 2002a). It is very interesting in terms of the scientific assessment’s effectiveness as well because the latest management procedure for sustainable whaling, namely the Revised Management Procedure (RMP), even survived one of the most emotional and highly confronted international negotiations, namely the International Whaling Commission (IWC), and adopted by the parties as the single official procedure (but not implemented yet). Our case specific argument would be that it is the learning process that enabled the IWC’s Scientific Committee (SciCom) to develop a management procedure robust enough to address the bipolar concerns of the anti- and pro-whaling countries. Our alternative analytical framework enabled us to characterize the learning process and identify the stimulating factors of such process.

2. Revising Siebenhüner’s analytical framework

2.1. Examination of Siebenhuner’s framework

Siebenhuner starts with defining ‘scientific assessment,’ drawing on the Global Environmental Assessment Project (GEA), as “the entire social process by which expert knowledge related to a policy problem is organised, evaluated, integrated, and presented in document to inform policy or decision-making” (GEA, 1997, p. 53). The important point in the definition is that it avoids *production* of expert knowledge. Thus, Siebenhuner seems to view “the production and validation of knowledge as largely independent of the use of the knowledge in policy decision” (Jasanoff and Wynne, 1998). This is reflected in his distinction between procedural and substantive knowledge as well (Siebenhuner 2002a, p. 412). However, part of the literature of science and technology studies consistent with constructivist accounts suggest that “scientific knowledge and political order are *co-produced* at multiple stages in their joint evolution, from the stabilization of specialized factual findings in laboratories and field studies to the national and international acceptance of causal explanations offered by science and their use in decisionmaking” (Jasanoff and Wynne, 1998). Co-production is clearly exhibited in the case of the Convention on Long-Range Transboundary Air Pollution (LRTAP) which Siebenhuner applies his framework; the integrated assessment model developed under the LRTAP, namely the RAINS model, was co-produced by the modelers and policymakers who were expected to use the model (Alcamo et al. 1990). Our case study also suggests that knowledge production should be an integral part of the analytical framework of assessments’ learning endeavor to better account for the effectiveness of assessments. Therefore, we incorporate knowledge production to the GEA definition of scientific assessments.

Another fundamental question would be: What is to be learned in assessments? As already mentioned, Siebenhuner distinguishes procedural from substantive knowledge and focuses only on the former. We maintain our argument that this distinction should be avoided for the reason provided above. Moreover, he sets three criteria to "indicate the direction in which learning in a qualified sense could take place" (Siebenhuner, 2002a): saliency, credibility, and legitimacy. These criteria are in line with the GEA work and are argued to be necessary conditions to design influential scientific assessments in decision-making processes. He summarizes as "scientific assessments learn when they change the way the assessment is conducted in order to become more salient, credible and legitimate or when participants acquire general abilities to conduct more salient, credible and legitimate assessments which are founded on changes in knowledge and beliefs" (Siebenhuner, 2002a, p. 413). However, there are other criteria that have been proposed for the same end (for example, Jasanoff (1995) and Ishii (2001a,b; 2004)). More problematic is that he focuses only on beneficial learning and not on detrimental learning (to learn in the direction of diminishing effectiveness of assessments), despite that, in order to enhance effectiveness of assessments, avoiding detrimental learning is equally important as facilitating beneficial learning.

Closely related to the 'what?' question above and equally important is the typology of learning modes. They are distinguished according to the direct effects on organizational behavior resulting from learning and thereby enable us to detail the overall cause-and-effect relationship between learning and effectiveness of assessments. Siebenhuner employs the topology of Argyris and Schon (1996) developed in the field of organization learning: Single-loop, Double-loop, and Deutero learning. However, each learning theory has "different origins and describe different aspects of the learning process and it is important to note the areas to which they apply and those to which they do not if a usable set of concepts for theory construction and evaluation is to be maintained"[1] (Bennett and Howlett, 1992, p. 278), and Siebenhuner does not justify that Argyris and Schon's conception fits into the context of scientific assessments. This is particularly important since scientists belong to 'the Republic of Science' which has significant contextual differences in terms of norms, rules, and structure compared to most organizations (e.g. corporations) studied by organizational learning scholars. This difference is reflected in our alternative framework.

Turning to the factors influencing learning bahavior, Siebenhuner sets contextual, cultural, personal, and structural factors as important factors. While we agree to most of them, we think it is more appropriate to distinguish between endogenous and exogenous factors, and to incorporate process-dependent factors. As for the former, this dichotomy is largely employed in the learning literature (e.g. Brown and Kenney, 2006). Siebenhuner (2002a) assigns exogenous factors to contextual ones, but there are endogenous contextual factors as well. For instance, our case study

[1] This was argued citing Sartori (1968).

shows that one of the critical factors stimulating learning was the internal factor reflecting the bipolar context of the whaling debate; advisory scientists were somewhat divided between precautionary and pro-whaling scientists, and this division should not be interpreted as a personal but as a contextual factor of the relationship among scientists. As for the process-dependent factors, we suggest to incorporate individual/collective perceptions of assessment participants about the success or failure of past operations of the assessment in question because it "is essential for productive organizational learning" (Brown and Kenney, 2006, p. 7). While the details are given below, it can be reasonably assumed that significant failure may stimulate learning by incentivizing assessment participants to recover the failure or impede learning by depriving their confidence. Our case study clearly exhibits the former case.

2.2. An alternative conceptual framework

With espousing Siebenhuner's justification of analyzing learning as a collective behavior (Siebenhuner, 2002a, p. 413), we first begin with defining the learning endeavor of scientific assessments alternatively to Siebenhuner's. We draw on the definition of social learning provided by Hall (1993) and adapt it to the context of scientific assessments: learning of scientific assessment is a deliberate attempt to adjust the goals or methodology of scientific assessment in response to past experience and relevant knowledge. It is important to note that this does not exclude the possibility of detrimental learning nor presume the objective of scientific assessments. This definition automatically answers the 'what to learn' question: past experience and relevant knowledge. It can capture more diverse cases of 'learning assessments' than Siebenhuner's, which we think is essential for the study of 'learning assessments' to provide more generalizable contribution and deeper understanding of scientific assessments because the field is in infant stage yet.

We now turn to learning modes, where we also follow Hall (1993) because it is compatible with the definition of learning suggested above and fits into the context of scientific assessment well. Hall (1993) establishes three orders of learning regarding policy change: first, second and third order learning. First-order learning occurs when policy instrument settings are changed in the light of experience and relevant knowledge, while overall goals and instruments of policy remain the same (Hall 1993, p. 278). Second order learning changes the policy goals and instruments (Brown and Kenney, 2006, p. 6). Third order learning requires change in overall goals and instruments of policy according to a shift in worldview or a paradigm shift in the organization in question (Hall 1993). These could be easily adapted to the context of scientific assessment by changing the words "policy" and "organization" to "scientific assessment". For more clarity and direct reference to the traits of the learning modes, we name those first to third order learning as *adaptive learning*, *reformative learning*, and *paradigmatic learning*, respectively.

Hall (1993) conceptualizes third order learning referring to Thomas Kuhn's concept of scientific paradigm. This makes Hall's overall ordering of learning especially suitable for applying it to scientific assessments because it is very likely that scientific activities are required to shift the paradigm of normal research science to produce scientific advice relevant to policy. Jasanoff (1995) argues that advisory science for regulatory decision-making has a different mode from research science: regulatory science. Though in its early stage, Ishii (2001; 2004) developed the conceptual model of 'diplomatory science' which is the mode for advisory science directly involved in environmental diplomacy. Double-loop learning which Siebenhuner employs in his framework does not distinguish changes in objective from changes in values, norms, and belief systems that can lead to paradigm shift. In other words, changes in objectives with and without paradigm shift cannot be clearly distinguished by applying the single- and double-loop learning typology. Besides that, deutero learning, which is the metal-level mode of 'learning how to learn,' will be retained but named as 'fourth order learning' in line with Hall's wording (see Brown, 2006, p. 25).

For the factors affecting learning, we employ those of Siebenhuner's framework with some addition. It is easily imaginable that the learning endeavor of scientific assessments could be affected, sometimes significantly, by collective and individual perception of the performance of past operations as failure or success. Siebenhuner's framework does not incorporate these factors of 'the shadow of the past,' therefore we add the following *process-dependent factors* to our alternative framework: evaluative perception of the past operations of the scientific assessment itself; of the relationship between science and politics related to past operations of the assessment; and, of the policy performance of the scientific advice provided by the past operations of the assessment. Indeed, "[m]ost analysts agree that organizations are more likely to learn as a consequence of failure than of success" (Brown and Kenney, 2006, p. 7), but "profound failure and crises may have particularly negative consequences for productive learning" (Brown and Kenney, 2006, p. 7). However, these arguments are yet to be tested empirically with case studies of 'learning assessments,' and this underscores the importance of incorporating evaluative perception of past operations into the alternative framework. Moreover, while Siebenhuner incorporates reflective mechanisms into his framework, we emphasize that evaluative perception at both the individual and collective level could be a direct influence to learning behavior without any reflective mechanism institutionalized.

Another addition relates to the aforementioned division between endogenous and exogenous contextual factors. While exogenous factors would be the contextual ones Siebenhuner submits, we submit *conflict/cooperation between research fields related to assessments*, *relationship between science and politics in the assessment's context*, and *new scientific findings* as the endogenous factors. As for the first two elements, scientific assessments mostly involve interdisciplinary research and interaction between science and politics, and therefore cannot escape the context of relationship between the research fields in question and between science and politics. Their conflictual relationship may inhibit learning by disfunctionalizing leadership structures, and/or by preventing the establishment of informal communication networks which Siebenhuner sets as one of the cultural factors facilitating learning. However, even in this conflictual situation, there is some possibility that conflict sparks off paradigmatic learning, which changes the whole conduct of assessment including problem framing and methodologies of assessment, to ease or avoid tense relationship. If their relationship is of cooperative nature, it can generally lead to productive learning (Argyris, 1999, pp. 84-87). As for the element of *new scientific findings*, this inclusion reflects our definition of scientific assessments that they not only organize, evaluate, integrate, and present in document but also produce knowledge to inform policy or decision-making.

As argued above, limiting learning to only when it results in enhancing effectiveness of assessments should be avoided. Instead, we propose to identify cases of learning without focusing on the normative evaluation of the consequences of learning and then to evaluate the effect of the changes resulting from learning with the effectiveness index adjusted from those proposed by Underdal (2000):

Level 1. Decision-makers recognize only the *existence* of the output of scientific assessments[2]

Level 2. Decision-makers recognize the *relevance* and *usefulness* of the kinds of knowledge that scientific assessments produce, and look to the relevant scientific community for information, models and theories

Level 3. Decision-makers accept as *valid* or *tenable* the substantive conclusions that meet the standards of the relevant scientific community itself

Level 4. Decision-makers accept not only factual conclusions but also what might be called the *'policy implications'* of these conclusions, and respond positively to more explicit advice offered by the relevant scientific community

It should be noted that this scale of effectiveness levels "can be considered cumulative in the sense that higher levels can be reached only through the preceding step(s)" and that "at each level adoption is a matter of degree" (Underdal, 2000, p. 9). This set of index is applicable to any cases. If the learning endeavor results in change that enhances the assessment's effectiveness, we call it *beneficial learning*, and, if opposite, *detrimental learning*.

Consistent effectiveness index allows us to compare case studies on a common ground to produce generalizable knowledge about the learning assessments including: what are the conditions for productive/detrimental learning?; are the conditions for productive and detrimental learning symmetrical?; are there any conditions that is proportional to the effectiveness index?; and, are there any certain types of detrimental learning that may constitute a necessary condition to productive learning? Siebenhuner's framework makes these important research questions unanswerable.

[2] This level of effectiveness was added by the authors.

3. A brief case study of the international whaling regime

There is no need here to repeat the history of the IWC which has swung between two extremes: from the tragic overexploitation of cetaceans to the transformation of IWC into a preservationist regime. However, because of this stalemate, the scientific assessment by the Scientific Committee (SciCom) of the IWC to establish a robust management procedure has attracted little attention. In this very brief case study, we apply the above framework to the series of management procedures which the SciCom has established and implemented so far.

The first management device of the IWC was the Blue Whale Unit (BWU). The objective of this scheme was obviously to manage the price of whale oil through measuring the oil production (and not the whale resources). This incentivized the whale industry to hunt selectively the biggest cetacean in the world —blue whale— in pursuit of catch effort efficiency. This led to the so-called 'Whaling Olympics' which turned bigger whales to be endangered species. The Scientific Committee recommended reducing the quota expressed in terms of BWU. However, the first reduction of quota was done when the catch could not reach the BWU quota because of scarce resources. At this time, the management objective was the detection of endangered species by trying to count the whales in the real world.

The BWU was abandoned in 1972 when the first proposal of moratorium on commercial whaling was made by the US. However, the proposal was refuted by the SciCom as having no scientific basis and had not been adopted before 1982. The compromise in 1974 was to establish a New Management Procedure (NMP). The NMP reflected new thinking developed in the fisheries management such as that the target of resource abundance set to the Maximum Sustainable Yield (MSY). Although not explicitly stated, this was in a sense more precautionary than the previous BWU because its objective is to optimize whale catch so that the catch is maximized under the condition of no depletion of whale stocks. However, the actual calculation remained to aim at finding the MSY by reproducing the real biological system of whales. In the beginning, the scientific advice was relatively well accepted and implemented by the policymakers, However, soon after the application of the NMP to actual whale stocks, this positivistic approach caused problems; as some of the scientists became more precautionary and some remained as pro-whaling scientists, it became more and more difficult to agree on the actual situation, especially between those two sides. The former scientists argued for more research and not producing any advice because of uncertainty, and the latter scientists argued for producing practical advice. The attempt to adjust the NMP failed roughly on the same grounds.

The NMP could be characterized as reformative learning because the objective and accordingly the method changed while the paradigm of positivism remained the same. This was of course partly motivated by the disastrous outcome of the 'Whaling Olympics', which is a factor that could be categorized as the perception of the past performance of past conduct of assessment. The effectiveness of the NMP on the policymaking process is at Level 4 at least for the initial period. Moreover, the attempt to 'band-aid' the NMP could be characterized as adaptive learning with having the effectiveness index of Level 1.

After some silent period in the SciCom alongwith the implemented moratorium on commercial whaling, the members of the SciCom began to seek for a Revised Management Procedure (RMP) in the Comprehensive Assessment mandated in the decision of the moratorium. The scientists soon agreed on a minimum data approach reflecting the poor data availability and to take uncertainty into account by investigating robust parameters. Based on that understanding, five teams competed with and learned from each other to develop the management procedure. A simulation technique using a model that was only assumed to be reflecting the real world was adopted to test candidate management procedures with the same performance criteria. The philosophy of the whole process is neatly described by one of the developers (Justin Cooke):

> The approaches developed here treat the MSY and MSY level as purely notional quantities used to determine catch limits. The catch limits are derived in a direct mechanical fashion from the data obtained: there is no scope for argument about the biological correctness or otherwise of the parameter values used, nor about whether the catch limits set by the procedure are justified by the data available. The procedures even set catch limits in cases where the only information available is an estimate of absolute stock size, without any information on the sustainable or replacement yield. The main criterion for choosing the parameter values in the models used is not the relative biological plausibility of the values chosen, but the consequences of adopting them even if they turn out to be wrong.
>
> The catch limits set in individual years by the procedure of the kinds presented here are not claimed to be based on specific scientific evidence, and are justified only in the context of the performance of the given procedure as a whole. Even if the procedure performs satisfactorily on average, there will be years when catch limits are 'unnecessarily' raised or lowered. The procedures involve the deliberate abandonment of the conventional requirement that the need for each specified management action be scientifically demonstrated before any such action can be taken. It may be that this is a necessary price to pay if sustainable exploitation of whale stocks is to become a feasible objective." I think that this statement is clearly demonstrating the epistemological shift of scientific credibility in order to achieve the diplomatic aim and to take the diplomatic context into scientific assessment.

Finally, the procedure developed by Cooke was adopted by consensus within the SciCom and recommended for adoption as the official management procedure. While it has not been implemented yet, the procedure was adopted by the polarized IWC with almost no criticism.

The RMP is clearly a product of paradigmatic learning. The paradigm changed from positivism to management-oriented science. Accordingly, the objective changed from optimization based on MSY to sustainable whaling feasible even under significant uncertainties, and the methodology from biologically plausible models to simulation. The significant explanatory factor is the experience of failure in the NMP applications. Despite of the paradigmatic learning, the effectiveness index reaches not the highest level but Level 3.

4. Conclusion

We have applied the alternative framework to the whaling issue and demonstrated that it can provide a more holistic and detailed explanation of the learning endeavors of scientific assessments. Needless to say, the alternative framework should be re-evaluated and updated by further case studies; exploring the learning process is itself a learning process.

References

Argyris, C. (1999) *On Organizational Learning*, 2nd ed. Blackwell Publishers, Oxford.

Argyris, C., Schön, D.A. (1978) *Organisational Learning: A Theory of Action Perspective*. Addison-Wesley, Reading MA.

Argyris, C., Schön, D.A. (1996) *Organisational Learning II. Theory, Method, and Practice*. Addison-Wesley, Reading MA.

Brown, M.L. (2006) Learning and Food Security in the European Union. In M.L. Brown, M. Kenney and M. Zarkin (eds.) *Organizational Learning in the Global Context*, Ashgate.

Brown, M.L. and M. Kenney (2006) Organization Learning: Theoretical and Methodological Considerations. In M.L. Brown, M. Kenney and M. Zarkin (eds.) *Organizational Learning in the Global Context*, Ashgate.

Cooke, J.G.. 1989: "Simulation studies of two whale stock management procedures," *Rep.Int. Whal Comn. Special issue 11: The Comprehensive Assessment of Whale Stocks: the early years*, 147-56.

GEA (The Global Environmental Assessment Project), 1997. A Critical Evaluation of Global Environmental Assessments: The Climate Experience. CARE, Calverton.

Hall, P.A. (1993) Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain. *Comparative Politics* 25(3), 275-296.

Ishii, A. (2001a) Merging the EU Acidification Strategy: Evaluating the 1999 Gothenburg Protocol to Abate Acidification, Eutrophication and Ground-level Ozone, *Review of European*

Community and International Environmental Law, Vol. 10, No. 2, pp. 210-226.

Ishii, A. (2001b) Fukusū-Osen-Busshitsu-Giteisho (1999): "Gaikō-Kagaku" ni yoru "Kō-sho no Riseika" [The Multi-effect, Multi-pollutant Protocol (1999): Rationalizing diplomacy by "diplomatory science"]. Studies: Ningen/Seimei /Shakai [Studies: human being/life/society], No. 5, Social Life Science Programme, Mitsubishi Kasei Institute of Life Sciences, Machida, Japan (in Japanese).

Ishii, A. (2004) Diplomatory science: Modeling the hybrid character of policy-advisory science for diplomacy. In F. Biermann, S. Campe, and K. Jacob (eds.) (2004). *Proceedings of the 2002 Berlin Conference on the Human Dimensions of Global Environmental Change "Knowledge for the Sustainability Transition. The Challenge for Social Science,"* 261-270.

Ishii, A. (2006) Gaikō no Bunmyaku o torikomu Kagaku-Kenkyū: Gaikō-Kagaku (Dipulomatorī Saiensu) Moderu no Kōchiku to sono Seisakuteki-Gan'i [Evaluating the usefulness of scientific assessments in the diplomatic context: The "diplomatory science" model and its policy implications]. *Japan Journal for Science, Technology & Society* 14, 39-61.

Jasanoff, S. (1987) Contested Boundaries in Policy-Relevant Science, *Social Studies of Science*, Vol. 17, No. 2, 195-230.

Jasanoff, S. (1995) *The Fifth Branch: Science Advisers as Policymakers*. Harvard University Press.

Jasanoff, S. (ed.) (2004) *States of Knowledge: The Co-Production of Science and the Social Order.* Routledge.

Jasanoff, S. and Wynne, B. (1998) "Science and Decision-making," in S. Rayner and E. Malone (eds) *Human Choice & Climate Change*, *Volume 1*—"*The Societal Framework*," pp. 1–88. Columbus, OH: Battelle Press.

Okubo, A., and A. Ishii (2004) Kokusai-Hogei-Iinkai ni okeru Fukakujitsusei no Kanri: Jisshō-Shugi kara Kanri-Shikō no Kagaku e [Management of Uncertainty in International Management of Whaling: Shifting from Positivism to Management-Oriented Science]. *Journal of Science and Technology Studies* 3, pp. 104-115 (in Japanese).

Polanyi, M (1962) The Republic of Science: Its Political and Economic Theory. *Minerva* 1:54-74.

Siebenhuener, B. (2002a) How do scientific assessments learn? Part 1. Conceptual framework and case study of the IPCC. *Environmental Science and Policy* 5, 411-420.

Siebenhuener, B. (2002b) How do scientific assessments learn? Part 2. Case study of the LRTAP assessments and comparative conclusions. *Environmental Science and Policy* 5, 421-427.

Underdal, A. (2000) Science and politics: the anatomy of an uneasy partnership. In Andresen, S., Skodvin, T., Underdal, A., Wettestad, J. (eds.), *Science and politics in international environmental regimes; Between integrity and involvement*, Manchester University Press, Manchester, UK.