Integrated Assessment **2:** 17–30, 2001.

Agent-based integrated assessment modelling: the example of climate change

Scott Moss [a], Claudia Pahl-Wostl [b] and Thomas Downing [c]

[a] *Centre for Policy Modelling, Manchester Metropolitan University, Aytoun Building, Aytoun Street, Manchester M1 3GH, UK*
[b] *Swiss Federal Institute of Environmental Science and Technology (EAWAG), Überlandstrasse 133, CH-8600 Dübendorf, Switzerland*
[c] *Environmental Change Institute, University of Oxford, 1a Mansfield Road, Oxford OX1 3SZ, UK*

Received 24 March 1999; revised 17 May 2000

Current approaches to deal with the socio-economic implications of climate change rely heavily on economic models that compare costs and benefits of different measures. We show that the theoretical foundations underpinning current approaches to economic modelling of climate change are inappropriate for the type of questions that are being asked. We argue therefore that another tradition of modelling, social simulation, is more appropriate in dealing with the complex environmental problems we face today.

Keywords: social simulation, integrated assessment, climate change, complexity, agent-based modelling, stakeholder participation

1. Introduction

Climate change brings into sharp relief the most difficult problems in addressing the interaction between science and policy.

Instead of trying to solve an existing environmental problem, the aim is to devise policies to prevent a problem that scientists expect to emerge. The process leading to the problem is not itself well understood. There are partial physical theories but no usable theory of the whole of the relevant physical system. There are no useful and relevant social theories and, yet, the source of anthropogenic climate change is ultimately the behaviour of social systems. Both systems are characterised by enormous complexity and the interaction between those systems cannot but add to that complexity. Moreover, the information available to policy makers makers is plagued by large inherent uncertainties due both to measurement errors and conceptual ambiguities. Reality is complex and the prediction of future developments is beyond our capacities now and for the foreseeable future. As a result, the normal approaches of the natural sciences are not applicable to a problem posing one of the greatest potential threats to the future of the planet so far identified.

In situations where decision stakes are high and uncertainty looms large, conventional scientific approaches may not be relevant. We argue that it is necessary to step back from the development and application of high theory and return to observation and classification in order to find a new basis for devising public policies. In particular, we argue that economics based approaches to integrated assessment from the DICE model [27] onwards are fundamentally flawed. Never having been based on observation and thus paying little attention to ascertaining conditions of application, economic theory is irrelevant on its own terms to the analysis of climate change. We make this point at considerable length in sections 2 and 3 because of the influence of these models in the political process, the importance we attach to the hazards of climate change and the need for useful and relevant guidance for policy formation.

The approach we offer in the place of economics is drawn from agent-based social simulation. The agent in social simulation is a software representation of real actors. Agent-based social simulation models are concerned with the ways in which social structures emerge from interactions among individuals and how those structures influence and constrain individual behaviour thereby to alter or reinforce the social structures.

The advantage of agent-based social simulation is that it can combine the problem orientation and commitment to observation of the sociologist and anthropologist with more formal approaches and arguably more careful methodology of the natural scientist. We argue that agent-based social simulation supports a new methodology that itself provides a suitable framework within which to collect observations of the social and physical systems, to generalise from those observations and to identify relationships and processes that must be understood before policies to deal with climate change and its effects can usefully be formulated.

2. The integrated assessment experience and the interface between physical and socio-economic modelling

The interface between physical and the social sciences in general and between physical and socio-economic modelling in particular is shaped by the prevailing dominance of the economist's view on how the decision problem should be framed. In this view dealing with climate change is mainly a cost-benefit problem. The costs of measures for preventing climate change have to be compared to the benefits of preventing potential damages from climate change. Physical models provide climate scenarios, economic models serve to quantify the costs and benefits for reducing greenhouse gas emissions (GHGs).

In the field of integrated assessment one attempts to integrate physical and economic aspects within one modelling framework to be able to provide more meaningful information to decision makers. Integrated assessment models range from highly aggregated models such as the DICE model of Nordhaus [27] to process based models where processes from climate to ecosystem change to the response of humankind are addressed in a detailed fashion [34]. The DICE model is a dynamic optimisation model for estimating the optimal path of reductions in GHGs. An aggregated global welfare function is optimised where choice is limited to consuming goods and services, to investing in productive capital or to slowing climate change.

Such approaches lead to disputes and arguments that do not address the core problem. Among economists, for example, there is a dispute over the appropriate level of the discount rate and to what extent cost-benefit considerations are useful for informing political debate. Several attempts were made to resolve the unrealistic assumption of a single decision maker by using multi-actor models, e.g., to model adaptive management of climate change mitigation in order to determine the optimal carbon tax or trading patterns (e.g., Hasselmann [13]). However, there is a notable absence of alternative approaches which try to tackle the problem by starting from the realisation that we simply do not understand the relationships involved or their consequences.

The interface between physical and social modelling has so far rested on something like a damage function which either entered a cost benefit analysis or served as a target to measure the effectiveness of response strategies. Current approaches, following standard economic modelling practice, imply greater predictability in the environment and the consequences of physical–social interaction than experience shows to be warranted – even (or particularly) with economic models on their own. These models do not allow for new behavioural patterns and social processes to emerge.

The essential problem is that economists start with a given set of modelling techniques and representations and then specify the object of the analysis in a manner that is suitable for those techniques and representations. What we need to develop is a framework within which to anticipate opportunities and threats and which will support changes in analytical approaches when the events that do emerge turn out to be different from our anticipations. No modelling technology that could support the analysis of such emergent phenomena will easily be reconciled with the equilibrium models of economists.

3. The economics problem

In this section, we address the issue of cost-benefit analysis by applying it to the use of integrated assessment models drawing on economic theory. The particular question we address is whether, in principle, the expected benefits from using such models to determine optimal social policies such as carbon tax levels exceed the expected costs. We conclude that no such assessment can be made for economic models or, we believe, for any model at all.

The argument turns on a formal analysis by Moss [22] of the costs and benefits of using any model for policy analysis.

The essence of cost-benefit analysis is to maximise the expectation of net benefit. A model should be used for policy analysis in a manner which maximises the policy analyst's expectation of policy benefit net of all costs associated with the analysis and policy implementation. In particular, an economist who satisfies his own definition of rationality will maximise the benefit of the policy actions less any costs of implementing the policy or any costs of identifying whether the model generating the policy were valid. Thus we shall say that a particular policy is implied by a model whenever the model is the best available and, at least, is no worse than any previous or current policy model.

The fundamental issue here is model validation. The terminology comes from computer science where a program or system is validated by demonstrating that its designed mechanism works as intended. Obviously, no program, including computational or simulation models, are applicable in all circumstances. An element of the validation process is to determine the circumstances in which the program design is appropriate. In applying cost-benefit analysis to the use of any particular model, we are in effect engaging in a process of validating the model quantitatively and/or determining the conditions in which the model is valid.

The formal argument is presented in the appendix. A verbal description of the argument is presented here.

Both the costs and the benefits of actions are assumed by cost-benefit analysts to have monetary values. In the case of climatic change, the value of the benefits of a successful policy will include the value of outputs of goods and services that would not have been produced in the absense of the policy (perhaps the greater provision of skiing facilities), the value that households would pay for leisure activities that would not have been possible due to climatic changes (additional skiing time, for example), the value of damages from natural disasters that would arise from climatic change. Costs would include any additional costs to industry from the investments required to reduce GHG emissions and reduced outputs as a result of the policies (sun blocks, for example).

An additional type of cost would be the cost of determining whether the policy itself is likely to have the consequences claimed for it. If the policy is implied by some social or economic theory or model, the consequences claimed for the policy will be realised if the theory or model is "correct". Indeed, if the criterion of goodness of a theory or model is that it should yield correct predictions, then a particular application of that criterion is that the consequences of policy-based actions predicated on that theory or model should have the consequences implied by it.

It is conceivable that some theory or model is so general and well founded as always to yield the correct predictions (including predictions of the consequences of adopting some policy) always and everywhere. It is also possible that a theory or model will sometimes yield correct predictions and sometimes yield incorrect predictions. When dealing with complex systems and, even more importantly, when dealing with the interactions among several complex systems such as physical, biological and socioeconomic systems, determining the circumstances in which one can have confidence that a particular policy will have the intended consequences is a highly non-trivial task. If it is possible, it may well be expensive. This expense, if incurred, would be an additional cost of the implementation of a policy.

To represent this problem as one of cost-benefit analysis, it is necessary to assume that there is an expected benefit from the policy conditional upon the validity of the underlying theory and a different expected benefit in the event that the underlying theory is invalid. If the theory is thought to be valid in given conditions – the theory's *conditions of application* – then the expected benefit if the model is correct will be conditional on the satisfaction of those conditions of application. Similarly, the expected benefit if the model is not applicable will be the expected benefit conditional on the failure of the conditions of application. The derivation from this specification of a useful formula is reported in the appendix. The formula is:

$$\big(1 - \Psi(\Phi)\big)E(B|\neg C) + c(\Phi) \geqslant 0, \tag{A.6}$$

where Φ is the set of conditions of application that is investigated to determine whether or not they hold, $\Psi(\Phi)$ is the subjective probability that all of the conditions in Φ will be satisfied, $E(B|\neg C)$ is the expected benefit of a policy given that its conditions of application do not hold and $c(\Phi)$ is the cost of ascertaining whether all of the conditions are true in the set of conditions of application Φ. There would be a condition of this form for every combination of conditions of application.

The inequality (A.6) defines the circumstances in which the rational economic policy analyst would *not* ascertain whether or not the conditions in Φ, the conditions of application of the theory, hold. The first term on the left side of the inequality is the product of the subjective probability that the conditions of application in Φ would not all be true and the expected benefit from the model generated policy if it were indeed the case that not all conditions of application were satisfied. Consequently, if the policy analyst were convinced that, for any set or subset of conditions of application, there were actually or virtually no chance of the conditions of application not being true, then the policy should be applied without incurring any costs of validation of the model. In effect, such a policy analyst would be convinced that his model were completely general in application. Alternatively, the policy analyst might accept that there are circumstances in which the model were inapplicable but that, even in such cases, the benefits following from the policy action would not be so bad as to make it worthwhile to incur any costs to validate the model. Finally, it might be that the cost of validating the model in any given set of conditions were large in relation to the costs of using the wrong model to determine policy.

In short, there are three issues to be taken into account when applying cost-benefit analysis to the question of whether to validate a policy generating model. These are:

- The costs of validation.
- The likelihood that the model would fail the validation procedure.
- The cost of implementing policies derived from the model if the model were not in fact valid.

3.1. Validation issues with respect to the DICE model

Sanstad and Greening [35] note that three types of economic modelling approaches are applied to climate modelling:

- neo-classical growth theory based on aggregate production functions,
- neo-classical general equilibrium theory,
- large-scale energy-sector models.

All of these approaches rely on some measure of social welfare represented as a social welfare function which takes as inputs the utility functions of individual agents in the economy. And they use a production function. We use the DICE model first published in Nordhaus [27] as a typical representative of this class of models. The DICE model is an intertemporal, optimal-growth model. The model maximizes social welfare by choosing values for three decision variables (consumption, investment and emissions control) subject to several economic and geophysical constraints. The exact form of the objective function is given by:

$$\max U = \sum_t \frac{P(t)\ln(C(t)/P(t))}{(1+\rho)^t}. \tag{1}$$

Discounted utility, U, is a function of per capita consumption, C/P, discounted at rate ρ.

Global output, $Y(t)$, is given by a Cobb–Douglas production function:

$$Y(t) = \Omega(t)A(t)L(t)^{1-\gamma}K(t)^{\gamma}, \tag{2}$$

where $A(t)$ represents technology, $L(t)$ is labor, $K(t)$ is capital, and γ is capital elasticity. $\Omega(t)$ relates production to the cost of controlling greenhouse gas emissions, TC, and damage from climatic change, d:

$$\Omega(t) = \frac{1 - TC(t)}{1 + d(t)}. \tag{3}$$

Both TC and d are expressed as fractional loss of global output.

These three equations determine the essential decision-making issues captured by the DICE and all economics-based integrated assessment models. These are the trade-off between damage to the environment as a result of production activities and the investment required to reduce that damage and the tradeoff between that investment and household consumption. The idea is that less investment now means more environmental damage (e.g., as a result of higher greenhouse gas emissions generating higher mean global temperatures) but also leaves more output for household consumption which is identified with greater social welfare. Reducing household consumption allows for more investment and therefore less environmental damage.

Evidently, the equations capturing these constraints must correctly reflect the nature of production, the ability to enhance or restrain social welfare and the relationship between production, investment and environmental damage. If this is not possible *in principle* then there is no credible argument that the DICE or any other model relying on these relationships can be valid. For this reason, we take these three equations to define conditions of application of the DICE model. Since policy benefit is the value of the social welfare function (1), the model can hardly be applicable to circumstances in which social welfare functions are themselves inapplicable. Also, the relationship specified in equation (2) between inputs of capital and labour, on the one hand, and aggregate output, on the other hand, allowing for technological change over time as captured by $A(t)$, must be able to correspond to actual production relations and technological change if the modeled effect of GHG emissons is to have any validity. If the social welfare function cannot be applied and the specification of production relations and technical change is inherently flawed, then it would be hard to justify the use of a model that relies on either or both of these relations.

In addressing these issues, we rely on theorems proved by respected, mainstream economists and published in core economics journals. The one theorem published by a heterodox economist was in a core journal and accepted in its formal aspects by the Nobel Laureate in economics who published the first version of equation (2) above.

3.1.1. The social welfare function

Any assessment of the conditions of application of an equilibrium economic model turns on *the Pareto conditions*. These are the first order conditions for the maximisation of utility by households and profits by firms. To achieve these, the composition of each household's consumption must be such that there is a common ratio of marginal utility to price for each commodity consumed. Similarly, each firm must hire inputs to the production process so that there is a common ratio for each such input of marginal product to input price. The second-order conditions (so that utility and profit are maximised rather than minimised) are satisfied if all of the utility functions and production functions are convex.

It is always possible to add a constraint to any equilibrium system such that at least one of the Pareto conditions cannot be satisfied. One way of doing this is by imposing a tax on one or more commodities. If taxes are imposed on several commodities and they are imposed at different rates, then there will be several, complex constraints preventing the satisfaction of a multiplicity of Pareto conditions. An obvious and relevant example of a system of such taxes are carbon taxes.

It has been known clearly and unambigiously since 1956 that, in any regime of taxes on commodities, the effect on social welfare cannot be determined by any general means. In the present case, it is not possible to say that a higher carbon tax rate will have any particular effect on social welfare. It might increase social welfare or it might diminish social welfare. Economic theory says only that we cannot say. This is a conclusion of the general theory of second best due to Lipsey and Lancaster [17]. A key passage from the relevant paper is this:

> "The general theorem of the second best states that if one of the Paretian optimum conditions cannot be fulfilled a second best optimum situation is achieved only by departing from all other optimum conditions. It is important to note that in general, nothing can be said about the direction or the magnitude of the secondary departures from optimum conditions made necessary by the original non-fulfillment of one condition. Consider, for example, a case in which the central authority levies a tax on the purchase of one commodity and returns the revenue to the purchasers in the form of a gift so that the sole effect of the tax is to distort relative prices. Then all that can be said in general is that given the existence and invariability of this tax, a second best optimum can be achieved by levying some system of taxes and subsidies on all other commodities. The required tax on some commodities may exceed the given tax, on other commodities it may be less than the given tax, while on still others a subsidy, rather than a tax, may be required."

The implication of this passage – and there are others like it in Lipsey and Lancaster [17] – is that the imposition of any tax on commodities (including a carbon tax) is sufficient to render any social welfare function otiose. If you believe the theory, then you cannot believe that the conditions of application of the social welfare function are satisfied. And, of course, if you do not believe the theory then you should not be using it in the first place.

3.1.2. The production function

The core element of the DICE model production function is the *Cobb–Douglas* (1928) *function* of the form $Y = aK^{\gamma}L^{1-\gamma}$, where a is any positive constant, K is the stock of capital and L is the employment of labour in the economy as a whole. The function has properties which are extremely desirable from the point of view of economic theory. The principle such feature is that if labour is paid the equivalent of its marginal product ($\partial Y/\partial L$) and capital is paid the equivalent of its marginal product ($\partial Y/\partial K$), then the total payments to labour and to capital are equal to the value of output. This is consistent with elementary properties of national income accounting. Additional, essential properties of a production function is that the second derivatives of the function with respect to capital and labour should be negative so that profits are maximised rather than minimised. The Cobb–Douglas function and all functions of its class (i.e., linear homogeneous functions) have this property, too.

There is, however, a problem. It has been known since the mid 1960s (and chronicled by Harcourt [12]) that these properties of the production function hold in only a few circumstances. Two of these circumstances pertain only in a general equilibrium with steady growth so that all values in the economy such as population, stocks of machines, outputs of goods and services all grow at the same constant rate. If there is a single, constant rate of profit generated by production in all firms in such an economy and that rate of profit is equal to the rate of growth, then formally the rate of profit and the wage rate will be equal to the marginal product of capital and labour, respectively. The second case is where, in every production process, the proportions in which machines are used is identical and the outputs of machines are in the same proportions as those in which they are used as inputs. In other words, there is a unit collection of machines of given proportions and all production sectors of the economy use the equivalent of a number of such unit collections to produce their outputs. And taking all such outputs together, they amount to another number of such unit collections. A little more concretely, every production process requires one standard blast furnace, two standard oil refineries, six standard sugar refineries, 492 standard spades, ... The total output of the economy must then be x standard blast furnaces, $2x$ standard oil refineries, $6x$ standard sugar refineries, $492x$ standard spades, ...

In the first place, no such equilibrium has ever been observed. In the second place, the point of carbon taxes is to encourage investments that will change the composition of capital equipment over time from equipment embodying more to less greenhouse gas emitting technologies. Consequently, even in the fantastic land where the composition of inputs and outputs were the same initially, the point of the taxation regime would be to change that composition thereby to violate the conditions in which the Cobb–Douglass or any other aggregate production function of its (linear homogeneous) class had the required properties.

There is one remaining case which has been identified as one where the aggregate production function used in the DICE and similar models is applicable. That is where there is one commodity produced in the economy, it is used both as a capital good to produce more of itself and as the only good that households consume, it is costlessly malleable, infinitely divisible and, *by assumption*, the additional output resulting from additional inputs of this commodity declines as the amount per input of labour increases. In this case, it is possible for the economy continuously to be in an equilibrium described by the Cobb–Douglas or any linear homogeneous production function and with the consequent bonus that the Pareto conditions are satisfied and a social welfare function such as equation (1) above is applicable.

Now add the element of the DICE model production function accounting for technological change – $A(t)$. This specification of the production function was devised by

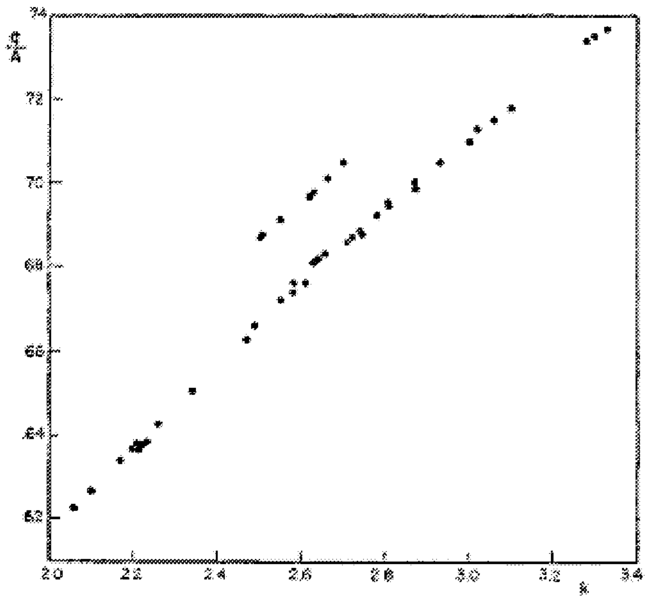

Figure 1. Solow's production function – "technical change" removed (reproduced from Solow [38]).

Solow [38], who argued that statistical differences between observed outputs and those implied by a production function could be interpreted as being the result of technological change represented as a shift of the production function itself.

By making a whole host of assumptions – Solow was completely honest about this – he virtually invented a series of values for capital[1] and assumed a particular (and peculiar) characteristic of technical change and then separated out changes in outputs due to that technical change and changes in outputs due to changes in the amount of capital employed per unit of labour. Figure 1, showing the relationship between output per worker and capital per worker after eliminating the effects of "technical change", is reproduced from Solow [38].

Note that seven points are above and to the left of the main body of points forming a slightly curved line. This set of points stumped Solow until Hogan [14] pointed out that they were due to an arithmetical error in calculating those data points. Hogan looked for the arithmetical error because he knew that the closeness of the relationship between output and capital, as calculated by Solow's technique, depends only on the extent to which the distribution of income (i.e., the fractions of income going to wages and profits, respectively) is constant over time. During the period for which Solow took his data, the share of wages in income in the USA was virtually constant at about 65%.

[1] "The capital time series is the one that will really drive a purist mad", wrote Solow. "Ideally, what one would like to measure is the annual flow of capital services. Instead one must be content with a less utopian estimate of the stock of capital goods in existence." Moreover, "lacking any reliable year-by-year measure of the utilisation of capital", Solow assumed the unutilised proportion of the capital stock is the unemployment rate.

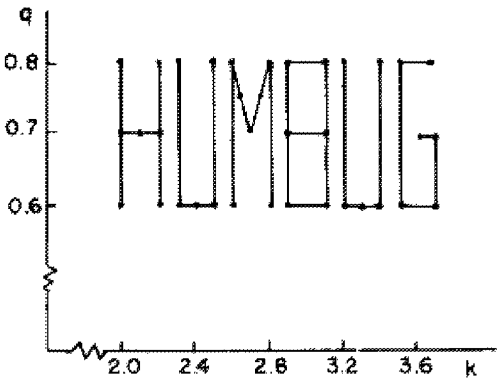

Figure 2. Shaikh's "data": the Humbug economy (reproduced from Shaikh [37]).

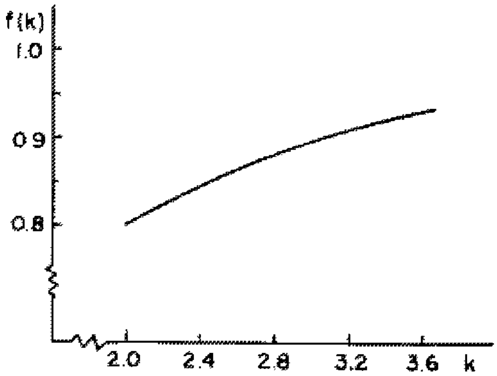

Figure 3. Shaikh's "production function": underlying Humbug production function (reproduced from Shaikh [37]).

Hogan's critique was followed 16 years later by Shaikh [37], who proved Hogan's verbal demonstration algebraically. He then took an invented set of points spelling out the word "HUMBUG" to relate output to capital. Shaikh's diagram of this "data" is reproduced as figure 2. He added the assumption that the profit share was constant at exactly 35% and applied Solow's technique to the data. The resulting "corrected" production function is reproduced as figure 3.

The conclusion is inescapable: Solow's technique for distinguishing between the effects of technical change and the effects of capital investment does no such thing. At best, it provides a complicated measure for the constancy of income distribution.

For integrated assessment modelling, the conclusion must be that even if a single, divisible, malleable good were used for both production and consumption, if there were technological change going on then the production function of the DICE model would not relate to production but only to the outcome of the social process determining the distribution of income between labour and the owners of capital.

3.2. General equilibrium models

Although economists have long since consigned the foregoing results to the scrapheap of history, an early response was frequently that the results concerning the aggregate production function did not affect general equilibrium theory in which capital goods were represented individually. It is therefore appropriate to consider whether results in general equilibrium theory suggest conditions of application in which a social welfare function can be meaningfully defined. As already noted, this would be an equilibrium in which all of the Pareto conditions were satisfied. Since all, but only, general equilibria are characterised by the satisfaction of all Pareto conditions, it is sufficient to consider conditions in which such equilibrium is possible.

In the most general versions of general equilibrium theory, all transactions are agreed at the start of time. Each transaction is contingent on a set of events pertaining at the time the transaction is to be completed. For example, an individual might contract to buy an umbrella in Duluth, Iowa on 12 January 1944 provided that it is raining and the individual is in Duluth on that day. Since all of these models require there to be a given set of individuals defined by utility functions and probability distributions for the occurrence of every possible event at every date, any individual around at the Creation is assumed to be around forever – including on the date 12 January 1944.

The first of these models was due to Arrow and Debreu [3].

The first economist to pursue the consequences of assuming that individuals in such a system could agree and complete transactions at every one of a sequence of dates was Radner [33]. The implication found by Radner for the existence of a general equilibrium when there are spot markets (i.e., markets in which goods are exchanged for money at the same time as, for example, in a shop) is this:

Agents can have rules to determine their supplies and demands for the various goods and services available to them in different circumstances. In the original general equilibrium model of Arrow and Debreu [3], such rules were not necessary because all the information that would ever be available was conveyed by prices. But when trading takes place and can be revised over time, then the ways in which other individuals behave will affect the consequences to you of your own behaviour. Moreover, the amount of information about other agents and their rules of behaviour grows without limit. Unless individuals all have sufficient computational capacities to calculate their own best rules of behaviour by identifying the rules used by the other individuals (and the effects of those rules on themselves), there can be no general equilibrium. If, however, computational capacities are limited, then eventually there will be more information available than the individuals can use. Consequently, general equilibrium cannot exist unless individuals have unlimited computational capacities.

3.3. Economic approaches to integrated assessment of climate change: conclusion

Our conclusions could hardly be simpler. The aggregate production function itself has no meaning in a world of heterogenous fixed capital and, if it did, the effects of technological change on the relationships between inputs of labour and capital and outputs could not be identified empirically. General equilibrium approaches are not applicable unless it is argued that all individuals have effectively unlimited cognitive and computational capacities.

In terms of the cost-benefit analysis described at the start of this section, the probability that any of these conditions of application will be satisfied is arguably nil. We already know from Roughgarden and Schneider (1999) that the "benefits" implied by the model being incorrect in relation to the parameters of the damage function, could be large and negative. The resources being devoted to integrated assessment and the development of climate policy indicate a widespread and influential view that failure to deal appropriately with anthropogenic climate change could have very high costs indeed. Failure to deal appropriately with climate change might result from conclusions drawn from economic models that operate outside of their range of application. At the same time, the costs of determining that we do not all have unlimited cognitive capacities and that capital equipment is not homogeneous and costlessly malleable are simply the costs of common observation and are, therefore, very small.

It follows that even the economist who is rational in his or her own terms will reject the prevailing economic approaches to integrated assessment.

4. Integrated assessment modelling *de novo*

The influence and importance of economic models in the integrated assessment research programme suggests that the rejection of that approach makes a new start reasonable. Our objective is to outline a research programme in relation to the intended destination rather than somebody else's existing point of departure. It is therefore appropriate to identify some key issues for the development of a new integrated assessment modelling technology.

We start by recognising that our target is to develop tools to inform climate policy analysis concerning the interactions between natural and social systems, each on a global scale. Each of these systems is immensely complex. We have good, reliable models of some relatively small elements of the natural (physical and biological) systems but we have no prospect of a predictive theory or model of those systems as integral entities (e.g., a human body or an ecosystem) or, even more surely, of the physical and biological systems together (e.g., the global carbon cycle).

As for the social system, we have no reliable models of any elements of the system and, what is more, there is no prospect of such models. The reason is clear. Many social structures and institutions have developed in order to enable individuals to make decisions in the face of their inability to predict future developments and the effects of those developments on the consequences of the individual's actions. Some social structures have developed to reduce uncertainty, the prime example being property rights and the norms of social behaviour that support respect for property rights. Societies typically develop formal representations of the more important norms of behaviour and other elements of the social structure. Most of us would not commit murder whether or not murder were an illegal act but the uncertainties generated by murderers are such that all societies have clear sanctions to remove murderers from their midst. Property laws are partly to enable individuals to know that they will not normally lose what they have or what they produce but there are many property laws and considerable case law to clarify issues so that the detailed effects of the relevant social structures are relatively clear and certain.

Other structures and institutions recognise that some events are inherently unpredictable. Thus we have stock exchanges, commodity (futures) markets, insurance schemes and the like. We also have governmental institutions that slow the pace of change in the social structures to lend stability to the social system by avoiding "knee-jerk" reactions to passing events.

It is not hard to multiply examples of social structures and corresponding institutions that exist to mitigate the effects of the unpredictability of social systems and, so, to realise that any social theory that were accurately to predict social system behaviour would change the conditions that gave rise to the system structures and forms in the first place.

We conclude that, while developing the ability to predict the course of events in physical systems does not change the systems themselves, there is no possibility of predictive social theories because such theories would themselves change the structure of the system. At the same time, all cost-benefit analyses depend on being able to identify all possible future outcomes, to predict their values (positive or negative) and to assign to each such event a probability of its occurrence. History and experience combine to suggest that the social system has developed to enable individuals to undertake commitments in the face of their inability to formulate probability distributions for the occurrence of all possible relevant futures. This is a different kind of system from that in which cost-benefit analysis is applicable.

4.1. Model validation and verification

If one does not rely on predictions to test a model's reliability what else can be the criteria to build trust in a model's quality and plausibility. We suggest to use a correspondance to issues of software validation and verification in computer science. Integrated assessment models are in practice computational models implemented as computer programs.

A program is said to be validated if it is demonstrated to do in practice what it was designed to do. A program is said to be verified if it is demonstrated to be consistent and sound relative to some appropriate formalism. The purpose of validation is to ensure that programs do what is expected of them in actual operating conditions while the purpose of verification is to ensure that the properties of the program design are clearly understood and the program properties are not contradictory.

In social simulation modelling, validation is the demonstration that the model as program produces outputs that correspond to observable properties of real social systems. They do not capture every property and do not replicate in detail every event of real social systems but they capture some properties and replicate some events. A validated model would relate *specified* conditions to social processes producing observed outputs. Those specified conditions are the conditions of application of the model.

A verified model is one that is demonstrated to be sound and consistent with respect to some logical formalism. Several research teams in the field of multi agent systems (which includes social simulation) have adopted a *compositional verification* methodology. In software engineering, the purpose of compositional verification is to start with desirable system properties, break up the system into components of manageable size in order to identify the properties of those components that will guarantee the desired system properties. The next step breaks up each of those components into manageable sub-components and then those into sub-sub-components, and so on, until the properties of the lowest level of component can be proved formally.

To take a physical example, Horton's Law states that the number of links among river segments increases as a power law in the order of river segments (i.e., the number of links to other segments that has to be passed before the river reaches the sea) (Bak, 1997). For purposes of modelling surface water flows over an area of land, the power law distribution can be used. But if it is intended to change the course of a river or to build a dam, then a more detailed model of the river system in question and its relation to population or industrial centres, etc. will surely be required. Similarly, in markets there is both empirical and simulation evidence that market shares have a power law distribution (Moss [25] and Moss et al. in prep.) and that the longevity of firms is distributed as a power law. In coarse grained models, it will sometimes be appropriate to represent markets by the appropriate, empirically estimated, power law distributions. But in modelling the behaviour of participants in such markets, an explicit representation of actors and their behaviours will be required. In this case a test of the consistency of these more fine grained representations with the more coarse grained representations will be whether the structure of river segments or the market shares of the more fine graned models corresponds to the same, empirically estimated power law distribution as is used in the more coarse grained models. More generally such approaches will serve to investigate the link between system properties at the macro-scale and the composition of the constituents at the micro-scale (Pahl-Wostl [29]).

Verification and validation should go hand in hand. To explicate this proposition, we turn first to the relationship between validation and conditions of application. Once that relationship is clear, we turn to issues of verification and argue that compositional verification in particular supports the development of simulation models that are readily comprehensible to modeller and user alike and, in addition, are demonstrably consistent with well validated representations of individual and group behaviour.

4.2. Validation and conditions of application

In the physical sciences, models are frequently validated experimentally. In the Popperian ideal of real science well designed experiments should allow the falsification of any theoretical statement. However, this ideal has its severe limitations, already in the physical sciences. Where big, complex systems are involved, as in the case of meteorological models, some model components are frequently already well validated (for example, thermodynamic equations) while the model as a whole can only be validated statistically against experience. Models are "tuned" by estimating parameters so as to capture on average and with minimum variance the time patterns of movements and characteristics of weather systems.

Models of social systems cannot in practice be validated experimentally although there is considerable and interesting research applying psychological experimental techniques to assess the extent to which individuals' behaviour corresponds to economists' assumptions. In general, it does not (Allais [1]; Kahneman and Tverski [15]).

Since the target systems of integrated assessment models are large, complex physical, biological and social systems, experience would not lead us to expect experimental validation of the models. Moreover, the very size and complexity of these systems individually as well as their combined complexity ensures that no model that we can comprehend will be correct and complete. Our models can be no more than a set of descriptions of what we observe together with suppositions about currently unobservable relationships that help to make our descriptions of the observable more comprehensible to us. So our purpose is not to find a correct theory in the sense of a theory providing accurate predictions of future events but rather to find descriptions of the phenomena of interest that help us to formulate plans of action to mitigate adverse outcomes and to enhance positive outcomes.

An obvious issue here is to determine the appropriate descriptions. Models of the relevant natural systems will naturally cohere with experimentally and observationally established relationships. Models of relevant social systems can only cohere with observationally established relationships. Some of these relationships will be described by independent observers and some by stakeholders. The point is to capture these various descriptions in a manner that does not distort them but that supports both the development of a consistent common description of the relevant relationships and that supports an exploration of the consequences of actions, given the correctness of the consistent common description of those relationships.

Instead of supposing there to be a "representative household" or a single policy-setting figure, there is clearly some advantage in descriptively accurate representations of the relevant actors and the ways in which they interact. We submit that a powerful approach is to represent the actors by means of computer programs that take as input representations of perceptions by the real actors, that process those perceptions in order to produce outputs corresponding to the actions of the real actors. These computer programs are known to the distributed artificial intelligence community as agents and this approach to describing actors and interactions among them has been developed by social scientists as agent-based social simulation. What is novel about the approach suggested here is the direct involvement of stakeholders as participants in the specification and evaluation of the agent-based social simulation models. A feature of this involvement is that the stakeholders specify the conditions leading to various observed behavioural patterns and are then involved in either the corroboration of the outputs from the models as accurate or at least plausible or, if they cannot offer such corroboration, the identification of any misspecification of behaviour or relationships that will have generated the inaccurate or implausible results. Stakeholder participation in this way provides an important basis of model validation.

Stakeholder participation also insures the identification of socially constructed reality and the mental models of individual actors. A person's behaviour is to a large degree affected by expectations, which are in turn based on a person's model of reality. Humans not only construct reality in their minds; their behaviour also causes this reality in their minds to become reality in their environment. Group model building has shown to be a powerful instrument to reveal the importance of mental models and to deal with messy problems (e.g., Vennix, 1999).

A clear difference between this participatory agent-based social simulation (PABSS) and the existing conventions of micro-economic, optimising models typified by DICE and its genre is that validation in this case is a reflexive process involving both modellers and stakeholders whereas validation (if done at all) in the DICE and similar models starts from formal specifications of behaviour and relationships derived from economic theory and compares the model outputs with available statistical data. The participatory nature of PABSS ensures that validation is a more thoroughgoing process involving also discussions between stakeholders and modellers. In this respect we follow the tradition of soft systems methodology where the quality of a model depends on its ability to structure the discussion about a problem (Checkland [8]).

The conditions of application of the models apply explicitly to the nature of the process as well as to the state of the environment. It is important that different dimensions of validation are combined namely the validation of the model against data, against knowledge of domain experts and against stakeholders perceptions in the participatory settings. This will be particularly important if we intend to identify relevant descriptions for the future behaviour of systems that is outside the realm of what has been observed in the past. Stakeholders may differ considerably in their perceptions of behavioural changes and in the expectations they hold regarding the behaviour of others. Some of these perceptions may also be implausible or even impossible.

Consider the example of consumer behaviour. Domain experts may contribute their expertise in studying the behaviour of a wide range of consumers in socio-psychological studies. Individual consumers may judge the representation of their own behaviour in a model. Statistical data on trends in consumer behaviour give empirical background on what has been observed in consumer populations over the past years. Combining these different dimensions of evidence and validation is not a trivial task and requires advances in method. Figure 4 shows the different contributions to developing and validating an agent-based model. A process to do so is currently developed within the EU-project FIRMA (Freshwater Integrated Resource Management with Agents) where particular emphasis is given to the involvement of stakeholders in the process of model building and development.

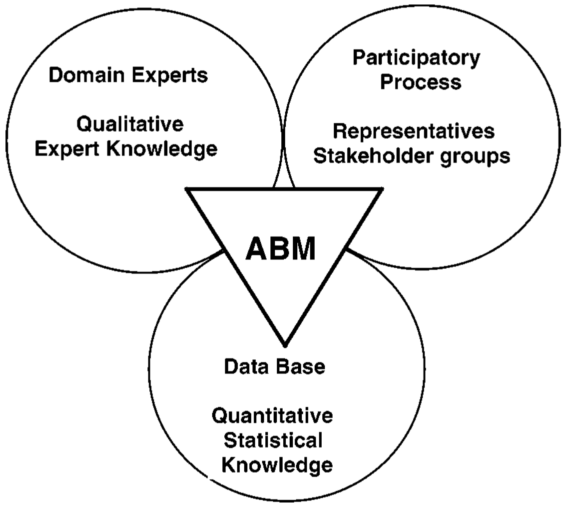

Figure 4. Different dimensions contributing to development and validation of an agent-based model (ABM).

4.3. Validation and aggregation

It has long been known (Miller, 1956) that people can hold in short-term memory between five and nine "chunks" of information at a time. Such "chunks" can be more or less complex depending on the experience of the person with the information and relationships being considered. This cognitive limitation behooves modellers to restrict the number of entities represented in their models either by implementing them as aggregates or by seeking to capture only a few of the features and characteristics of the target entities. In the language of computer science and artificial intelligence, detailed representations of target entities are fine grained while aggregate or partial representations are coarse grained.

The proposal here is that the grain of analysis of any model should be only so fine as modellers, stakeholders and other model users can understand. Swamping users with detail or large numbers of agents does not help either in the elicitation of descriptions of relevant behaviour and relationships or in the understanding of results. There seems little point in substituting a poorly understood artificial social or natural system for a poorly understood real system. In par-

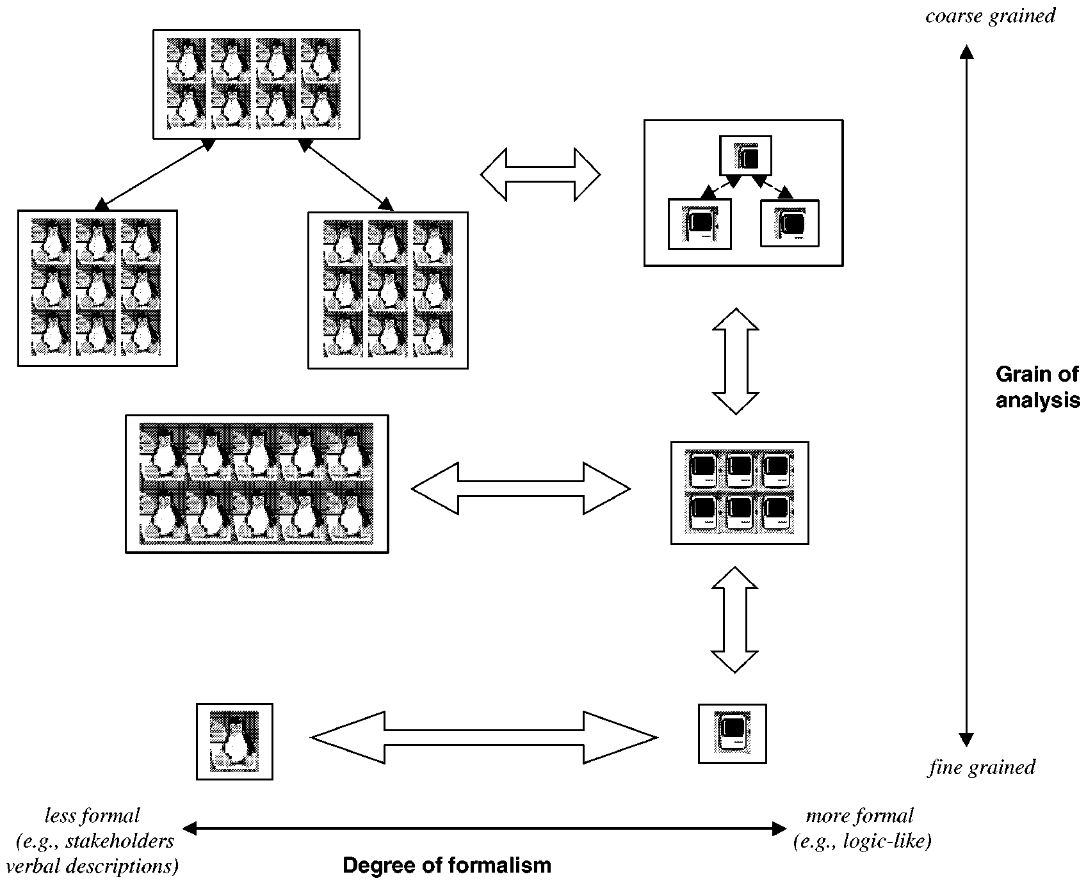

Figure 5. Grain of analysis and degree of formalism in agent-based models.

ticular the problem of over-parameterization is not trivial for agent-based models.

An important aspect of our vision of integrated assessment modelling is that the models should be convincing to stakeholders and, to make them more convincing, that the stakeholders should be able to query the models to determine *why* particular outcomes emerge from the models. Our method follows the method of compositional verification. To explain phenomena emerging at a coarse grain, the components of the coarsest grain model are themselves modelled to give a validated description of the processes, relationships and behaviour generating the outcomes observed at the coarser grain. For more detailed explanations, the components of the finer grained models can themselves be modelled to describe the sources of the previously finest grain behaviour. This decomposition of each model into components which are themselves then modelled can proceed until there is a validated model comprising agents with cognitive processes verified with respect to an appropriate and experimentally validated cognitive theory.

This implies for an agent-based model to be useful in a participatory setting that it can be explored at several hierarchical levels to show the user different levels of details of the description of the behaviour of relevant agents. And the user should be able to explore the implications of choices in parameter settings and the uncertainties that may emerge from such changes.

4.4. Validation and verification: a summary

The purpose of the verification and validation procedures discussed in this and preceding sections is to provide a suite of models at different grains of analysis such that each model is consistent with the models at finer grain (indicated by the vertical block arrows in figure 5) and every model is validated with respect either to stakeholder understandings (such understandings perhaps developed through participation in the modelling process) or with respect to independent, perhaps empirical, observation. As in figure 5, moving from finer to coarser grain typically involves implementing agent representations of fewer actors than are actually observed though it is always important to implement enough agents to capture the qualitative nature of the interactions among observed actors.

5. Characteristics of a constructive procedure

A constructive procedure for integrated assessment of climate change requires the development of an analysis of social behaviour, including its economic aspects, that has both empirical validation and qualitative plausibility. This is a pragmatic but essential issue. Instead of assuming with economists that unobserved relationships are always correct, we look to externally validated elements for our models and then assess the outputs from our models in terms of goodness of correspondence to both available statistical records and the qualitative historical record as assessed by domain experts. Those externally validated elements of the models define the models' conditions of application.

This approach is closer to that of physical modellers, who are concerned with the conditions of application of their models, than to economic modellers. Conditions of application refer often to the empirical relationships that have to be included. Confidence is generally enhanced by a model's ability to reproduce empirical data. However, our concern is not just with the replication of the fundamental features of statistical data series but also with qualitative relationships as understood by domain experts.

A difference between the modelling procedures proposed here and physical modelling is that we do not propose to predict outcomes. Given the uncertainties in global climate and the long-term unpredictability[2] of weather, there exist also severe limits to predictability in physical models (e.g., Pahl-Wostl et al. [31]). However, we would like to emphasize to take the limits to predictability one step further. Our modelling rationale is the development of tools for counterfactual experiments and for what-if analyses to inform and help focus discussions about policy measures and also to help identify social and perhaps physical processes that analysts might not otherwise have considered.

So we start our analyses from plausible accounts of important relationships in social responses to climate change. By plausible, we mean that observed qualitative conditions and statistical descriptions put into the models yield outputs which also correspond to observed qualitative features and statistical descriptions of the phenomena of concern. An important feature of such models is that they capture representations of changing social relations. These relations encompass the social embedding of an individual and the complexity of an individual's exchange with his environment. Such changing relations would include institutional changes in exchange, changing organisational structures, the development of new mental models by agents and how these affect policy assessments.

5.1. Investigation of new behavioural patterns in social systems

A new modelling approach should account for the different kinds of interaction among agents as well as with their social and natural environments. In current approaches, the interface between social and physical modelling is mainly given by the price mechanism. Potential damage of climate change serves as input into economic growth models. The consideration of measures is mainly limited to conventional policy measures such as carbon or energy taxes. However, mitigation of, or functional responses to, climate change will entail processes of social learning where environmental awareness and the formation of values are important. For strategic planning, one has to take into account the formation of expectations that may be informed by results from climate forecasts and expected policy measures. This implies that one has to account for the flow of information other than market prices. Rather than building large fully integrated physical-social models we will investigate the effects of different types of information. We will further represent the influence of information, the attitude towards and the perception of risks, different levels of individual and social values, the importance of uncertainties for the processes of policy formation.

Representations of some agents will be based in established theories of cognition. The agents represented in this way are those engaged in strategic behaviour including planning, generating and modifying social policies, guiding the process of technological and institutional innovation, determining the scale and direction of investment, etc. Other agents can be represented more simply without loss of descriptive accuracy. We have consumers in mind here particularly. Here changes in preferences, the formation of values, the emergence of product images will be of major importance. The development of models for consumer behaviour will be based on belief networks and means-end chain theory that is empirically well grounded in marketing research (Kottonau and Pahl-Wostl, in preparation). This approach focuses in particular on the mutual relationship between collective and individual learning. We consider changes at the level of procedural knowledge (rules) as the major driving force for the evolutionary socio-economic change required for sustainable development. Therefore, we will put a major emphasis on the social embedding of individual action.

We cannot say in advance what the necessary degree of reductionism must be – that is, how fine grained must be our representations of agents. Initially, we represent enterprises as engaging in activities determined by actors represented as problem space architectures, who learn by generating, testing and evolving models of their environments and other agents, converting these models into rules of behaviour so that declarative knowledge (the models) become procedural knowledge (the rules) in accordance with Newell's unified theory of cognition (implemented as Soar) or Anderson's theory of memory (implemented as Act or more recently ACT-R).

We will investigate means of aggregating over these agents in order to reduce computational expense without loss of accuracy of our representations with respect to relevant observations. At the level of aggregated groups of agents (e.g., consumers) we need to develop descriptions of the stylized behaviour that reflect the heterogeneity of the ensemble. Procedures how this can be accomplished are known from ecosystem modelling (e.g., Pahl-Wostl [29]). We will develop alternatives to the "representative agent device" used in many CGE models. Even within the rational actor paradigm Mantel and Debreu showed already in 1974 that only in highly restrictive cases of nearly identical consumers the aggregate demand has the same mathematical properties as individual demand (Debreu [10], Mantel [18]). In addition,

[2] This unpredictability can, for example, be modelled as either chaos or self-organised criticality.

we will represent markets as emergent trading relationships and practices. We will not assume market structures, degrees of competition or the effects of competition on the abilities or inclinations of individual agents to set prices or determine sales volumes. In this, we follow Marshall [19, pp. 323–330], Kaldor [16] and Moss (1981).

5.2. The importance of institutions

The theoretical analysis of major innovations turns on the relationship between the technology of exchange and the institutions that support exchange. For example, ships are bought and sold under a completely different set of arrangements than are chocolate bars. The reasons are that ships are generally not well standardised and they are expensive to store. Consequently, building ships for stock to sell to any customer that comes by would imply a long time lag between production and sale with a rapidly increasing price to cover storage and financing changes. There is less risk to the shipyards and less expense to the purchasers if ships are built to order. Because of the high costs of maintaining shipyards, the yards maintain order books so that their work is planned out for several years ahead. The order book enables the shipyards to be fully utilised in the face of fluctuating demands for ships. Chocolate bars, on the other hand, are cheap to store and they are (partly as a result of branding) highly standardised and they are quite durable. As a result, they can be sold to passing customers whose identity is not important and they can cheaply be held in stock so that the inventory fluctuations absorb minor differences in production and sales. The basis of this difference is the technologies involved in storage, transportation and communication.

Technological changes in these activities can radically change the nature of exchange. In the last century, refrigeration replaced the prevailing system of selling live animals to local butchers with the creation of huge slaughtering factories and the transport of chilled meat to the local butchers. This and many similar examples have been reported by Porter and Livesay [32]. The effects of the Internet and electronic trading could – indeed should – be analysed in these terms.

It is plausible that an important factor in the scale and pattern of energy use will involve changes in the practices of exchange. Such induced institutional change may be beneficial or not. A sucessful climate policy should foster beneficial changes enhancing productivity and at the same time reducing energy consumption. In order to investigate the importance of this possibility, we propose to model the relationships between trading patterns and practices, on the one hand, and the technologies of storage, transportation and communication (which collectively comprise the technology of exchange), on the other hand. A particular issue to be investigated at an early stage in the project will be whether an effect of a carbon tax or a tax (say) on diesel fuel would make transportation of some goods so expensive as to encourage local production and markets with consequent effects on economies of scale in production. Can electronic control and robotics reduce minimum efficient scales in production thereby to make it possible to increase transportation costs by some kind of tax and, as a result, encourage local production without loss of scale economies? If the degree of standardisation of products were unaffected (because electronic control programs can be shared via the Internet), would there be a savings in the combined costs of exchange and production without adversely affecting consumers? What kinds of market arrangements might we expect to emerge from such taxes? Would it be possible by public provision of infrastructure to facilitate such changes? And, of course, how important might be the influence of such changes on anthropogenic carbon dioxide emissions and thus on the course of climate change?

6. Future developments

We can imagine a range of different approaches which would meet the criteria of our methodological regime in application to climate change.

The issue of scale in both time and space raises questions not only about the use of physical models but also about how to combine a physical climate model and a social model. Decision making is local and short term. Climate change is in the end a global phenomenon and long term. However, the overall climate policy process started also at the global scale. Bringing these two scales together to achieve meaningful results requires much conceptual work. We offer a few suggestions for an initial set of models integrating representations of learning processes with established models of climate. These would provide both substantive output and a focus for developing the required modelling technology. Examples of applications are:

How the effects of adaptability and emergent technology on a regional scale diffuse globally. The consequences of different speeds of response associated with different dynamics of the physical/ecological systems and how these might influence technological change related to changing (or not) patterns of energy use under different policy regimes. Analysis of the principle of robust action where one should choose short term decisions such that long term degrees of freedom are maintained.

Investigation of the mutual relationship between global negotiation processes and local processes of decision making and innovation. We expect climate policy to be an iterative process where local decisions are influenced and shaped by expectations of global developments and *vice versa*.

Modelling the process of knowledge generation about the climate system, the way this knowledge affects people's belief about climate and climate change. We believe such models must take into account that uncertainties about both the physical and the social processes make it possible (even likely) that individuals will hold contradictory beliefs. Models which allow that qualitative aspects are important for consumer choice and where shared environmental awareness influences decision making.

How might expectations of future developments determine investment and life-style decisions which would support functional responses to the threat of unfavourable climate change?

In general terms, we are proposing a modelling methodology and technology to support the exploration of different global scenarios where one has different response strategies and different scenarios of change in the climate system. Our approach is based on a view of socio-economic systems where the notion of an equilibrium state does not make sense.

The legitimisation of a modelling approach for policy advice derives not only from internal criteria within science, in particular not from being based on theoretical foundations that may be subject to dispute. A model must produce plausible results regarding empirical relationships. And it must be plausible for the non-expert audience – extended peer community and an embedding of the whole modelling process into a social process with a dialogue with non-scientific experts and/or citizens being concerned as consumers and decision-makers. It is essential in this process to include local domain knowledge and the subjective assessments of the people concerned (Pahl-Wostl et al. [31]). This paper motivates and describes a framework for taking such requirements into account.

Appendix. Conditions of application for the rational economist[3]

To consider the determination of the subjective expectation of net benefit from a model-based policy, we define a policy as a set of individual actions A. We suppose that any model used to generate policy recommendations has a set of conditions of application C. Since C could be the null set, this supposition includes the possibility that no conditions of application have been specified. There will be n such conditions C_i where n is a non-negative integer and $C_I \in \{\text{true}, \text{false}\}$.

Let B be the image of the mapping $[(A|C) \rightarrow \Re]$ the value of the benefits expected from the set of policy actions in A given that a set of conditions C are satisfied.

The "observation tag" for the ith condition is $\phi_i \in \{\text{true}, \text{false}\}$ which takes the value *true* if it is intended to observe the ith condition and *false* otherwise. The intention of the policy analyst to observe conditions of application is captured by the set

$$\Phi = \bigl\{\phi_i = \text{true } (i = 1 \dots n)\bigr\}.$$

In addition, we denote by $C(\Phi)$ the cost of observing all of the conditions $\phi_i \in \Phi$.

To complete our notation we require some means of representing degrees of prior belief in the satisfaction of the conditions of application which it is intended to observe. The standard representation is in terms of subjective probabilities. For this reason, we adopt the mapping $\Psi(\Phi) \notin [0,1]$ which we interpret as the subjective probability that all conditions of application in Φ will be found to be satisfied.

By hypothesis, if all of the conditions of application of the theory are true, then the acts in A will imply some expected benefit $E(B|C)$. Otherwise, some different benefit $E(B|\neg C)$ will result. Since the benefit will be net of the cost of ascertaining whether conditions of application are satisfied, we define the benefit as $B = B\ (\Phi)$.

Evidently, the prior expected benefit of A when the set of conditions to be observed is empty is

$$E(B|\Phi = 0) = E(B|C)E(C) + E(B|\neg C)\bigl(1 - E(C)\bigr). \tag{A.1}$$

More generally, the expected benefit given any arbitrary set of conditions of application to be observed will be

$$\begin{aligned} E(B|\Phi) = \Psi(\Phi)\bigl\{&E(C|\Phi)E(B|C) \\ &+ \bigl[1 - E(C|\Phi)\bigr]E(B|\neg C)\bigr\} - c(\Phi). \end{aligned} \tag{A.2}$$

In equation (A.2), $c(\Phi)$ is the cost of observing the conditions in Φ. The expression $E(C|\Phi)$ is the expectation that all of the conditions of application are satisfied given that the individual conditions in the set Φ are known to be satisfied. Expanding equation (A.2), we get

$$\begin{aligned} E(B|\Phi) = {}&E(C|\Phi)\Psi(\Phi)E(B|C) \\ &+ \bigl[1 - E(C|\Phi)\bigr]E(B|\neg C)\Psi(\Phi) - c(\Phi). \end{aligned} \tag{A.3}$$

From the definition of C, $E(\Phi|C) = 1$.

Therefore,

$$E(C|\Phi)\Psi(\Phi) = E(\Phi|C)E(C) = E(C),$$

where the first equality is Bayes' Law. In consequence, equation (A.3) can be written

$$\begin{aligned} E(B|\Phi) = {}&E(B|C)E(C) + \bigl(1 - E(C)\bigr)E(B|\neg C) \\ &- \bigl(1 - \Psi(\Phi)\bigr)E(B|\neg C) - c(\Phi). \end{aligned} \tag{A.4}$$

Substituting into equation (A.4) from equation (A.1),

$$E(B|\Phi) = E(B|\Phi = 0) - \bigl(1 - \Psi(\Phi)\bigr)E(B|\neg C) - c(\Phi). \tag{A.5}$$

The interpretation of equation (A.5) is that, taking the case where no conditions of application are verified as the base case, the expected benefit is reduced by the expectation of benefit when the conditions of application are not satisfied and by the cost of ascertaining whether those conditions are satisfied. If this is true for every possible combination of conditions of application, then it is rational never to test for model applicability.

Evidently, the rational modeller who accepts the implications of economic theory for rational behaviour and cost-benefit analysis will investigate the conditions of application of the model only when

$$\bigl(1 - \Psi(\Phi)\bigr)E(B|\neg C) \geqslant -c(\Phi). \tag{A.6}$$

[3] This appendix is taken entirely from Moss [22].

indicating that the expected benefit of the policy implied by an inapplicable model is negative to an extent which is greater in magnitude the costs of determining whether the model is applicable. This result accords with economic reasoning.

References

[1] M. Allais, Le Comportement de l'Homme Rationnel Devant le Risque: Critique des postulats et Axiomes de l'Ecole Americaine, Econometrica 21 (1953) 503–546.

[2] J.R. Anderson, *Rules of the Mind* (Lawrence Erlbaum Associates, Hillside, NJ, 1993).

[3] K.J. Arrow and G. Debreu, Existence of an equilibrium for a competitive economy, Econometrica (1954) 265–290.

[4] K.J. Arrow and F.H. Hahn, *General Competitive Analysis* (Oliver and Boyd, Edinburgh, 1972).

[5] S. Bernow, A. Rudkevich, M. Ruth and I. Peters, A Pragmatic CGE model for assessing the influence of model structure and assumptions in climate change policy analysis, Report No. #96-190, Tellus Institute Boston (1998).

[6] K. Binmore, M. Piccione and L. Samuelson, Bargaining between automata, in: *Simulating Social Phenomena, Lecture Notes in Economics and Mathematical Systems*, eds. R. Conte, R. Hegselmann and P. Terna (Springer, Berlin, 1997) pp. 113–132.

[7] K. Carley, M. Prietula and Z. Lin, Design Versus Cognition: The Interaction of Agent Cognition and Organizational Design on Organizational Performance 1 (1998); http://www.soc.surrey.ac.uk/JASSS/1/3/1.html.

[8] P. Checkland, *Systems Thinking, Systems Practice* (Wiley, Chichester, 1993).

[9] J.B. Clark, *The Distribution of Wealth* (Macmillan, New York, 1893).

[10] G. Debreu, Excess demand functions, J. Mathematical Economics 1 (1974) 15–23.

[11] S. Funtowicz and J. Ravetz, Science for the post-normal age, Futures 25 (1993) 735–755.

[12] G.C. Harcourt, *Some Cambridge Controversies in the Theory of Capital* (Cambridge Univ. Press, Cambridge, 1972).

[13] K. Hasselmann and S. Hasselmann, Multi-actor opimization of greenhouse gas emission paths using coupled integral climate response and economic models, in: *Earth System Analysis – Integrating Science for Sustainability*, eds. H.-J. Schellnhuber and V. Wenzel (Springer, Berlin, 1998).

[14] W.P. Hogan, Technical Progress and Production Functions, The Review of Economics and Statistics 40 (1958) 407–411.

[15] D. Kahneman and A. Tverski, Prospect theory: an analysis of decision-making under risk, Econometrica 47 (1985) 263–291.

[16] N. Kaldor, Speculation and Economic Stability, *Review of Economic Studies* [reprinted in: *Kaldor Essays on Economic Stability and Growth*] (Gerald Duckworth & Co., London, 1939 [1960]) pp. 17–58.

[17] R.G. Lipsey and K.J. Lancaster, The general theory of the second best, Review of Economic Studies 24 (1956) 11–32.

[18] R.J. Mantel, On the characterisation of aggregate excess demand, J. Economic Theory 7 (1974) 348–353.

[19] Marshall, *Principles of Economics* (Macmillan, London, 1893 [1961]); (8th ed. (variorum edition) published in 1961).

[20] S. Moss, *An Economic Theory of Business Strategy* (Basil Blackwell, Oxford, 1980).

[21] S. Moss, H. Gaylard, S. Wallis and B. Edmonds, SDML: a multi-agent language for organisational modelling, Computational and Mathematical Organization Theory 4 (1998) 43–69.

[22] S. Moss, The Economics of positive methodology, in: *Economics as Worldly Philosophy*, eds. R. Blackwell, J. Chatha and E. Nell (Macmillan, New York, 1992).

[23] S. Moss, Critical incident management: an empirically derived computational model, Journal of Artificial Societies and Social Simulation 1 (1998); http://www.soc.surrey.ac.uk/JASSS/1/4/1.html.

[24] S. Moss, A cognitively rich methodology for modelling emergent socioeconomic phenomena, in: *Computational Techniques for Modelling Learning in Economics*, ed. T. Brenner (Kluwer Academic, Amsterdam, 1999).

[25] S. Moss, Applications-centred multi agent systems design (with special reference to markets and rational agency), in: *Proc. 4th Int. Conf. Multi Agent Systems*, IEEE (2000).

[26] A. Newell, *Unified Theories of Cognition* (Harvard Univ. Press, Cambridge, MA, 1990).

[27] W. Nordhaus, *Managing the Global Commons. The Economics of Climate Change* (MIT Press, Cambridge, MA, 1994).

[28] N. Oreskes, K. Shrader-Frechette and N. Belitz, Verification, validation and conformation of numerical models in the Earth sciences, Science 263 (1994) 641–646.

[29] C. Pahl-Wostl, *The Dynamic Nature of Ecosystems: Chaos and Order Entwined* (Wiley, Chichester, 1995) p. 288.

[30] C. Pahl-Wostl, Integrated assessment of regional climate change and a new role for computer models at the interface between science and society, in: *Prospects for Integrated Environmental Assessment: Lessons Learnt from the Case of Climate Change*, eds. A. Sors, A. Liberatore, S. Funtowicz, J.C. Hourcade and J.L. Fellous, European Commission, DG XII, Toulouse (1996) pp. 156–160.

[31] C. Pahl-Wostl, C.C. Jaeger, S. Rayner, C. Schär, M. Van Asselt, D.M. Imboden and A. Vckovski, Integrated assessment of climate change and the problem of indeterminacy, in: *Views from the Alps: Regional Perspectives on Climate Change*, eds. P. Cebon, U. Dahinden, H.C. Davies, D.M. Imboden and C.C. Jaeger (MIT Press, Cambridge, MA, 1998) pp. 435–498.

[32] G. Porter and H. Livesay, *Merchants and Manufacturers* (Johns Hopkins Univ. Press, Baltimore, MD, 1971).

[33] R. Radner, Competitive equilibrium under uncertainty, Econometrica 10 (1968) 31–59.

[34] J. Rotmans and H. Dowlatabadi, Integrated assessment of climate change: evaluation of methods and strategies, in: *Human Choice and Climate Change*, eds. S. Rayner and L. Malone (Battelle Press, Columbus Ohio, 1998).

[35] A.H. Sanstad and L.A. Greening, Economic models for climate policy analysis: A critical discussion, Environmental Modelling and Assessment 3 (1998) 3–18.

[36] S. Shackley, P. Young, S. Parkinson and B. Wynne, Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools?, Climate Change 38 (1998) 159–205.

[37] A. Shaikh, Laws of production and laws of distribution: the humbug production function, The Review of Economics and Statistics 56 (1974) 115–120.

[38] R. Solow, Technical change and the aggregate production function, The Review of Economics and Statistics 39 (1957) 312–320.

[39] R. Solow, Laws of production and laws of distribution: the humbug production function: a comment, The Review of Economics and Statistics 56 (1974) 121.

[40] D. Tillman, T. Larsen, C. Pahl-Wostl and W. Gujer, Modelling the actors in water supply systems, Water Science and Technology 39, 203–211.

[41] M. Ye and K. Carley, Radar Soar: towards an artificial organization composed of intelligent agents, Journal of Mathematical Sociology 20 (1995) 219–246.