Validity is a matter of confidence-but not just in system dynamics

Article

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Validity is a Matter of Confidence – but not just in System Dynamics

David C. Lane[1]
Henley Business School, England

INTRODUCTION

The paper by Nistelrooij *et al.* (2015) describes some of steps taken with a client to ensure that the model being developed was something that they could feel confident using. Their title elegantly entwines two ideas; that their modelling involves eye treatment and that the notion of model validity is a fundamentally subjective construct involving the opinion of specific individuals. It is this later point idea that I will explore in this piece.

SYSTEM DYNAMICS STANDS ALONE?

Trained as a mathematician, I came to the world of System Dynamics with a different notion of 'validity'. I looked at models and learned proofs that were 200, 300, 400 years old, developed by people from cultures I could barely understand. Today, still, I can read a proof by Archimedes (and I

---

[1] Correspondence to: Prof David C Lane, Henley Business School, Whiteknights, Reading, RG6 6UD, England. E-mail: d.c.lane@henley.ac.uk

heartily recommend this; try Heath, 2002) and see that we share (i.e. I was carefully educated to have) the same notion of universal truth.

When I came across System Dynamics in the 1980s I found something different: the idea that validity was all about having 'confidence' in a model, that is, making sure that the people trying to use a model to formulate policy would be confident in so doing. This was new to me. Moreover, from my reading it seemed that System Dynamics stood in splendid isolation in this stance. The position seemed to be that System Dynamics was a special case, that it had deliberately and necessarily distanced itself from many of the ideas of the broader Operational Research / Management Science (OR/MS) community.

However, as I read further I discovered that others took this approach too. From the world of decision analysis, I came across Phillips' notion of a 'Requisite Model', the idea that what needed to be in a model was simply what was required by a specific group of people to solve a specific problem at a specific time (Phillips, 1984). This was similar to 'confidence', surely? So not completely isolated. When I read Checkland (1981) I saw the link between the 'confidence' idea and his position that the most interesting things to model were socially constructed, the result of inter-subjective agreement. The isolation of System Dynamics eroded further. Other 'problem structuring methods' had a similar stance, and I argued that engagement with these ideas could help System Dynamics (Lane, 1994).

As I then read further into the literature of simulation I discovered that similar ideas were in use there. The situation was made hazy by the forest of different words used to describe very similar things but the message was clear: though there were detailed aspects of System Dynamics that required particular treatment of 'validation', the core idea of creating 'confidence' was something that System Dynamics shared with a range of affine techniques.

In this Discussant's piece I want to present that argument. Not with the aim of belittling System Dynamics for its lack of specialness but rather to strengthen its views on validation. The chief means of doing this follows in the next section: a brisk journey through three or four key decades of OR/MS research on validation. These are the focus because they show significant changes in our conceptualisation of 'validation'.

VALIDATION IN MAINSTREAM SIMULATION: EMERGENCE AND TRANSITION

The first significant problem confronting a researcher trying to understand how the concept of validity evolved in mainstream simulation is the bewildering and non-standardised range of different terms that are in use. We might all readily agree that we want a model to be 'valid', however, as soon as we look for an actual definition of that word a forest of terms springs up and leaves you stumbling through thickets of synonyms and variants. We ask that models are: 'verified', 'realistic', 'reliable', 'credible', 'effective', 'convincing', 'legitimate', 'acceptable', 'accurate', 'good', 'apt', 'plausible', 'implementable', 'representative', 'trustworthy', 'useful', 'usable' or 'used'. And that is just trying to get an idea of what 'valid' means. When we move to specific tests of 'validity' - however defined – there are further difficulties. Volume is one source. Balci and Sargent, two authors who have contributed insightfully to the area, were able to identify more than 300 tests (Balci & Sargent, 1984). As they observe, there is no standard terminology. Perhaps more

importantly, there is little clarity on what constitutes a valid model or on what is the one way to validate a model. This says something about the significance, complexity and subtlety of the validity issue. However, there is a route through this tangle.

To give a sense of this terrain I will re-employ terms from Lane (1995) which help to identify the level of operation of any comment regarding validation. The terms macro, meso and micro are useful. 'Macro' level statements concerning validity operate at a philosophical or theoretical level. They seek to establish the epistemological basis of knowledge claims and thereby commit to a formal model of science. 'Meso' level statements operate in a more practical register, involving an operational perspective which aims to clarify what, in general terms, is meant by the concept of validity. 'Micro' level contributions take for granted implicit theoretical considerations in favour of applicability and practicality. The interest here is in the specific processes that one would actually go through in the quest for 'validity'.

Representativeness at the fore

Whilst the concept of validity was thought to be important from the very creation of OR, it is notable that it was considered to be readily understood. Working at the meso level, in Ackoff's founding work (1956) what is very clear is that a model should represent the system under study. The essence of validation was then about establishing that representativeness. However, Landry *et al.* (1983) note that even here, usefulness was an additional measure of validity but that that aspect was taken for granted - to the extent that it was barely mentioned.

A highly influential contribution came from Naylor and Finger (1967) as they focused on the predictive powers of models. Their meso concept of validity (they use the term 'verify' throughout) meant proving a model to be true and their 'multi stage approach' to validation drew from three strands of theory to achieve this. Rationalism should be used to choose model variables and parameters. However, these are strictly considered as hypotheses; to empirically 'verify' them, statistical techniques are proposed. The ultimate decision on model validity, the macro stance, is then based on predictions.

Landry *et al.* (1983) give an account of the softening of this position. When models began to be used for studying the consequences of different possible courses of action, they argue, validation with a predictive approach ceased to be so overpoweringly relevant and 'usefulness' was increasingly brought into the validation debate as a subjective measure. In their review of macro ideas in OR, Déry *et al.* (1993) see this shift from prediction to understanding as an important move from a critical rationalist, or falsificationist, approach to a utilitarian, instrumentalist philosophy.

However, a very strong representativeness line was again advanced by Fishman and Kiviat (1968): if a model describes some hypothetical system then no validation can occur. Operating at the meso level, their definition of validation sees it as a test of whether a simulation model reasonably approximates a real system. Consistent with that, their micro level contribution is the careful specification of a number of statistical tests that may be used to analyse model output. This view is repeated in Fishman (1973), the motivation for such tests being very clear since validation involves checking consistency with the true system as it exists before any change is made. The success of this validation establishes a basis for confidence in the results that the model generates under new

conditions. Rivett (1968) supports representativeness in his discussion of the OR/MS 'attitude'. However, when discussing simulation, he sees a tension between the realism of a simulation-based 'game' and its playability, thereby introducing a notion of usefulness.

Van Horn (1971) can be seen as making contributions at all three of the levels as he opened up discussion. He qualifies and extends the ideas of Naylor and Finger. His macro level contribution supports the rationalistic approach but makes the important observation that since models concern people, physical processes and organisational structures the model elements used to represent these will arrive with varying degrees of *a priori* 'confidence' (a word he uses repeatedly). He supports an empiricist stance but proposes that a good substitute is sensitivity testing. Matters become even more striking when we come to his meso assumptions. Validation, he argues, "is the process of building an acceptable level of confidence that an [insight] about a simulated process is . . . correct . . . for the actual process" (pp. 47-8). He concludes that, "Seldom, if ever, will validation result in a "proof" that the simulator is a correct or "true" model of the real process" and that "There is no such thing as "the" appropriate validation procedure. Validation is problem-dependent" (both p. 248). This is born out in his micro level contribution: he offers eight validation actions, each with a value-to-cost ratio, and concludes, "The real task of validation is finding an appropriate set of actions" (p. 257).

In contrast, Sagastri and Mitroff (1973) offer only the meso-assertion that establishing the degree of correspondence between a scientific OR/MS model and reality constitutes validation. Shannon (1975) further adapts the macro contribution of Naylor and Finger to produce a 'utilitarian' multi-stage approach: 'modified rationalism' (face validity of model structure - involving the use of observation, previous research, theory and intuition - to check that the assumptions make sense), empiricism (operationalised by Fishman and Kiviat's tests) and 'absolute pragmatism' (model usefulness in predicting behaviour). Shannon's meso stance is that there are no absolute proofs of validity, rather an extended process of accumulating evidence to ensure representativeness and credibility.

A noteworthy *crie de cour* regarding the limitations of representativeness outside the circle of modellers comes in the title of a paper by Watt (1977): "Why won't anyone believe us?" He diagnoses the failure of large-scale simulation models to influence public policy makers. Whilst citing 'The Limits to Growth' work (Meadows, Meadows, Randers & Behrens, 1972) as a significant exception, he suggests that few models have been validated in the sense of proving useful in influencing public policy because they are credible.

Usefulness emerges

What can be seen as emerging doubts are supported by Quade (1980) who cites the overarching 'belief that a model can be proved correct' and 'attempting to really simulate reality' as pitfalls of analysis. He argues instead for the need to balance realism with 'relevance' by focusing on an issue not a system. He re-crafts this idea by describing the need to ensure that a problem is appropriately conceptualised prior to model construction. He then proposes 'invalidation procedures' as a means of increasing confidence in a model

According to Sargent (1982), "The state of the art regarding model validation is unfortunately far from satisfactory" (p. 159). He offers a framework (not dissimilar to one proposed by Sagastri and Mitroff) for representing the relationships between different sub-types of validity: 'conceptual model', 'data' and 'operational' and also 'computer model verification' (the last is a widely used term for determining that a computer model does run as expected). These combine into a reasonably clear meso level conception of validity similar to those seen so far and there are no macro level considerations.

The 'usefulness' line of thought is very much to the fore in Gass (1983). He is primarily concerned with large models intended for use by non-technical decision-makers for policy-analysis on not yet existing - that is, future - systems. Issues of validity, assessment and utility naturally arise. He agrees with Quade's warning and supports Shannon's views. His meso comments are helpful: validation concerns the confidence that those outside the model building process have in a model; confidence evolves over time and confidence is an attribute not of the model but is a judgement made by users with a model purpose in mind. His hierarchy of tests shows an operational focus. Landry *et al.* (1983) offer another framework, based on Sargent. Finlay (1985) considers many forms of modelling in addition to simulation and draws on work by Gass. Verification involves checking that a program does what is required of it, validation involves the application of tests which increase the reasonable probability that the relationships in a model are appropriate and that the model can help to tackle specified real world problems; in other words, usability and perceived representativeness again. Moving down from this meso level he offers two detailed sub-measures.

Balci (1990; 1994) advances matters at the meso level, describing 'credibility assessment' but his prime contribution is a detailed and coherent structuring of the 10 stages and activities that make up the 'life cycle of a simulation study' and the development of a 'hierarchy of credibility assessment stages' which build up into his concept of 'acceptability of simulation results'. In the first paper he lists numerous (micro) validation tests that may be used and in the second he notes the applicability of 45 such tests to each of the 10 stages.

At the macro level, Pidd (1992) takes a hypothetico-deductive approach to the creation and testing of theories. However, he then supports a multi-stage approach because he identifies two contrasting approaches to validation. 'Black box' validation is used as a final stage when the workings of a model remain unknown to the end user and predictive power is the goal. The focus is on replicative validity (implemented using statistical tests) and predictive validity to confirm predictive power. Contrastingly, 'white/transparent box' validation concentrates on confirming a plausible structure and employs face validity and conceptual validity as tests. Representativeness plus usefulness are behind these ideas but Pidd's careful division is reaffirmed in his general concept of model validity, which he sees as testing whether a model is wholly adequate and appropriate for the task for which it is intended.

Usefulness to the fore?

  We return to the macro level and the study of Déry *et al.* (1993), even though this moves the discussion out of the field of simulation and into the broader practice of OR/MS. The instrumentalist, utilitarian approach to OR/MS had emerged by this time. Perceived

representativeness was now a means to an end; the true measure of validity was the usefulness of models as intellectual instruments. However, during the 1970s the notion of usefulness was increasingly seen as implying usefulness in a specific social context and this led to the relativist, or paradigmatic stance that the knowledge claims derived from a model were determined not solely by the relationships of the model but also by the social relationships within which it was built. It is from ideas such as these that we can say that problem structuring methods (PSMs) emerged, and interest shifted to the importance of these social relations. The work of Phillips and Checkland has been cited already but PSMs offer other views. For example, Eden and Sims (1979) argue that 'coercing' a management team's understanding of the world using the undoubted 'truths' of OR/MS is inappropriate and the other extreme of 'empathising' with their worldview so much that all notions of rigour and empiricism were lost is also discarded. Instead a 'negotiative' approach is advanced in which OR/MS ideas were used to elicit and structure a team's ideas and to formulate a course of action which both solved their agreed understanding of the problem and 'attended to social realities'. The latter involved proposing solutions that were implementable within the organisational culture that obtained and that were acceptable to the participants. Similarly, Friend and Hickling (1987) describe the importance of moving beyond 'agreeing' a course of action to being 'committed' to embarking upon it, whilst Checkland and Scholes (1990) describe the need to perform a stream of 'cultural analysis' which deals with 'norms, roles and values' in a situation and so create 'culturally feasible' changes.

The emergence of PSMs and the richness of its methods are described elsewhere (Checkland, 1972; Rosenhead, 1989; Eden & Radford, 1990; Lane, 1994; Mingers & Rosenhead, 2001). For our purposes two strands are important. Firstly, the attention given to the social context of, and responses to, modelling; problems are not 'solved' but 'finished' (Eden, 1987). Secondly, the dual need to view humans as capable of appreciating, interpreting and actually creating their own social realities and to supply modelling tools that support this. Checkland (1995) returns us to the key issue. In validating 'soft' OR/MS models or approaches it may only be necessary to agree that they are 'relevant' to a structured debate or 'defensible' as a means of illuminating a worldview. This validation approach would seem to presage the ultimate triumph of usefulness over representativeness.

Obviously treatments of validation in OR/MS do not end here. Contributions continue, at the macro level (e.g. Kleindorfer *et al.*, 1998), at the micro level (e.g. Leemis, 2004) and across all levels (e.g. Robinson, 2014). Nevertheless, what is presented here serves to show the diversity of views in play and prepares the ground for what follows.

VALIDATION IN SYSTEM DYNAMICS RECONSIDERED

In the light of the material presented in the previous section, we can now re-visit the idea of model validity in System Dynamics. It is by looking at the three levels of discussion that we see commonalities and differences.

At the macro, or theoretical, level, we find a degree of variation. For around a decade the theoretical stance was left implicit. Bell and Bell (1980) then performed a search for a suitable philosophical approach. Rejecting instrumentalism, refutationism is advanced as the appropriate

theory since causal models offer clear test points; causal and behavioural hypotheses could be stated and tested, an approach then detailed by Bell and Senge (1980). The strong practical thrust that System Dynamics shares with OR/MS led Forrester and Senge (1980) to the more balanced conclusion that Naylor and Finger's multi-stage approach to validation theory, with the macro level assumptions that go with it, was appropriate for System Dynamics.

From this subsequently grew what Lane (1999) calls an 'interpretive division' in approach. In philosophical terms the debate pivots around more and less objective interpretations of the notion of 'validation by confidence'. Some of the employment of statistical techniques (see Sterman, 1984), seem to aspire to the austerity of logical positivism: 'confidence' is created by such 'truly scientific' means provided by experts. Radzicki (1990), offering one of the most clear and detailed arguments in the literature, describes an alternative. Confirming the poor esteem in which economist hold System Dynamics, he diagnoses that a basic difference in research philosophy is located in the utilisation by the majority of economists of the logical empiricist approach, whilst System Dynamics can be seen as an example of pragmatic instrumentalism. Barlas and Carpenter (1990) similarly reject logical empiricism. However, they employ a careful reading of Forrester and his notions that model validation is achieved solely by owner confidence to support the proposal that a Quinian, relativistic approach to model validation is appropriate. The meso level corollary is that validation is a gradual process of building confidence in the usefulness of a model - inherently a social, judgmental, qualitative process: models cannot be proved valid but can be judged to be so.

So, whilst some might still locate System Dynamics within a functionalist social theory, there is an argument that an interpretative form of the approach is practised. The use of group intervention methods championed by Vennix (1996), and exemplified in the paper by Nistelrooij *et al.* (2015), rests on an inter-subjective stance at this macro level. What is now clear is that this debate at the macro level is similar to that of OR/MS (as described in Déry *et al.*, 1993). This is born out in a study of System Dynamics in which the social theoretic assumptions inferred from its practice are seen to stretch from objectivism and across social system theory into social action theory (Lane, 1999) and, arguably, are consistent with social theories which aim to integrate objective and subjective positions (Lane, 2001; Lane & Husemann, 2002; 2008).

Turning to System Dynamics' meso position, this has been clear from the start. The process of model building is a means of making a group's assumptions explicit and logically complete in order to facilitate individual and organisational learning (Forrester, 1961; 1968a; 1968b; 1969; 1971). The purpose is to aid systems re-design. As Sterman (1988) re-states it; "the primary function of model building should be educational rather than predictive . . . modeling [is] a process rather than . . . a technology for producing an answer" (p.165). Validation is spoken of in terms of the 'confidence' that those using the model have in it, that confidence being created by various tests which add to the model's 'plausibility'; "In the system dynamics approach validation is an on-going mix of activities embedded throughout the iterative model-building process" (Richardson & Pugh, 1981, p.311). Additionally, "The validity . . . of a model should be judged by its suitability for a particular purpose" (Forrester, 1961, p.115). A corollary is that the usefulness, and hence validity, of such models can only be appropriately judged in a personal way; "The evaluation of improved managerial effectiveness will almost certainly rest on a subjective judgement rendered by

managers in regard to the help they have received [from engaging with a model]" (*loc. cit.*). At the meso level, again, the links with broader OR/MS ideas are clear.

In contrast, it is at the micro level that the System Dynamics approach becomes quite distinctive. Two works supply the best information on the micro level tests recommended in the System Dynamics field for building confidence in models. A series of tests is proposed in detail by Forrester and Senge (1980) and they have been usefully organised into a table (and slightly extended) by Richardson and Pugh (1981).The general nature of these tests merits comment , even though this borders on a return to meso level considerations.

A System Dynamics model constitutes an assembly of causal hypotheses about the relationships between variables which then support time-evolutionary behaviour. The shorthand for this idea is: 'the right behaviour for the right reason'. As Forrester observes:

> *"The importance of justifying model detail rests on a fundamental working assumption . . . that if all the necessary components are adequately described and properly interrelated, the model system cannot do other than behave as it should. The converse is not true; an endless variety of invalid components and structure exist to give the same apparent system behaviour, but these incorrect structures usually would not open the way to better system designs."*
>
> <div align="right">(Forrester, 1961, p. 117)</div>

Behaviour tests are done in association with appropriate tests of structure. Building confidence in a model is a process of considering both the behaviour of the model and its structure in an iterative way, and so the model testing process may be considered as two related activities. First, the interacting variables are validated by judging whether an effective choice of variables has been made in order to express the desired activities and whether they have been interconnected well. Each detail of the structure must be examined, equation by equation, policy by policy since, "The individual expressions in a model should always have meaning in the context of the real system" (Forrester, 1961, p.129). Second, the confirmation of structure is then tied in with the behaviour of the model, this being required to have general characteristics close to that of any observed data.

As with the treatment of general simulation, the description here is not intended as an exhaustive account of validation in System Dynamics. Rather, it provides sufficient ground for the main conclusion to this piece.

CONCLUSION: SYSETM DYNAMICS IN GOOD COMPANY

We close by drawing together the two accounts presented here. At the micro level, System Dynamics can feel very different from other simulation techniques in its approach to validation. This is what made system dynamicists feel different, even unique, in their approach. This difference is largely an illusion. System Dynamics is a distinctive approach – something clearly acknowledged and appreciated by contemporary external reviewers when *Industrial Dynamics* first appeared (for examples, see Lane, 1997). The field has distinctive concerns about the structure of its models and how their resulting behaviour over time can be compared with data. These concerns are reasonable

and consistent with the information feedback approach which defines the field. These concerns also derive from the System Dynamics interest in using models to learn rather than predict (Randers, 1980); its interest in participative approaches and tackling big issues (Lane, 2010); and its quest for qualitative policy insights (Lane, 2012).

However, such distinctiveness at the micro level is no reason to ignore commonalities at the other two levels. At the macro level System Dynamics shows a range of approaches quite similar the back and forth of debate and emphasis in OR/MS. Quite, simply, there is a variety of stances available and none of those adopted by System Dynamics is unique. Finally, at the meso level, the System Dynamics focus on the notion of 'confidence' is completely in line with other simulation approaches, even other OR/MS approaches.

The conclusion follows directly. Whilst system dynamicists can feel proud that their field alighted on the 'confidence' approach to validation idea arguably before other parts of OR/MS, today the field shares that idea. What is needed now is a sense of the solidarity with those other approaches. System Dynamics does not stand alone but rather in the company of a range of effective and well-grounded simulation and other OR/MS approaches. It can be more confident in itself as a result.

REFERENCES

Ackoff RL. 1956. The Development of Operations Research as a Science. *Operations Research* **4**: 265-295.

Balci O. 1990. Guidelines for successful simulation studies. In *Proceedings of the 1990 Winter Simulation Conference* eds. O Balci, RP Sadowski and RE Nance, IEEE: Piscataway, NJ; pp. 25-32.

Balci O. 1994. Validation, verification and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research* **53**: 121-173.

Balci O and Sargent RG. 1984. A bibliography on the credibility assessment and validation of simulation and mathematical models. *Simuletter* **15**: 15-27.

Barlas Y and Carpenter S. 1990. Philosophical roots of model validation: two paradigms. *System Dynamics Review* **6**: 148-166.

Bell JA and Bell JF. 1980. System dynamics and scientific method. In *Elements of the System Dynamics Method* ed. J Randers, MIT Press: Cambridge, MA; pp. 3-22.

Bell JA and Senge PM. 1980. Methods for enhancing refutability in system dynamics modeling. In *System Dynamics. TIMS Studies in the Management Sciences Vol. 14* eds. AA Legasto, JW Forrester and JM Lyneis, North-Holland: Oxford; pp. 61-73.

Checkland PB. 1972. Towards a systems-based methodology for real-world problem solving. *Journal of Systems Engineering* **3**: 87-116.

Checkland PB. 1981. *Systems Thinking, Systems Practice*. Wiley: Chichester.

Checkland PB. 1995. Model Validation in Soft Systems Practice. *Systems Research* **12**: 47-54.

Checkland PB and Scholes J. 1990. *Soft Systems Methodology in Action*. Wiley: Chichester.

Déry R, Landry M and Banville C. 1993. Revisiting the issue of model validation in OR: an epistemological view. *European Journal of Operational Research* **66**: 168-183.

Eden C. 1987. Problem Solving or Problem Finishing? In *New Directions in Management Science* eds. MC Jackson and P Keys, Gower: Aldershot, UK; pp. 97-107.

Eden C and Radford J. 1990. Tackling Strategic Problems: The Role of Group Decision Support. Sage: London.

Eden C and Sims D. 1979. On the nature of problems in consulting practice. *Omega* **7**: 119-127.

Finlay PN. 1985. Mathematical Modelling in Business Decision-making. Croom Helm: London.

Fishman GS. 1973. Concepts and Methods in Discrete Event Digital Simulations. Wiley: New York.

Fishman GS and Kiviat PJ. 1968. The statistics of discrete-event simulation. *Simulation* **10**: 185-195.

Forrester JW. 1961. *Industrial Dynamics*. MIT Press: Cambridge, MA.

Forrester JW. 1968a. Industrial Dynamics - A response to Ansoff and Slevin. *Management Science* **14**: 601-618.

Forrester JW. 1968b. *Principles of Systems*. MIT Press: Cambridge, MA.

Forrester JW. 1969. *Urban Dynamics*. MIT Press: Cambridge, MA.

Forrester JW. 1971. "The" model versus a modelling "process" (republished 1985). *Dept. memo D1621-1, system dynamics group, MIT, republished in System Dynamics Review* **1**: 133-134.

Forrester JW and Senge PM. 1980. Tests for Building Confidence in System Dynamics Models. In *System Dynamics: TIMS Studies in the Management Sciences* eds. AA Lagasto, JW Forrester and JM Lyneis, North-Holland: Oxford; pp. 209-228.

Friend J and Hickling A. 1987. Planning under pressure: The strategic choice approach. Pergamon: Oxford.

Gass SI. 1983. Decision-aiding models: validation, assessment, and related issues for policy analysis. *Operations Research* **31**: 603-631.

Heath TL. 2002. The Works of Archimedes (republished edition of 1897 work of the same name and of 'The Method of Archimedes, recently discovered by Heiberg' from 1912). Dover: Mineola NY.

Kleindorfer GB, O'Neill L and Ganeshan R. 1998. Validation in Simulation: Various Positions in the Philosophy of Science. *Management Science* **44**: 1087-1099.

Landry M, Malouin J-L and Oral M. 1983. Model validation in operations research. *European Journal of Operational Research* **14**: 207-220.

Lane DC. 1994. With A Little Help From Our Friends: How system dynamics and 'soft' OR can learn from each other. *System Dynamics Review* **10**: 101-134.

Lane DC. 1995. The Folding Star: a comparative reframing and extension of validity concepts in system dynamics. In *Proceedings of the 1995 International System Dynamics Conference: Volume I - Plenary Program* eds. T Shimada and K Saeed, Gakushuin University and the International System Dynamics Society: Tokyo; pp. 111-130.

Lane DC. 1997. Invited Review and Reappraisal: 'Industrial Dynamics' by Jay W. Forrester. *Journal of the Operational Research Society* **48**: 1037-1042.

Lane DC. 1999. Social Theory and System Dynamics Practice. *European Journal of Operational Research* **113**: 501-527.

Lane DC. 2001. *Rerum cognoscere causas:* Part II - Opportunities generated by the agency/structure debate and suggestions for clarifying the social theoretic position of system dynamics. *System Dynamics Review* **17**: 293-309.

Lane DC. 2010. Participative Modelling and Big Issues: Defining features of system dynamics? *Systems Research and Behavioral Science* **27**: 461-465.

Lane DC. 2012. What Is A 'Policy Insight'? *Systems Research and Behavioral Science* **29**: 590-595.

Lane DC and Husemann E. 2002. Steering Away From Scylla, Falling Into Charybdis: The importance of recognising, simulating and challenging reinforcing loops in social systems. In *Entscheiden in komplexen Systemen* ed. PM Milling, Duncker & Humblot: Berlin; pp. 27-68.

Lane DC and Husemann E. 2008. Steering without Circe: Attending to reinforcing loops in social systems (2007 Jay Wright Forrester Award address). *System Dynamics Review* **24**: 37-61.

Leemis LM. 2004. Building Credible Input Models. In *Proceedings of the 2004 Winter Simulation Conference, Vol. 1* eds. RG Ingalls, MD Rossetti, JS Smith and BA Peters, Institute of Electrical and Electronics Engineers: Piscataway, NJ; pp. 29-40.

Meadows DH, Meadows DL, Randers J and Behrens WW. 1972. The Limits to Growth: A report for the Club of Rome's project on the predicament of mankind. Earth Island: London.

Mingers J and Rosenhead JV (eds.). 2001. *Rational Analysis For A Problematic World Revisited*. Wiley: Chichester.

Naylor TH and Finger JM. 1967. Verification of computer simulation models. *Management Science* **14**: B92-101.

Nistelrooij LPJv, Rouwette E, Verstijnen I and Vennix JAM. 2015. The Eye of the Beholder: A case example of changing clients' perspectives through involvement in the model validation process. *Systems Research and Behavioral Science* **32**: to appear.

Phillips LD. 1984. A Theory of Requisite Decision Models. *Acta Psychologica* **56**: 29-48.

Pidd M. 1992. *Computer Simulation In Management Science* (3rd ed.). Wiley: Chichester.

Quade ES. 1980. Pitfalls in formulation and modelling. In *Pitfalls of Analysis* eds. G Majone and ES Quade, Wiley: Chichester, UK; pp. 23-43.

Randers J. 1980. Guidelines for model conceptualization. In *Elements of the System Dynamics Method* ed. J Randers, MIT Press: Cambridge, MA; pp. 117-139.

Richardson GP and Pugh AL. 1981. Introduction to System Dynamics Modelling with DYNAMO (republished edition). Productivity: Cambridge, MA.

Robinson S. 2014. Simulation: The practice of model development and use. Palgrave Macmillan: Basingstoke.

Rosenhead JV (ed.). 1989. Rational Analysis For A Problematic World: Problem structuring methods for complexity, uncertainty and conflict. Wiley: Chichester.

Sagasti FR and Mitroff II. 1973. Operations research from the viewpoint of general systems theory. *Omega* **1**: 695-709.

Sargent RG. 1982. Verfication and validation of simulation models. In *Progress in Modelling and Simulation* ed. FE Cellier, Academic Press: London; pp. 159-169.

Shannon RE. 1975. *Systems Simulation: The art and science*. Prentice-Hall: Englewood Cliffs, NJ.

Sterman JD. 1984. Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica* **10**: 51-66.

Sterman JD. 1988. A skeptic's guide to computer models. In *Foresight and National Decisions* ed. L Grant, University Press of America: Lanham, MD; pp. 133-169.

van Horn RL. 1971. Validation of simulation results. *Management Science* **17**: 247-258.

Vennix JAM. 1996. Group Model-building: Facilitating team learning using system dynamics. Wiley: Chichester.

Watt KEF. 1977. Why won't anyone believe us? *Simulation* **28**: 1-3.