Fundamentals of Validation

Brian R. Gaines
Knowledge Science Institute
University of Calgary
Alberta, Canada T2N 1N4
gaines@cpsc.ucalgary.ca
http://ksi.cpsc.ucalgary.ca/KSI

Abstract

A methodological framework for validity is developed based on the separation of physical, mental and knowledge worlds. Each world has different criteria for truth and valid inference, and an system's interaction with each world may be evaluated using differing criteria. The framework presented in this article provides a place for all approaches to, and all aspects of, system development, and makes explicit the relationships between them. It is a general framework for the growth of human skills and knowledge, and also for the analysis of a mature science of the validation and verification of knowledge-based systems.

1 Introduction

The objective of this article is to provide a methodological framework within which to analyze notions of validity. It is an obvious question to ask any system developer "how can you show your system is valid?" However, it is rather less obvious how the counter-question "what do you mean by valid?" should be answered. Does validity come from being effective in operation, from emulating the behavior of human experts, from being derivable from generally accepted principles, or some other such criterion? What are the truth criteria for validity, what are the underlying presuppositions that make them truth criteria, and how can these be shown to be valid?

There is no single or simple answer to these questions, but as we develop increasingly complex information systems and combining human and computer systems, it is becoming important to clarify the notion of validity and provide it with clear, operational foundations that capture and structure its rich connotations.

2 Modeling and Control

Figure 1 shows the basic situation for which validity criteria are required. Some *system* interacts with some *world*. The use of the term 'system' is intended to encompass the full range of human and information systems, computer services, simulations, control systems, scientific theories, experts, expert systems, learners, computer-assisted learning systems, and so on.

If the system is primarily concerned with *modeling* some part of the world through its *perception* of it then validity considerations have to do with the extent of the *prediction* by the system's model of the behavior of that part of the world. If the system is primarily concerned with *controlling* some part of the world through its *action* upon it then validity considerations have to do with the extent of the *power* of the system's control to determine the behavior of that part of the world.

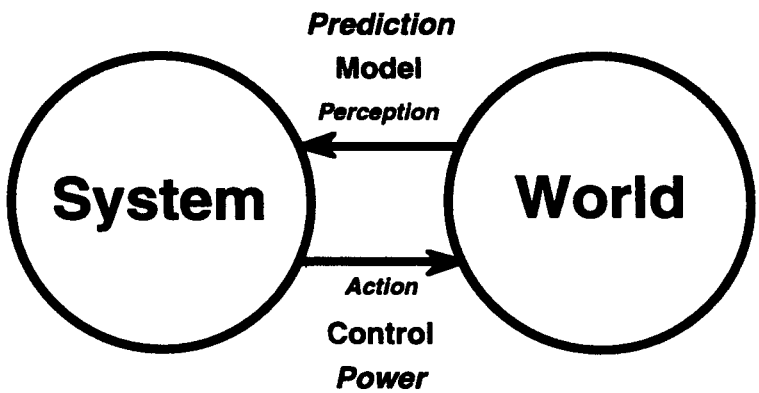

Figure 1 Validity criteria for a system modeling or controlling a world

Note that the system's evaluation of the power of the control that it exerting on the world comes only through its perception of the world. This is the fundamental principle underlying Powers (1973) modeling of behavior as the "control of perception." The system is controlling its perception of the world and hence its capability to control is critically dependent on the *construct system* (Kelly, 1955) through which it perceives the world.

In general, modeling can be analyzed as *construction without action*, and control as *construction with action*. One can subsume control and modeling into generalized modeling by thinking of control as modeling which changes the world to fit the model. However, in practical modeling it is usually necessary to take actions intended to gather data, stabilize the situation, explore different situations, and so on, while not changing the world being modeled, only the relation of the system to it.

Hence, it is also natural to subsume modeling into control, and differentiate them along a dimension that assesses how much the world is being changed through the modeling process. Following Kelly (1955), one may use the term *anticipation* to subsume both modeling and control.

3 Anticipating a World Containing Anticipatory Systems

If the world being anticipated (modeled or controlled) itself contains other anticipatory systems that model and control then anticipation becomes more complex. Anticipation of such systems has to take into account the ways in which their behavior changes with their experience, and it often appropriate to model an anticipatory system teleologically in terms of its goals rather than causally in terms of its input-output relations.

Hence, an important distinction made in anticipating a world is to separate the *mental* sub-world of other anticipatory systems from the *physical* sub-world of non-anticipatory systems. Other systems are recognized by a system as analogies of itself, themselves anticipating the world, modeling and controlling it in a way unlike causal physical systems. In addition other anticipatory systems act as role models, knowledge sources, and creators of new knowledge, in a way that makes the dimensions of interaction with the mental world very different from those of the physical world.

Figure 2 recognizes this separation and shows the system having communication with a *mental* world in which perception involves the comprehension of communication, and action involves communication to persuade other systems. The term 'mental' could be replaced with 'social' or 'cultural' in characterizing the world of systems. Colloquially, one uses the term 'mental' for the internal processes of a system, 'cultural' for the influence of other systems upon those processes, and 'social' for the coordinated behavior of groups of systems.

The separation of the mental and physical worlds as part of the modeling processes of a system is not necessary—systems can function without it—and it is appropriate to characterize a system which makes this separation as a *reflective* system. A reflective system recognizes that its anticipations are not causally necessitated but derive from choices that it makes, as do those of other systems.

It is possible to collapse the mental into the physical and treat the mental processes of systems as if they were the physical processes of objects, attempting to predict and control them. Habermas (1981) terms this 'strategic' social action, involving including the decisions and behaviors of other systems in a means-end calculation. Systems in the strategic mode relate to other systems as objectifiable means or obstacles to the attainment of their aims.

Reflective systems model the mental world is one of choice within a framework of established conventions so that other modes of interaction with it are possible. For example, a system may communicate to others what family of conventions it intends to adhere to. Habermas terms this 'dramaturgical' social action involving the projection of a public image.

Systems may operate together as a social unit within a common framework of conventions. Habermas terms this 'normative' social action in which the parties involved fulfill reciprocal expectations by conforming their behavior to shared norms and values.

Habermas' fourth category of social action, *communicative action*, is particularly interesting because it involves the systems involved attempting to reach explicit agreement for the sake of cooperation. This introduces the concept of mutual comprehension suggesting a further distinction that separates the shareable distinctions involved in communicative action from the less shareable distinctions involved in individual activities and social interaction without mutual comprehension.

Figure 3 extends Figure 2 to make this further distinction, showing the system interacting with a 'knowledge world' also, in which perception involves the derivation of knowledge from other knowledge, and action involves the capability to create new knowledge.

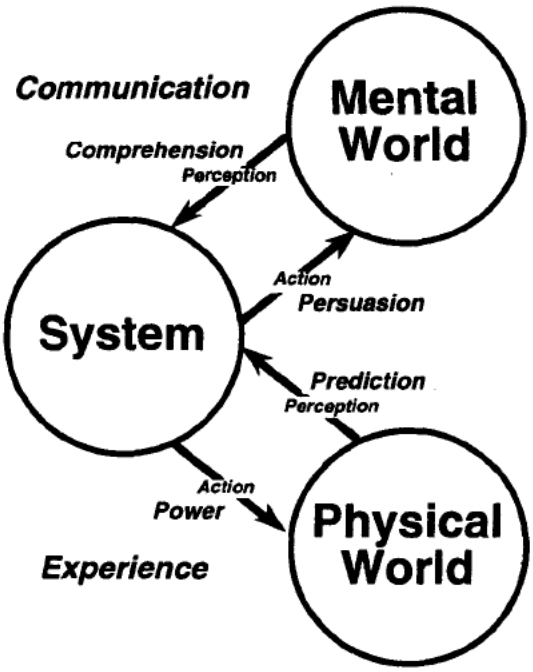

Figure 2 Separation of physical and mental sub-worlds

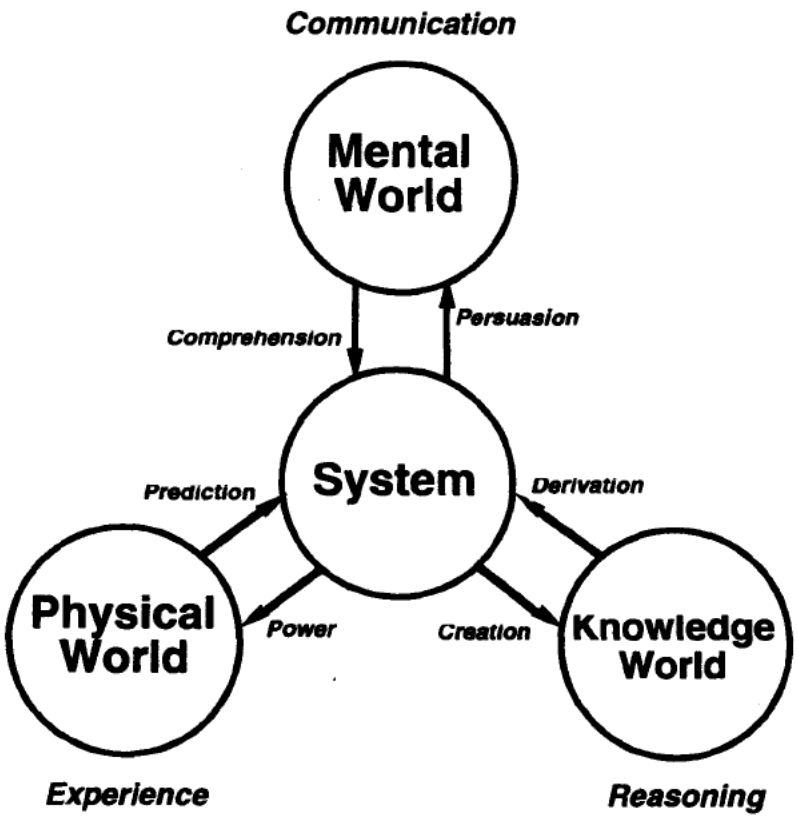

Figure 3 Separation of the physical, mental and knowledge sub-worlds

This knowledge world has many of the properties of Popper's (1968) *world 3* of *objective knowledge*. In particular, it is a product of the mental world, but distinguished by the extent to which it exists independently of the particular mental worlds from which it was created. There is an important conceptual link between Popper's criteria for objective knowledge and the existence of a social world valuing communicative action. The primary criterion is one of *fallibilism*—there must be tests that might show the knowledge to be false. However, this criterion which establishes the possibility of a critical community, also establishes the need for knowledge to be comprehensible so that it can be tested by the community (Notturno, 1985).

4 Validity Criteria in Physical, Mental and Knowledge Worlds

The separation into three worlds is significant for notions of validity because different criteria apply in each type of world. Validity involves truth criteria such that it can be decided whether a statement is true or false, and inference procedures which derive true statements from true statements.

The Physical World

The physical world arises from the assertion that there exist *necessary* distinctions, that is ones that are forced upon a system.

A key concept is that of *actuality*, that the distinctions made are actual.

The conceptual framework involved is that of *epistemology*, that the distinctions have to be drawn from the actual world.

Truth in the physical world is by *correspondence*, that the distinctions made correspond to the properties of physical systems.

Inference is *causal*, based on chains of physical cause and effect

The Mental World

The mental world arises from the assertion that there exist *chosen* distinctions, that is ones that are possible options for a system.

A key concept is that of *agency*, that distinctions have a maker.

The conceptual frameworks involved are those of *psychology* and *sociology*, that distinctions exist in the mental worlds of individuals and groups.

Truth in the mental world is *performative*, that the distinctions chosen will be continue to be chosen.

Inference is *conventionalist*, based on accepted chains of reasoning that are chosen within a culture.

The Knowledge World

The knowledge world arises from the assertion that there exist *independent* distinctions, that is ones that do not have to be ascribed to either the physical or mental worlds although they may be abstracted from them.

A key concept is that of *abstraction*, that systems of distinctions have their own existence.

The conceptual framework involved is that of *ontology*, that some distinctions are made independent of any criteria other than intrinsic structure.

Truth in the knowledge world is by *coherence*, that the systems of distinctions have an acceptable intrinsic structure.

Inference is *structuralist*, based on chains of reasoning that conform to the structure.

5 An Overall Model of Validity

Figure 4 extends Figure 3 to show these criteria within each world, and also shows the validity criteria appropriate to the relations between the worlds. It shows that the validation of a system depends on a number of different perspectives.

If the system is evaluated with regard to its interactions with the physical world it is in terms of its capability to predict and power to control that world.

If the system is evaluated with regard to its interactions with the mental world it is in terms of its capability to comprehend and power to persuade that world.

If the system is evaluated with regard to its interactions with the knowledge world it is in terms of its capability to derive knowledge from, and power to create knowledge in, that world.

If the system is evaluated as an emulator of existing relations between a system in the mental world and one in the physical world, then it is skill that is being evaluated and the criteria are pragmatic ones of whether it performs well, regardless of how it does it.

If the system is evaluated as an emulator of existing relations between a system in the mental world and one in the knowledge world, then it is rationality that is being evaluated and the criteria are hermeneutic ones of whether it reasons well.

If the system is evaluated as an emulator of existing relations between a system in the knowledge world and one in the knowledge world, then it is closeness to reality that is being evaluated and the criteria are those of positive science.

In practice, a system might be evaluated from all or any combination of these perspectives.

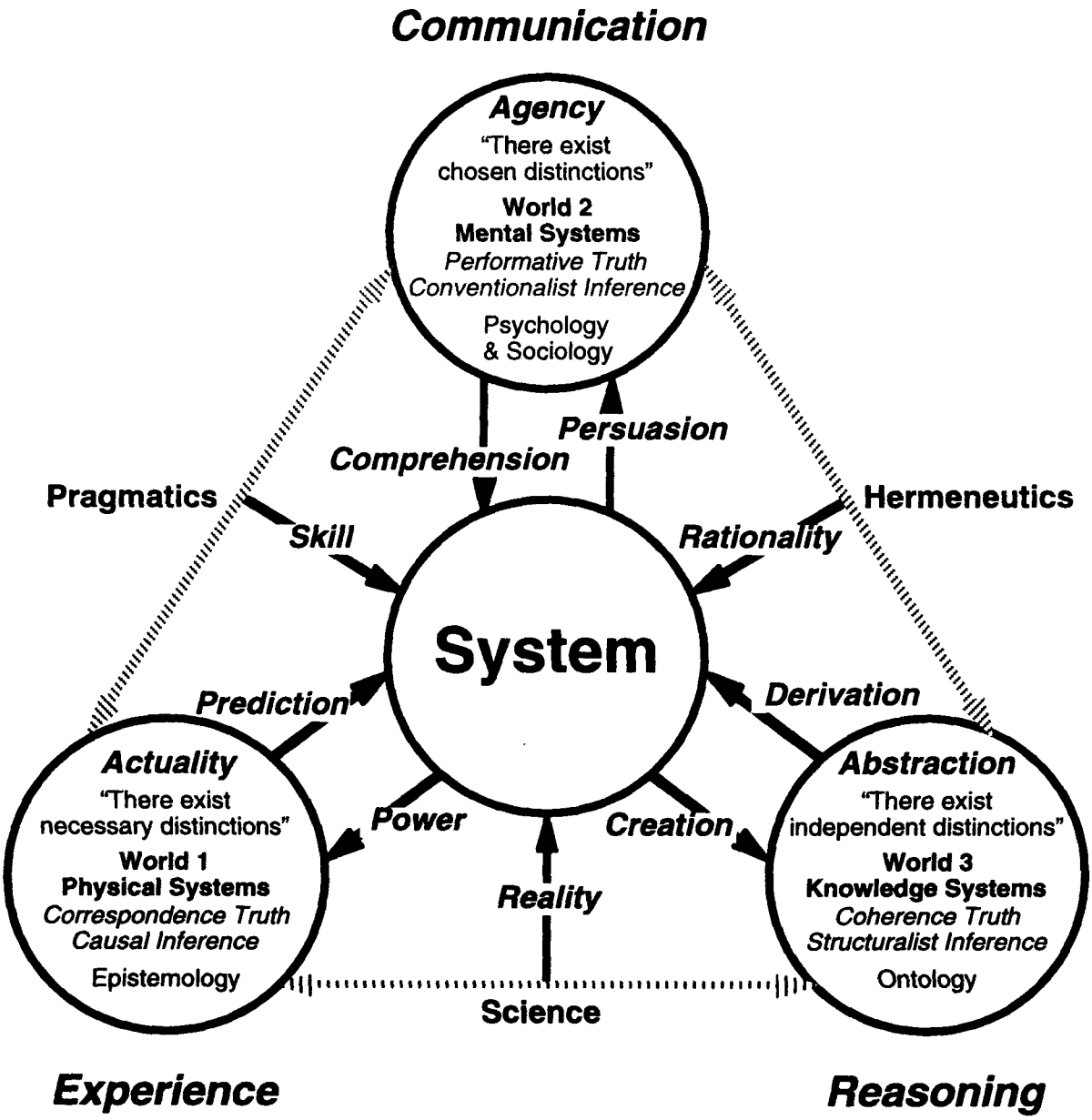

Figure 4 Dimensions of validation

5 Implications for Validation and Verification of Expert Systems

The analysis of validation developed above provides a methodological framework within which to classify studies of validation and verification of knowledge-based systems.

The expert systems paradigm may be seen as a shift in emphasis from the design of modeling and control systems from knowledge of the physical world, to design based on the emulation of human skills—that is, a shift from the *science* dimension at the bottom of Figure 4 to the *pragmatics* dimension on the left.

The assumption that expert systems would function as decision-support systems for people led to a secondary emphasis on *comprehension* and *persuasion* capabilities—that is, on question answering and explanation capabilities.

The pragmatic development of 'knowledge bases' that performed well through elicitation from human experts or learning through empirical induction of rules and neural networks does not usually result in knowledge systems that satisfy the coherence criteria of the knowledge world. Hence, much validation and verification research has focused on improving the internal coherence of knowledge bases whilst preserving their performance.

Additionally, even coherent knowledge bases generally lack derivation from underlying principles and thus lack status as 'scientific' models. Hence, another line of research is to develop the 'deep' knowledge underlying an effective knowledge base.

These last two developments replicate those in the study of human knowledge in general where formal logic is studied as the foundation of reasoning and reduction to basic principles is studied as the foundation of science. However, it is useful to remind ourselves that the 'knowledge level' is not an integral part of expertise. People perform complex tasks without explicit knowledge bases, let alone coherent or principled ones. People coordinate activities through communications that also are often not coherent or principled.

Abstraction to the knowledge level is a powerful and important technique but it is not always possible and it may be misleading. There may be many ways of making a knowledge base coherent or deriving it from principles and there may be no means of evaluating which of these is projectible in terms of future capabilities. The choice among the possibilities is itself a fallible inductive process.

6 Conclusions

With the growth in the literature on validation and verification in the past decade it would be timely to revisit the fundamentals of expert system development and assess the state of the art in all areas of validation. There is a danger that a shift in emphasis to 'knowledge-based systems' may reduce the domain to being one of information systems based primarily on logical inference rather than numerical calculation. There is a danger that emphasis on the internal coherence of knowledge bases may neglect considerations of comprehensibility of explanation.

The expert systems paradigm was originally intended to address those system development requirements where precise logical/numerical models were not available or derivable. It provided an alternative approach based on the emulation of human skill. It is important that this new paradigm not become lost in a quest for rationality based explicit and principled logical models.

This is not in the least to denigrate the knowledge level. Objective knowledge is a major human invention that has allowed human skills to be enhanced through technology, and the rationalization of systems that emulate human skills provides new insights into those skills, possibly allowing them to be emulated over more general domains.

The methodological framework presented in this article provides a place for all approaches to, and all aspects of, expert system development, and makes explicit the relationships between them. It is a general framework for the growth of human skills and knowledge, and also for the analysis of a mature science of the validation and verification of knowledge-based systems.

Acknowledgments

Financial assistance for this work has been made available by the Natural Sciences and Engineering Research Council of Canada.

References

Habermas, J. (1981). **The Theory of Communicative Action: Reason and the Rationalization of Society**. Boston, Beacon Press.

Kelly, G.A. (1955). **The Psychology of Personal Constructs**. New York, Norton.

Notturno, M.A. (1985). Objectivity, Rationality and the Third Realm: Justification and the Grounds of Psychologism: A Study of Frege and Popper. Dordrecht, Holland, Martinus Nijhoff.

Popper, K.R. (1968). Epistemology without a knowing subject. Rootselaar, B. Van, Ed. **Logic, Methodology and Philosophy of Science III**. pp.333-373. Amsterdam, North-Holland.

Powers, W.T. (1973). **Behavior: The Control of Perception**. New York, Aldine.