What is a computational social model anyway?: A Discussion of Definitions, a Consideration of Challenges, and an Explication of Process

Jessica Glicken Turnley, Ph.D. and Aaron Perls
Galisteo Consulting Group Inc.

September 2008

Defense Threat Reduction Agency
Advanced Systems and Concepts Office
Report Number ASCO 2008-013

Consulting
Group
Inc.

What *is* a computational social model anyway?:

A Discussion of Definitions, a Consideration of Challenges, and an Explication of Process

Jessica Glicken Turnley, Ph.D.*
jgturnley@aol.com
and
Aaron Perls
*Corresponding author

September 2008

Galisteo Consulting Group, Inc.
2403 San Mateo Blvd. NE
Suite W-12
Albuquerque, NM 87110
Voice 505.889.3927
Fax 505.889.3939
www.galisteoconsulting.com

We would like to thank Jennifer Perry of the Defense Threat Reduction Agency's Advanced Systems and Concepts Office (DTRA /ASCO) for her contributions and support to this piece. She provided creative ideas and thoughtful comments and critiques that added to and improved the substance of this work, including strong contributions to the problem formulation and the research design. She also provided outstanding project management, keeping us on our toes while smoothing out road bumps.

EXECUTIVE SUMMARY

There has been rapidly increasing interest by the national security community in the use of computer-based models of social phenomena. This discussion is an effort to make explicit the process of computationally modeling of social phenomena, and to challenge aspects of it. We define 'models,' and discuss issues regarding socio-cultural data, verification and validation, the role of theory, and the social roles necessary on every modeling team. In so doing, we illustrate some of the limitations of computational social models while identifying areas where they can serve important creative roles, identify points in the model-building process requiring greater rigor, and describe some possible research investment areas.

A model is not *all* objects and *all* relations in the target domain but a selection from them. If it were all objects and all relations, it would not be a model but would *be* the target domain. *This implies choice by someone in the model development process of which part of the domain to re-present through the model*. We suggest that this logic of choice of elements for inclusion in a model is analogic one (i.e. is based on analogy) and is expressed in theory.

An analogy is a relationship that posits that the parts and relationships of one system are 'like' the parts and relationships of a second. We call these analogic structures 'type A models.' A type A model is instantiated in time and space by populating it with 'real world data' which yields a type B model. Every type B model (every model of a particular instance) is based upon at least one type A model (is structured by theory).

The theory (or type A model) at play will suggest a modeling approach (e.g. social networks if the emphasis is on understanding structure, agent-based if the focus is on the identification of rules for behavior). The modeling approach serves as a lens to further focus attention on certain parts of the target domain and not others. The presentation medium (e.g. computers) also should be the result of a deliberate choice as the medium chosen puts constraints on data and other model aspects. Narrative or qualitative data, for example, which is the form in which most cultural data is collected, cannot be utilized in computational models which are mathematically based. Often what has happened in these cases is that data that is available is converted to the appropriate type through the use of quantifiable surrogates. Accuracy is sacrificed for precision, giving the model user an inappropriate level of confidence in the model.

The classic way to assess the value of computational models is through verification and validation. Verification refers to the internal consistency or goodness of the model itself, whether or not the code does what it is supposed to do every time. Various tools have been developed to ascertain the 'goodness of fit' of the code performance. Validation has much greater variation in definition than verification. Abstractly, it has been defined as a determination of the accuracy of the representation of the target system. It is usually performed through experimentation.

We suggest that validation only applies to type A models (models of theory), while type B models (models of an instance) can only be verified. We must first ascertain that a type B model (which contains 'real' data) is correctly reproducing the relationships enumerated in the underlying type A model. This is a verification question, a check of internal consistency. If the consistency check holds and if the type B model does not 'look like' the 'real world' when it is exercised, three possibilities can be explored: 1) the data itself is questionable; 2) the fidelity of the conversion of

data to a format that can be computationally manipulated is flawed; or 3) the underlying theoretical models (the type A models) are incorrect. In third possibility, the modeling team is asking whether it has selected the 'correct' entities and relationships to use as the underlying structure. This is a question about model validity.

Given the tremendous difficulties of validating (in a classic sense) computational social models since we cannot do the necessary experiments on human populations, it will be useful to rethink the purpose of these models and so devise other measures of goodness. Rather than fail as predictive tools, it is possible that computational social models could succeed as creative stimuli. Analogies (and so models) identify that which is important, unimportant, or not yet known to be important or unimportant in the target domain. They thus help us see what may be, in effect, hidden in plain sight – they actually can color what we see and do not see in that domain. If we focus on the creative power of analogies (and so of models), we might arrive at a different mechanism of assessing goodness.

Our look at the modeling process led us to identify five key social roles in any modeling process. They are: the *questioner*, who poses the question which initiates the process and establishes the model's purpose; the *user,* who exercises the model in a particular socio-technical environment, *disciplinary or theoretical expert* who identifies the elements to include in the model and the relationships among them; the *data provider*; and the *model builder* who translates relevant theory and data into the chosen medium. In some environments, there may be a sixth role, the *funder*, who may be behaviorally distinct from the questioner and the model user.

Every one of these roles is exercised in the construction of any model, although a single individual may fill more than one role. Individuals filling these roles should be vetted for excellence against the standards of each.

We identified five general areas of concern in the realm of computational social modeling.

Absence of a clear, shared understanding of model purpose. A clear *and shared* understanding of the model's purpose among all players in the development process is critical for successful utilization of the model. This is often not the case, leading to the inappropriate application of a particular model (to an application outside the design scope) or to the development of a model that does not suit the use environment or answer the original question. Enhanced process rigor would address this problem.

An implicit and unsupported theoretical base. Many computational social models are constructed with little or no involvement by social theorists, and are built around heuristics developed by the model builders. We would argue strongly for a requirement for a formal description of the social theoretical underpinnings of any computational social model in its presentation.

Data-related issues. Data may be incomplete or inaccurate. There are no good ways to account for these uncertainties in the modeling process or in its output, an area ripe for research. Data also may be based on non-observables and/or collected in narrative or natural language form, and translated into quantitative representations. There currently is no assessment of the cost in accuracy of translating of this sort. Development of such cost assessment techniques would be useful – but much more useful would be research exploring the extension of computing capabilities into areas that do not require the conversion at all.

Absence of justification of modeling approach. It often appears that the modeling approach is a function of the expertise of the modeler rather than a consequence of the theory driving the model structure. The requirement for a formal justification of modeling approach as a measure of model completeness is a function of process rigor.

Over-emphasis on validation. Computational social models cannot be validated with the same rigor and according to the same tenants as computational models of physical phenomena. Therefore, they cannot be used as predictive tools. The expectations and purposes of these models need to be revised in accordance with the value the models do bring to the table—their ability to contribute in significant ways to informed judgment by providing insight and stimulating creative thinking. This area is ripe for further research and is essential if the goodness of the models is to be assessed in any consistent fashion.

While these issues and concerns we have raised are significant, they can be overcome through more rigor in the modeling process and presentation or through investment in research. Models of any sort can be powerful creative tools, helping us see the world in new ways. Computational social models can help us manipulate large data sets over time and space and replicate problem-solving processes in places where that may be helpful. And they, too, can provide a new lens on familiar phenomena.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

BACKGROUND AND PURPOSE

There has been rapidly increasing interest by the national security community in the use of computer-based models of social phenomena, commonly called computational social models. This interest has been catalyzed by changes in the national security environment that have shifted the emphasis of threat assessments from analyses of adversary capabilities to studies of their intentions or motivations. During the Cold War, we assumed an adversary with malevolent intent and were primarily concerned with how that intent could be expressed in ways that could do us harm (that is, in assessing capabilities). Today, for reasons beyond the scope of this discussion, we are interested in why people develop malevolent intent and, if they are hostile, how aggressively and by what means they are likely to express it. The multiplicity of adversaries we face today vice the relatively unitary adversary of the Cold War period adds to the complexity of today's national security environment. This interest in computational representations of social phenomena has been further stimulated by recent advances in computational capabilities and the development of new theories of collective interaction based on the mathematics of complexity. These advances have allowed analysts to apply computational power to what historically have been purely text-based analyses of human communities.

The computational social modeling field is relatively immature, particularly when compared to the use of computers to construct models of physical and some biological phenomena.[1] Not surprisingly, then, many practitioners in the field of computational social science trace their intellectual and technical roots to the physical, life, and computer sciences, mathematics, and other applied technical fields such as engineering, epidemiology, and systems ecology. Since computational manipulation of this sort (i.e. computer-based) is relatively new in the social science field, there is not a significantly large cadre of mature sociologists, anthropologists and other social scientists who engage with these types of models. As a result, many of the methodological assumptions that underpin the development and utilization of computer-based models in the physical and life sciences have been implicitly adopted in the computational social modeling world.

This discussion is an effort to make explicit many of these heretofore implicit assumptions, and to challenge their applicability to the process of computationally modeling social phenomena. We are not arguing that computational social models are useless. To the contrary - there are volumes of work illustrating their contributions. We hope to provide a delimitation of the other side of the ledger sheet – a discussion of some of the limitations and boundaries of these types of approaches. Some of these limitations are inherent in the nature of the tool (its reliance on quantitative data, for example), while others are functions of the immaturity or lack of social theory which inform the models. That said, this discussion also underscores that the use of the tool (the computational model) can be an important mechanism for the development and advancement of social theory.

[1] We emphasize that when we use the term 'computational social models' (and grammatical and stylistic variants) we are referring specifically to computer-based models. We recognize that the application of computational techniques to social phenomena has a history that predates the use of computers. However, the rapid computational and advanced data storage and manipulation capabilities provided by computers has allowed significant advances in some areas (such as social network approaches) and the emergence of others (e.g. complexity theory) that has led to qualitative changes in the field.

DISCUSSION OVERVIEW

Mary Hesse asks,

> Does [scientific] "explanation" imply an account of the new and unfamiliar in terms of the familiar and intelligible, or does it involve only a correlation of data according to some other criteria, such as mathematical economy or elegance?[2]

We answer Hesse's question by saying that scientific explanation does both. In the course of so doing, we will propose that explanation uses two broad classes of models, and will describe each. Each type answers different sorts of questions and so is subject to different measures of 'goodness.'

The bulk of this discussion focuses on unpacking the definition of 'model.' However, it is important to understand at this point that when we use the term 'model' without the term 'computational,' we are referring to what Merriam-Webster calls "a structural design."[3] This design can be expressed mathematically, computationally, in natural language or in physical form.

Both classes of models we will describe can be simultaneously present in any computational model. One class of model, which we shall call type A, tells a story about the world and the way it works. Theories are models of this type. The other, type B, attempts to develop predictions about the world *within the framework of a particular story* or type A model. Thus every type B model includes elements from type A models, usually implicitly, sometimes explicitly. For example, in the physical sciences we might have a statement such as 'force equals mass times acceleration.' We can write this same statement mathematically as $f = m$*a. In the social arena, we might have a statement such as 'ties defined by biology and marriage form socially significant structures.' This also can be described graphically as shown in Figure 1.

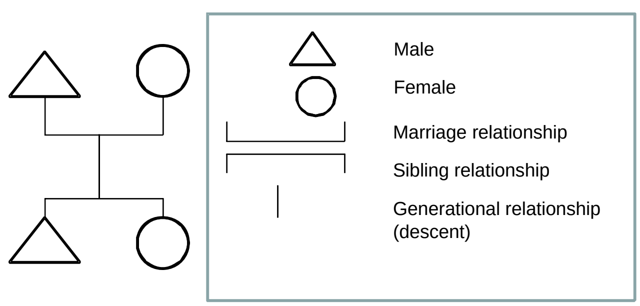

FIGURE 1: NOTIONAL KINSHIP CHART

Both $f = m*a$ and the notional kinship chart are stories about how the world works at an abstract or general level (i.e. not tied to a particular instance). We call this story (the logic structure and the elements) and its presentation (in our examples, in natural language [text], mathematically, and graphically) 'models.' These are models of the class we will call type A, and will describe in much more detail below. However, when we describe the trajectory of a particular object, or the specific kinship relationships that constrain the flow of social resources in a particular human group, we are presenting a model of the class we will call type B. We also will describe this in more detail below,

---

[2] Mary B. Hesse, *Models and Analogies in Science* (Notre Dame, IN: University of Notre Dame Press, 1966).

[3] http://www.merriam-webster.com/dictionary/model accessed May 2008

but note for our purposes now that in order to present our type B model, we must first *assume* the existence and truth of at least one type A model.[4] We thus can have a type A model without a type B – but never vice versa. Furthermore, we will suggest that in order to understand human action both classes are necessary but neither are sufficient, although different disciplines argue for the primacy of one or the other. We also will suggest that some of the difficulties surrounding the validation of certain computational social models arise from an incomplete understanding of the ways in which these two different classes of models are operationalized in a computational environment

There are two threads to our discussion. The first is a definitional discussion. It begins with the general notion of 'model.' We then refine the discussion by introducing first a detailed argument for our type A model, and then a similar argument for the type B model. We then move through the process of constructing computational social models, defining them as a model of a particular class of phenomena (socio-cultural) presented through a particular medium or vehicle (a computer). This leads us into a discussion of verification and validation of these types of models. The second thread is intertwined with the first. It addresses the different social roles (sets of expected behaviors) that are brought into play during the modeling process. This will further illuminate the nature of the model by illustrating the various disciplinary epistemologies that come into play during the construction of a computational model of social phenomena.

[4] The description of the trajectory of a particular object at some specific time at a defined point in space assumes $f = m*a$. To be a more complete description, it also must incorporate some assumptions about the medium in which the object is traveling, the presence/absence of other nearby objects, etc. In the same vein, the description of the kinship relations of a particular group of people at a particular time and place will incorporate considerations of the physical and social environment beyond kinship (e.g. markers of in-group membership, flows of power and other social resources, etc.), each of which is constrained by assumptions embodied in other type A models.

WHAT IS A MODEL?

In lay terms, a 'model' means something that re-presents something. It is not the thing in itself, for then it would *be* the thing and not a model. A map of the world is not the world. It is a re-presentation of selected elements and relationships in it.

There are three necessary components of the construct known as a 'model.' There is the 'something else,' i.e. the target domain or thing that is re-presented. This is usually some portion of the experienced world, e.g. a society, a family, the brain or an atom.[5] There is the representation, that is, what we call the model. And there must be some established relationship or connection between the two such that we say that the model is a model *of* one thing and not of another thing. This simple deconstruction is shown in Figure 2.

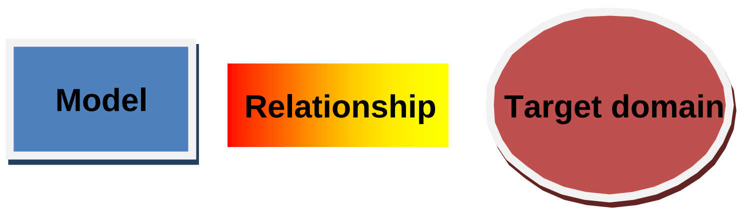

FIGURE 2: ELEMENTS OF THE 'MODEL' CONSTRUCT

SOME DEFINITIONS

It will be useful to pause here and define some of our terms. When we use the term 'model' without any modifiers it refers to the general class of re-presentations as shown in Figure 2, of which computational models are a subclass. A 'computational model' is a computer-based model (a re-presentation through a computerized medium) of any of several classes of phenomena (e.g. physical, biological, or socio-cultural) expressed mathematically. We will use the term 'computational social model' (or its appropriate grammatical variant) to mean a computer-based model of socio-cultural phenomena. Figure 3 shows this taxonomy.

[5] Note that this 'portion of the experienced world' can be something tangible or conceptual.

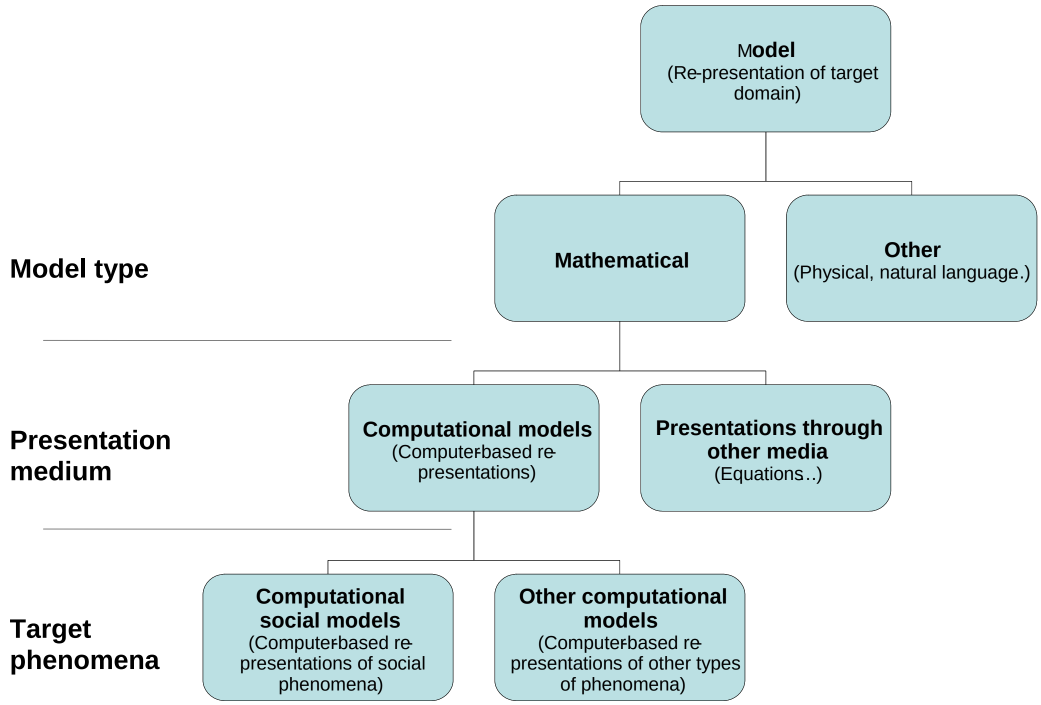

FIGURE 3: MODEL TAXONOMY

This section focuses on models (the highest level in the taxonomy), that is, that general class of re-presentations which includes (but is not limited to) computational social models. The remainder of the discussion will move us down to the lower left corner – computational social models.

We also need to describe our players. There are several social roles at play in the modeling process, where a 'role' is a set of socially defined expected behaviors.[6] Although a single individual may play several social roles during the construction of a computational social model, it is critical to keep the roles analytically distinct, and the notion of roles as sets of expected behaviors distinct from the individuals who play them. An individual can be evaluated on his ability to fill any given role. By extension, if he assumes multiple roles, he should be evaluated separately against the criteria for each.

Key roles in the modeling process are as follows: the questioner, the theoretical expert, the data provider, the model builder and model user.[7] As we explicate each of these roles, note that while a single individual may assume more than one role in any given process, all roles are present in any

---

[6] Robert K. Merton. 1949. Social Theory and Social Structure. The Free Press. New York, NY.p.116

[7] In a government environment there may be a sixth role, that of funder. The funder may add concerns about credibility (which are different than concerns about model validity) and accountability to the mix and may also dictate certain requirements regarding modeling approach or presentation medium.

modeling process. Fulfillment of the requirements of any particular role should be evaluated independently of the number of roles any given individual may fill.

The *questioner* poses the question which initiates the process. The modeling literature often refers to this as 'establishing the model's purpose.' Modeling literature and practice sometimes conflates the questioner with the *user* of the model. The user utilizes the model to serve some purpose. He may use only the output from the model to accomplish his purpose, may use insights gained through the construction of the model, and/or may combine what he gains from this model with other inputs. The user may inform the questioner and the roles of questioner and user may be played by the same individual in some cases but not others. In practice, interchange between the questioner and user often establish boundaries of required precision or accuracy (from the user to the questioner) or set requirements of theoretical rigor and justification (from the questioner to the user).

There is a role for a *theoretical expert*, as well as for a *data provider*. The theoretical expert brings to the project background in one or several social or behavioral science disciplines. These disciplines represent decades if not centuries of thinking, experimentation and observation on the structure of human interaction and behavior. Furthermore, since theory is contested in the social and behavioral sciences *and* since it is theory that drives model structure and therefore the type of data included (more on this later in the discussion), the selection of the theoretical expert can have a significant impact on the model structure and functioning. The data provider will have access to the articulation of these structures in a particular time and place – Weirdistan in the twenty-first century, for example. The data provider and theoretical expert roles are often conflated in the literature and in practice into the 'subject matter expert' (SME). In these instances, the data provider role often completely dominates the role of the theoretical expert, and the individual recruited to fill the role of data provider often *de facto* fills the role of theoretical expert whether or not he is qualified to do so. In other cases, the individual writing the code, developing the equations or constructing the narrative (the model builder) assumes this role. It is important to emphasize that although the theoretical expert's role often is invisible and completely implicit in the computational social modeling process, it is *always* present nonetheless and has a determining influence on model structure. We will argue strongly that the theoretical expert and the data provider are actually two separate roles which may be exercised by the same individual but which participate at different points and in different ways in the modeling process. Individuals filling each of these roles must be vetted for quality according to different criteria.

Finally, there is the *model builder,* the role that concerns itself with the translation of relevant theory into the chosen presentation format that will access and accommodate available data. In the case of computational models, this is a code builder. In the case of qualitative or narrative models, this would be an author. We have summarized these roles in Table 1.

Table 1: Roles in the modeling process

Role descriptor	Expected behaviors
Questioner	Establishes model purpose
User	Utilizes model for its intended purpose
Disciplinary or theoretical expert	Provides theoretical knowledge; constrains model structure
Data provider	Provides data relative to a specific instantiation of the theory
Model builder	Presents theory and relevant data in the chosen presentation medium

Note that most discussions of the construction of computational (social) models generally identify two or possibly three roles: model user, model builder (often called 'the modeler') and sometimes the subject matter expert (SME). The SME conflates the theoretical expert and the data provider. The 'model builder' is often conflated with the SME, with the model builder providing theoretical structure and/or data. We have identified five distinct roles: questioner, user, theoretical expert, data provider, and code builder. We show this alternative typology and its relationship to ours in Table 2.

TABLE 2: ALTERNATIVE TYPOLOGY FOR ROLES IN THE MODELING PROCESS

Role descriptor	Alternative typology	Expected behaviors
Questioner	Model user	Establishes model purpose
User		Utilizes model for its intended purpose
Disciplinary or theoretical expert	Subject matter expert, sometimes is conflated with the model builder	Provides theoretical knowledge; constrains model structure
Data provider		Provides data relative to a specific instantiation of the theory
Model builder	Model builder	Presents theory and relevant data into the chosen presentation medium

We will adhere to our terminology throughout this discussion, because we believe this parsing will help infuse rigor and quality into the process of the construction of computational social models.

We recognize that in practice a single individual may play several roles. An intelligence analyst is generally considered to be a user, for example. However, his discomfort with a 'black box' computational social model he is given to support his work usually comes not from his role as a user but from his role as a theoretical expert. The theory that defines the model's construction and operation is not transparent and so the analyst has no basis for judging its applicability, appropriateness or acceptability. The theoretical expert may also be the data provider, particularly in anthropologically-based models where theorists generally also have a regional specialization. Questioners and users also may be collapsed into the same individual, as is the case with (for example) a warfighter with a particular tactical or operational need.

TYPE 'A' MODELS

Our definition of roles involved in the modeling process begins our application of rigor to the process. We continue this development by returning to our simple notion of a model as presented in Figure 2 with its three elements: the target domain, the re-presentation of that domain (the model) and the particular relationship between them. The semantic or structuralist theory of models, which underlies most current discussions of models,[8] claims that "models are structures (where a structure, roughly speaking, is a collection of objects along with the relations in which they enter) and that they represent due to their being isomorphic to their target system."[9] However, we argue that a 'model' must be further defined by stating that it is not *all* objects and *all* relations in the target domain but a selection from them. If it were all objects and all relations, it would not be a model but would *be* the target domain. *This implies choice by someone in the model development process of which part of the domain to re-present through the model*. The locus and logic of this choice is rarely discussed either in the literature or in practice. We argue that the logic of this choice is a key element in the construction of the model, and must be made explicit.

We suggest that this logic of choice of elements for inclusion in a model is an analogic one (i.e. is based on analogy), where an analogy is an argument for isomorphism between part of one system and another, and is expressed in terms of a theory. We will go into the nature of analogies in greater detail later in this section. For now, we simply assert that an analogy is a relationship that posits that the parts and relationships of one system are 'like' the parts and relationships of a second. If we look again at Figure 2, and if the 'real world' is the target domain, the means of selecting elements and relationships from it to re-present in a model (the connection between the real world and the model) is analogy.

A theory is an analogy that we believe is more 'true' than another (more on this 'truth' later). It tells us which elements and relationships among them in the target domain are important or relevant. If our theoretical approach is one based on the notion of a rational man, the (perception of) costs and benefits and mechanisms of calculating tradeoffs will be of interest. If our approach is one driven by power of symbolic constructs, texts and other embodiments of affective meaning will be of interest to us. Affective rather than calculated social ties will be assumed to be the stronger in this latter case.

---

[8] See for example Richard Hull and Roger King. 1987. Semantic Database Modeling: Survey, Applications, and Research Issues. *ACM Computing Surveys* vol.19 no.3. Pp.201-260

[9] Roman Frigg. 2002. Models and Representations: Why Structures are not Enough. Center for Philosophy and Natural Sciences, Measurement in Physics and Economics Technical Paper 25/02, London School of Economics. London, Britain. P. 2

Analogy is the mechanism or logic; a theory is a statement of a particular analogy. An analogy explains the unknown in terms of the known. [10] An analogy "may be said to exist between two objects in virtue of their common properties."[11] To create an analogy, then, the theoretical expert must begin with a target system or domain which is not understood and an already known system or a system with known elements. He perceives some likeness between *objects* in a known or already understood system and those in the unknown system (the target domain). He also must perceive *relationships between objects* in the known system that are like relationships in the unknown system. Thus an analogy can be written as

$$\frac{\text{mass of a billiard ball}}{\text{acceleration of a billiard ball}} :: \frac{\text{mass of a gas molecule}}{\text{acceleration of a gas molecule}}$$

where the billiard ball is the known domain and the gas molecules the unknown.

There are several relationships embodied in this analogy. Stating it as an analogy posits that gas molecules are somehow 'like' billiard balls. This likeness consists in the possession of certain properties, i.e. mass and acceleration. It also consists in a particular causal relationship between mass and acceleration which can be defined in a certain way. The use of analogy posits that the same causal relationship holds for properties in both the billiard balls and the gas molecules. So there are posited similarities between properties and between relationships in the known and unknown domains.

We are particularly interested in predictive models. For an analogy to be predictive (or, to say it another way, falsifiable) not only must the objects and the relationships among them be similar, but the relationships among them must be causal. Models or analogies exhibiting relationships with such a causal nature we call theories (the causality allows prediction which allows falsifiability which is the hallmark of a scientific statement[12]). This eliminates from consideration as scientific models such analogs as

$$\frac{\text{father}}{\text{son}} :: \frac{\text{state}}{\text{citizen}}$$

This is a moral but not a scientific analog. There is no inherent causality in the father : son relationship.[13]

There are other types of analogical models that are not theoretical models. A scale model of the Eiffel Tower (a representational model), for example, extracts certain elements from the 'real' Eiffel Tower as relevant and not others. (For example, the composition of the materials may not be relevant while the proportional size of objects in/on the Tower is.) Idealizations, such as the 'rational man' assumption that underlies much economic theory or the moral statement we made earlier about fathers and sons and citizens and states also are analogical. It should also be made

[10] For clarity, analogies are statements such as 'a man is like a wolf,' or 'kinship relations are like power hierarchies' or 'gas molecules are like billiard balls.'

[11] Hesse. 1966. op.cit. p.58

[12] Karl Popper. 2002 (1934). The Logic of Scientific Discovery. Routledge Classics. New York, NY

[13] Hesse. 1966. op.cit. p.62

clear that type A models are not always formally stated and constructed. In fact, in many cases they are implicit and expressed as heuristics – but they are present nonetheless.

The full set of Type A models is as follows, although we are only concerned with theoretical models here.

Type A models (Analogic)

1. Representational
    1.1. Scale models
    1.2. Idealizations
2. Models of Theory

Table 3 annotates this list.

TABLE 3: TYPE A MODELS

Characteristics / Model type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
Analogical Models	Two things are analogous if there are similarities between them that are deemed relevant. Makes 'sense' out of the unfamiliar. Identifies key elements and relationships among them. Can answer a 'why' question by positing causal relations among elements.	Comfortable and generally familiar. Identifies and makes accessible key aspects of a phenomenon, either/both relationships or elements. Can be predictive Can create new knowledge, either through exploration of neutral analogy or by imputing characteristics of a familiar system onto an unfamiliar and causing us to 'see' the unfamiliar differently.	Generally imprecise. No attempt at reproducibility. There are no perfect analogies Initially subjective – system definition and identification of important elements is observer-determined until verified by type B model. Usually uni-dimensional, although dealing with multi-dimensional phenomena	Useful in completely unfamiliar situations Well suited for data-poor problems Can handle qualitative data easily	Things are more or less as we perceive them and intuit them. Meaning constructs become reified (analogies become metaphors).	Social science and other domains which deal with qualitative data Theory construction Analogical models use heuristics well: they allow modelers to make assumptions by supposing an analogy, rather than finding more data on the specific target

[14] Mary S. Morgan, "Imagination and Imaging in Model Building", *Philosophy of Science*, volume 71 (2004), pages 753–766

Characteristics / Model type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
Representational models	A representational model is simplification of phenomena, i.e. it represents what are perceived as core or essential aspects.	Representational models are relatively easy to interpret and are very accessible.	As idealizations or oversimplifications of the target domain, representations are never a complete picture, i.e. they illustrate what has already been assumed to be *true*.	Problems where the model must be quickly and easily grasped	Which aspects to include in the simplification must be decided *a priori*; generally the decision logic is implicit.	Simple objects and systems, i.e. systems and objects containing few parts or aspects that interact or function in an intuitive manner.
Scale models	A down-sized or enlarged representation of target domain	Will appear to closely resemble the target domain, thus making the target domain accessible and real.	See Representational Models above. There is no such thing as a perfectly accurate scale model	See Representational Models above.	See Representational Model above. It is often assumed that a scale model is an entirely accurate representation, with a scale change.	Simple objects and systems. In tandem with another model.
Idealization	An idealization is a deliberate simplification of something complicated, with the objective of making it more tractable.	Idealizing phenomena always makes them more intuitively accessible, and more easily mathematized.	See Representational Models above By definition, in treating a phenomenon as an idealization, there are aspects of the target being neglected.	See Representational Models above	If the striping away of what has determined to be non-essential is implicit, an idealization may be taken for the thing itself.	Situations where heuristics are sufficient Quickly simplifying the complex

[15] J.E. Summers, R.F. Gragg, and R.J Soukup. 2006. “Topography measurement of scale-model representations of the rough ocean bottom by touch-trigger probe and its implications for spectral characterization.” Oceans. (September . 2006) Pp. 1-6.

Characteristics / Model type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
Models of theory	Presents a story or description of the world in terms of causal relationships	Theories are generally meaningless without a model associated with them to describe the interactions of phenomena which are bound up in a given theory.	The contingent nature of the *theory* can be lost, and the theory misinterpreted as *fact*. If a model in theory is self-consistent and well made, this means it is a *good model*, not necessarily a *good theory*. The goodness of the theory rests in its ability to be validated.	'Why' questions Problems embodying causal relationships Illustrates a foundation upon which a potentially predictable model may be constructed.	Assumptions vary from theory to theory. Most theories have even deeper theoretical underpinnings (are contingent on other theories) that might not be made explicit within the model.	Making sense of unfamiliar situations Discovery

CONSTRUCTING ANALOGIES AND DEVELOPING THEORY

Returning to our discussion of analogy, we find that there are different types of analogy that are can illuminate the way models work. Hesse characterizes the relationship of things from the system that we know to those which are like it from the unknown system a *positive* analogy. There also are properties of billiard balls that we do not want to ascribe to gas molecules - color, for example. Hesse calls this relationship (these deliberately excluded elements) a *negative* analogy. Finally, there are elements about which we may be undecided, i.e. we are unsure if the isomorphism exists. She calls this relationship a *neutral* analogy.[16] A model is the re-presentation of the positive and neutral analogies.

> "The model is the imperfect copy *minus the known negative analogy*, so that we are only considering the known positive analogy, and the (probably open) class of properties about which it is not yet known whether they are positive or negative analogies"[17] (emphasis in original).

Hesse and others argue that it is in the neutral analogy that the creative powers of models lie. By exploring the goodness of fit of these elements about which we are undecided (the neutral analogy), we can learn more about the target domain. If, for example (to stay with our billiard ball analogy), we are undecided about whether or not shape is important to our understanding of how gas molecules act, we will construct a type B model (a model of a particular instance) to determine if shape is a positive or negative analogy. We will return to this point later in our discussion of type B models.

If we revisit Figure 2 with its three elements – the target domain, a representation of the target domain (the model) and the particular relation between them, we are suggesting that the relationship is an analogical one. Those involved in the modeling process, interested in addressing a problem through the creation of a re-presentation, must pick part of the world to re-present. They must create, as Margaret Morrison and Mary S. Morgan put it, a model which is "a partial representation that either abstracts from, or translates into another form, the real nature of the system..."[18]

If they do not make their logic of choice explicitly, we must impute – and then judge the appropriateness of – such a logic. Morgan noted that "…modeling requires making certain choices…"[19] This is a powerful statement, for it implies that those involved in the modeling process, through the construction of an analogy, can shape what part of the 'real world' (the target domain) we do and do not see. Thomas Kuhn, in his groundbreaking study of paradigms, provides an anecdote which illustrates the world-shaping force of such constructions.

---

[16] Hesse. 1966. op.cit. p.8

[17] Ibid. p.9

[18] Margaret Morrison and Mary S. Morgan. "Models as mediating instruments" in Models as Mediators Mary S. Morgan and Margaret Morrison, eds. Cambridge University Press. Cambridge, U.K. pp.10-37. p.27

[19] Mary S. Morgan. 1999. "Learning from Models." in Models as Mediators Mary S. Morgan and Margaret Morrison, eds. Cambridge University Press. Cambridge, U.K. pp.347-388. p.386

> "An investigator who hoped to learn something about what scientists took the atomic theory to be asked a distinguished physicist and an eminent chemist whether a single atom of helium was or was not a molecule. Both answered without hesitation, but their answers were not the same. For the chemist the atom of helium was a molecule because it behaved like one with respect to the kinetic theory of gases. For the physicist, on the other hand, the helium atom was not a molecule because it displayed no molecular spectrum. Presumably both men were talking of the same particle, but…their experience in problem-solving told them what a molecule must be."[20]

And as Claude Lévi-Strauss, a French anthropologist, found to his sorrow when he found the truly 'other' or the completely unknown in the heart of the Brazilian jungle:

> They were as close to me as a reflection in a mirror; I could touch them, but I could not understand them. I had been given, at one and the same time, my reward and my punishment….I had only to succeed in guessing what they were like for them to be deprived of their strangeness…if…they retained their strangeness, I could make no use of it, since I was incapable of even grasping what it consisted of.[21]

Because the 'other' was completely 'other' it was unintelligible. And the only way to make it intelligible was to domesticate it – to transform it into something already experienced. "By starting with a world we know, we are led into believing in the world in the model, because a number of its features still match those of the world we know."[22]

In fact, the process of modeling *is* the construction of something that is arguably 'like' the target domain in both elements and relationships. This requires that the target domain must (be perceived to) have some structure, consisting of both elements (or properties) and the 'laws' that govern the relationship of these elements to each other. However, as Frigg says, these domains themselves (he calls them 'systems') have no inherent structure. That structure (Frigg says) is ascribed by the scientist.[23] The theory is the story – a type A model – that describes the structure. The theory thus does not inhere in the thing described, but is the frame for understanding the scientist brings – the 'sensemaking' device, in Karl Weick's terms.[24]

Of course, while a model is 'like' the target domain it also is 'not like' that domain. A description of interactions among kin in the United States will show how a group of people (elements) are related (structure), but probably will not provide information on the economic exchanges in which they engage (another structure).

---

[20] Thomas S. Kuhn. 1970 (1962). The Structure of Scientific Revolutions University of Chicago Press. Chicago, IL. p.50-51

[21] Claude Levi-Strauss. 1977 (1955). Tristes Tropiques. John and Doreen Weightman (trans). Pocket Books, NY. p.376

[22] Morgan. 1999. op.cit. p.363

[23] Frigg. 2002. op.cit. p.23

[24] Karl E. Weick. 1995. Sensemaking in Organizations. Sage Publications. Thousand Oaks, CA

THE CREATIVE ROLE OF MODELS – THE USE OF THE THEORETICAL EXPERT

At this point we have established a process which is much more active or creative than the simple three part diagram we drew in Figure 2. We now begin with an unknown or target system that we want to explain. That system is composed of unknown elements in some random structure. We also have some system that we know that we say is 'like' the unknown system in some way. In creating this likeness, this analogy, we identify a unit of analysis (elements, properties, a person, a group), and relationships that are key in the known system, and posit similar units of analysis and relationships in the unknown system. That is, we *impute* structure to the unknown system: we *create* order. If those relationships in the known system are causal, we posit, through the power of analogy or isomorphism, that those in the unknown system also are causal. We thus have established a theory or what we are calling a type A model. Figure 4 illustrates the construction of a theory.

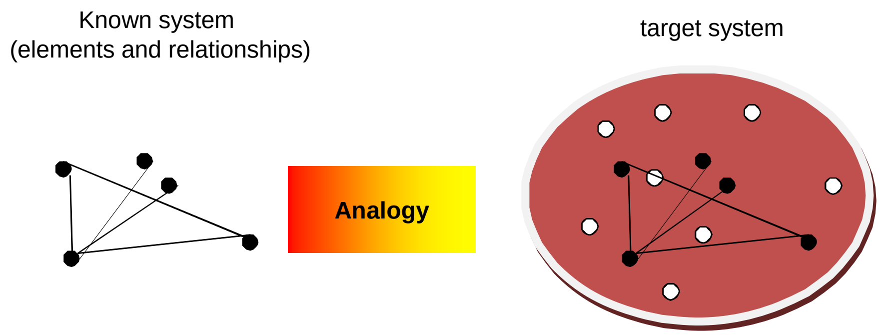

FIGURE 4: CONSTRUCTING AN ANALOGY

One clear way in which the scientist ascribes structure to the world is through the identification of relevant elements which is part of the definition of the focus units of analysis. Take the domain of human activity as an example. Is the basic unit of analysis the individual or the collectivity? If it is the collectivity, how is that collectivity defined? By the ties of blood and marriage we call kinship? By the abstract rules governing citizenship? The unit of analysis selected has been variously codified in different disciplinary approaches. Anthropologists and sociologists focus on collectivities, ranging in size and complexity from dyads to entities as abstract as nation-states; psychologists focus on individuals either as entities in and of themselves or as they interact in groups; neuroscientists focus on a particular organ (the brain); and so on. The quote from Kuhn cited above gives an example of the power of the selection of the unit of analysis in the construction of explanations.

Delimitation of the relevant unit of analysis and the relationships among these elements of the model is a powerful dimension of the 'creativity' of models. By saying that one thing is 'like' another, models may cause us to 'see' things we might otherwise not have.

> A given metaphor [where metaphor is defined as a strong analogy][25] highlights certain features of the source domain and hides others, depending on the intent of the author.

---

[25] An analogy is 'man is like a wolf.' A metaphor is 'man is a wolf.' In metaphor, the two parts are collapsed and are presented as identical.

> Often, however, some of the hidden elements are implied by the author or are inferred by the recipient, depending on context. It's just these implications that make metaphor a powerfully creative force in scientific reasoning.[26]

Hesse calls this the 'interaction view' of models.

> It is claimed in the interaction view that a metaphor causes us to 'see' the primary systems differently and causes the meanings of terms originally literal in the primary system to shift toward the metaphor.[27]

For example, says Brown, "the act of naming the process 'folding' [i.e. protein folding] *creates* similarities…it invites us to probe the cross-domain mapping between the literal, everyday act of folding and the changes that occur in a protein as it undergoes the transition we call folding" (emphasis in original).[28] If we assume that interaction among individuals in Iraq, for example, is defined by their common citizenship we will simply not 'see' the influence of other factors such as religion or kinship. Calling certain aspects of computer-based interaction 'cyberspace' causes us to treat them as we would other types of space (like air space, land, or sea). We can fight 'in' cyberspace, defend cyberspace, inhabit cyberspace…etc.

This construction of analogies, or the imposition of structure on the world which we have argued is the root of the construction of a type A theoretical model, describes the role of the *theoretical expert* in the modeling process. It is in the exercise of this role that the modeling process begins to exercise choice, to select certain elements and relationships from the target domain as important or relevant, and others as not. We suggest that the role that this role plays (to be somewhat redundant in language, although not in thought) has been largely ignored or, at best, significantly underplayed in the computational social modeling process. We will return to this point later through our discussion of validation and verification.

TYPE 'B' MODELS

The theoretical expert, through the application of theory (the construction of type A models) tells us what is relevant in the experienced world, given the particular problem the prospective model user is facing. Is it individuals? Collectivities? If collectivities, is it the nation-state or the *umma'*? Or both? And do collectivities interact on the basis of ideology or self-interest? Do we have a 'clash of civilizations' or are we concerned with *realpolitik*? Or one way in one type of situation and another in the second? Theories (type A models) help us both frame and answer these questions. But how, then, do we apply this general knowledge to (e.g.) Islamic fundamentalist groups in the Middle East in the 21st century? (And yes, we will get to computational models. Our point here is that computational models are a particular instance of a more general class of things… and we must understand that class in order to understand the instance.) This problem is compounded in the social sciences as there is no common shared set of generally accepted theories (such as the 'laws of physics' in the physical sciences or natural selection in the life sciences) to serve as a common starting point.

---

[26] Theodore L. Brown. 2003. Making Truth: Metaphor in Science. University of Illinois Press. Urbana, IL p.29

[27] Hesse. 1966. op.cit. p.167

[28] Brown. 2003. op.cit. p.25

We now posit a different class of models. Our type B model posits an instance of the general structure presented in the type A model in a real-world case by giving the objects and relationships in the analog location in time and space. So a type B model *is a particular instantiation* of a theory which is a type A model. We now construct a two-step diagram as shown in Figure 5.

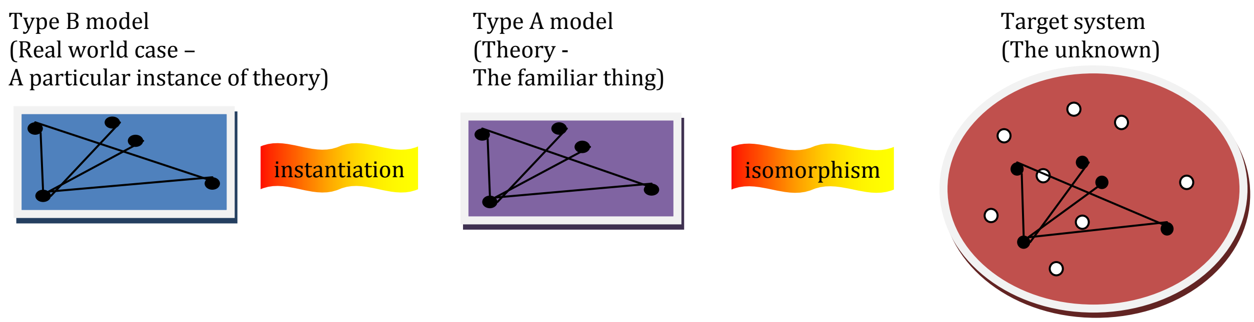

FIGURE 5: TYPE A AND B MODELS

Type B models do not tell a story or give an explanation for phenomena as do type A models. Rather, they work within the story or structure previously defined by the type A model.[29]

Type B models focus on the *content* of the elements and relationships of the target system or phenomenon, rather than on abstract instances which define them. The elements and their relationships have been defined *a priori* as part of the construction of the theory, or type A model. For example, a model in physics which predicts a comet's path might treat all objects as mass points because the only aspect of those objects which is assumed to be of value for calculating trajectory is mass. The model predicting the comet's path assumes that the taste, smell or color of an object will not directly affect its trajectory. We thus *start by assuming* that mass is important. We do not ask *if* mass is important. Our task is to acquire data about mass from the 'real world' to populate the model for a particular interest. By the same token, if our interest is in human collectivities, we may construct a model of a group that connects people on the basis of biological or marriage relationships. We would focus on data illustrating the particular kin relations a group emphasizes. We would not ask *if* kinship were important. However, in strong contrast with the physics model, the social sciences have many different operative 'type A' or theoretic approaches to social relationships. The social sciences *do* ask if kinship is important, or if (for example) the more rule-based constructs of solidarity defining nation-states are operative. If the latter case holds, the

[29] The type B model has its roots in the early enlightenment, the scientific revolution; it is in many ways the direct product of the thinking of Francis Bacon, the father of the modern 'scientific method.' The scientific method is an attempt to use tools and reason to isolate aspects of objects which seem to be governed by laws, and therefore predictable. It also is an effort to remove human error from analysis by looking at smaller aspects of the whole in a more rigorous, more consistent manner. The push is toward replicability of results, with variance attributed to human error. This is connected to the Galilean notion of idealization, which is predicated on the belief that there is a skeleton (model) of ideal, objective, phenomena underlying the subjective world we experience. This causes the objects in the world to behave in a certain manner. Again, variance from these expected behaviors is attributed to human measurement error. This general paradigm underlay scientific knowledge for over three centuries.

connections among people will not be marriage but common citizenship. In this social example, it is the type A model (the theoretic model) that is challenged. Its legitimacy must be established (or at least provisionally accepted) before data is collected to populate a type B model.

Type B models give us an opportunity to test the underlying assumptions of type A models. Suppose we provisionally accept that kinship is the primary locus of self-identification for some random collection of human beings. We would then collect data from that group about kin-based relationships, and use it to populate the model. Obviously, this is where the role of the *data provider* comes into play. We would then exercise or 'run' the model and see if it corresponds to what we see in the target domain. If a type B model's outputs conflict with data, the type A model that provided the underlying structure has been invalidated. *A type B model cannot be validated* in the strict sense of the term 'validation'. Absolute isomorphism, or strict validation, cannot be established between a type B model and the target phenomenon, because the type B model is constrained by the specific lens of theoretical thinking embodied in the type A model structure which informs it. Only a type A model or a theory can be validated (falsified). We will return to the question of validation later, but it will be useful here to develop a more complete example to illustrate the relationship between type A and B models.

We will use the kinship example we introduced in the introduction to this discussion as an example. Suppose we take an area of the world known as Weirdistan as our target system. Initially, it appears as some random collection of individuals who move through time and space. Our theoretical experts tell us that there is a 'story' we can tell about the way these individuals interact that takes their action out of a random and into an ordered space. That story revolves around what we call 'kinship relationships.' The *theoretical expert* tells us that groups are defined by relationships of blood, marriage and descent, organized into tribes (which themselves are defined by a narrative of common descent, and that these same relationships constrain the flow of social resources such as money and power. This is (at its simplest level) kinship theory. In Weirdistan, our *data provider* tells us that that people self-identify as belonging to the same group to which their father belongs. Our data provider also tells us that social resources (such as power and wealth) are passed through the male descent line. This mapping tells us that Weirdistan recognizes patrilineal descent (descent through the male line). If we map this relationship, we have Figure 6 , where color reflects group membership.

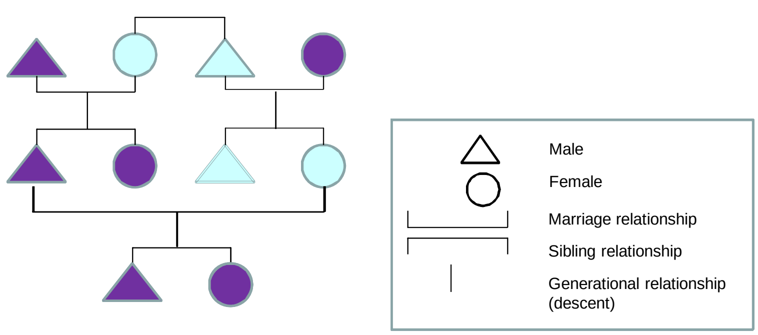

FIGURE 6: PATRILINEAL DESCENT

Kinship theory, which we get from our *theoretical expert*, allows us to 'predict' that a man is likely to marry a woman from his mother's brother's group (known as a 'cross cousin') as that would be a woman from a group that is NOT his own – or, at the very least, NOT to marry a parallel cousin, or a cousin from his own patrilineage. These kinds of 'predictive' statements (albeit 'soft' predictions as they are indicative not determinative) arise from the application of our type A model. That they are 'soft' predictions suggests that there are other type A models at play that we might not yet have identified (there might be other ways in which one chooses a mate such as notions of 'fictive kin' or adoption where an otherwise ineligible individual is 'made' eligible) or that this particular type A model has not yet been determined to embody causal relationships. We then look at actual data from Weirdistan and see if marriage patterns and social relationships match what our theoretical model leads us to expect.

This leads us to an important conclusion. The progression from isomorphism to instantiation argues that in what is conventionally called 'the modeling process,' *a theory (type A model) is an integral part of a model of a particular instance (a type B model).* This means that every model of a particular phenomenon is analytically two models, as the model of the particular instance (the type B model) is constrained by identifiable theoretical assumptions (the type A model). Our Figure 5 holds. The theoretical assumptions can be implicit or explicit. In most computational models of physical or biological phenomena, these theoretical assumptions are implicit and generally unchallenged. One doesn't question the 'laws of physics' when modeling the trajectory of a weapon, or the 'truth' of natural selection when working in population biology. But one should challenge the applicability of a particular social theory as there is not the same general acceptance of a small body of explanatory principles in the social sciences as there is in the physical or life sciences.

As we suggested briefly earlier, the theoretical landscape of the social sciences is far different from that of the physical or life sciences. There is a wide divergence of opinion on just the concept of 'culture,' for example, one of the foundational terms in anthropology. Definitions can range from the artifactual (the collections of 'things' and isolable, identifiable behavior patterns that characterize a particular group of people) to the highly ephemeral (ideologies and belief systems). It can be a term that characterizes a fairly stable phenomenon (as was implicit in British structural-functionalism that underlies much of our thinking about kinship) to a continuously negotiated, always emerging function of interaction (a premise underlying most 'sensemaking' approaches).

This is not the place to go too far into these different approaches in the social sciences. However, it is enough for our present purposes to emphasize that the nature of the type A models, or theories, that inform the type B model, or re-presentation of the target domain at a particular point in time, will affect not only how the type B model is constructed but also what is included in the modeling process.

Since the selection (construction) of the type A model is highly contested in the social sciences, we argue that it is critical that the type A model (the theoretical assumptions) always be made explicit in computational social models. In fact, we find this to be the exception rather than the rule. Without formal presentation and justification of the type A model, the decision maker does not know *why* our computational model of Weirdistan includes detailed kinship structures but little information on economic relationships (for example) and even less on religion. Without explicit delineation and justification of type A models, there is no means by which the modeling team can be held intellectually accountable for model structure. (And model structure will drive model data… there will be no data on religion if the theory – the type A model – does not include a religious dimension.) An explicit presentation of the type A model provides an explanation of and justification for the logic of choice.

CONSTRUCTING A MODEL

Jessica Turnley has suggested elsewhere that delimitation of the portion of the target domain selected for modeling as a type B model is constrained by three factors: the selected theoretical approaches (which she refers to as "notions of the construction and functioning of the target domain"), i.e. the type A model applied; the purpose of the type B model; and the modeling approach chosen (where 'modeling approach' is a function of both presentation and theory, a statement we will develop later in this discussion).[30] We argue here that the model's purpose provides the initiating constraint on model development and sets some of the strongest boundaries on its nature.

MODEL PURPOSE AND INSTRUMENTALITY

Models 'do' things. They are instrumental in various ways. However, the instrumentality involved in the creative role of models (that they can make sense or provide explanation in their own right) is not universally accepted. According to Morrison and Morgan, "literature on scientific practice still characterizes the model as a subsidiary to some background theory that is explicated or applied via the model."[31] In this sense, the model is simply a language or a vehicle to convey information, to translate the theory into a different presentation mode. It does not create new knowledge *ab initio*.

We have argued for a broader role for models through their analogical properties, and posit that they are sensemaking devices as Weick uses the term,[32] helping us to see the world as ordered (non-random). Models thus require a sensemaker or user in order to be a model. If an ant traces lines in the sand that have the same geometry as a map of Iraq, it is not a map of Iraq until a map-reader comes along. (This is another version of the old question - if a tree falls in the forest, does it make a sound? We would answer that in this context by saying it makes a noise because it does generate sound waves – but does not make a meaningful sound until there is a listener to give it meaning. In the same way, our lines in the sand are just that – lines in the sand – until a reader comes along to give them sense.) As it is with any human artifact, the sense is exogenous to the model. In the case of a model, it is imputed by the problem the model was designed to address. The problem the model is designed to address, a dimension of the model purpose, is a critical part of a model's sensemaking function. Thus the role of the questioner is critical to the shape of the model constructed.

By 'model purpose,' we mean the question the model is designed to address as well as the model's use environment. The question could be one of two types: it could be a general, abstract question

---

[30] Jessica G. Turnley. 2005. Validation issues in computational social simulations. Paper presented at 3rd UCLA Lake Arrowhead Conference on Human Complex Systems May 2005. http://www.hcs.ucla.edu/lake-arrowhead-2005/HCS2005_JessicaTurnley2.pdf

[31] Mary S. Morgan and Margaret Morrison. 1999. "Introduction" in Models as Mediators Mary S. Morgan and Margaret Morrison, eds. Cambridge University Press. Cambridge, U.K. pp.1-9. p.7

[32] Weick. 1995. op. cit.

about the nature of the world (a theoretical question), or it could be a question about a specific phenomenon located in time and space. Note that these correspond to our type A and B models, respectively. The ‘use environment’ is defined as the social and physical locations of the questioner and of the model user (refer back to .

Table 1 for definitions of these roles). While we recognize that this definition of purpose conflates several dimensions into one concept (the targeted problem and the multiple dimensions of the use environment into the concept of 'model purpose'), we believe this is analytically tractable and, in fact, analytically necessary.

Every statement of model purpose must include a statement of the problem to be addressed, and some definition of the user and his environment. The purpose could be formulated as, for example, "help the warfighter anticipate adversary movements in Weirdistan," proposing a type B model. This is a very different purpose or problem statement than "help the war planner (or policy maker) anticipate adversary movements in Weirdistan." The first statement implies a fast tempo of operations. It is also likely that the user of information provided by the model is operating in adverse and potentially technically unsophisticated conditions. Both these statements are not true for the war planner. His operations tempo by definition is slower. He may be operating in adverse conditions (if he is working in theater), but most likely in conditions that are somewhat more technically sophisticated than those of the warfighter. The policy maker is working in a completely different environment, and his operations tempo is significantly slower than even the planner's.

The model's purpose provides certain boundaries for the type A or theoretical model. If we are interested in the acquisition of weapons, for example, we will draw from different theoretical traditions than we would if we were interested in recruitment into terrorist organizations. This has implications (as we will see later in this discussion) for modeling approach. It also places constraints on the type of 'language' or vehicle in which the model is presented. Finally, it also suggests that it is highly unlikely that a model built in response to a particular purpose (for a particular user, as we have defined that role here) can be directly and simply transferred to another user and applied to a different problem in another use environment.

SELECTING A PRESENTATION MEDIUM AND A MODELING APPROACH

By 'presentation medium,' we mean (for example) a computer display, a narrative, or a physical representation. This establishes an important definition for our discussion. To say that something is a 'computational model' or a 'mathematical model' or a 'narrative or textual model' says nothing about either the thing that is modeled (the target domain) or the character or nature of the model itself. It is a description of the presentation medium, nothing more.

That said, presentation media do put constraints on the way in which data is manipulated and so affect the type of data that can be included in a type B or predictive model. Each medium also allows the analyst to manipulate the data in different ways, so different insights can potentially be drawn from the same data set. Presentation media are:

1. Physical materials (physical models)
2. Mathematical models
    2.1. Equations
    2.2. Computational models
3. Natural language

The presentation medium chosen should be a function of problem type, use environment, and data type (e.g. quantitative, qualitative, natural language, graphic…). It thus should be selected by the *questioner* (problem type), the *user* (use environment), the *data provider* (data type), and the *model builder*. It should not be selected by the model builder alone, although the model builder will provide input on the capabilities of the modeling language and the presentation medium of which the others may not be aware. Mathematics, for example, requires quantification; natural language does not. Computational models, which are a subset of mathematical models, can show results in pictures (graphics) in ways that pure equations and natural language cannot. Physical models also communicate graphically, but about only one target system. Computational models can be altered through data input to communicate about several target systems. They also can be easily modified in ways that physical models cannot, and can accommodate very large data sets and/or number of relationships. Natural language models can also be easily modified, and the elucidation of their structure and functioning is more easily grasped by most people than is that of computational models which often require deep background in mathematics. However, natural language models cannot accommodate large data sets or sets of structural relationships as easily as can computational models. There are many other distinctions among presentation media; suffice it to say here that the nature of the user's problem (which includes here the environment within which the model will be exercised) and the nature of the data (for a type B model) affect the choice of presentation medium.

The selection of a presentation medium almost always will involve tradeoffs. If, for example, the problem type and the use environment requires a computational format, but available data is qualitative and in a natural language format, some compromise will need to be made and justified in terms of the entire modeling process. In this case, the data can be 'converted' to quantitative format through the use of surrogates, where (for example) 'religious belief' is represented by church-going or prayer behavior. The quantitative precision required by the mathematics underlying computational social models thus trumps the accuracy found in the qualitative (natural language) description. We will return to this point about data later…but suffice it to say here that this tradeoff and the associated costs should be made explicit in the modeling process and in any presentation of the model results. In fact, this tradeoff is generally implicit. Most selections of computational formats are driven almost entirely by problem type, and secondarily by use environment. It generally is assumed that data will be available in the appropriate format, and the cost of transforming it to that format, if any, is not addressed.

Table 4 summarizes the strengths and weaknesses of different presentation media types.

TABLE 4: TYPES OF PRESENTATION MEDIA

Characteristics / Model Type	Description	Strengths	Weaknesses	Comments
Physical models[33]	Physical models are models constructed from physical material.	Physical models closely resemble their target, thus making the target accessible and real.	Cannot represent abstract elements of a system Can be difficult to construct	Well-suited for problems involving physical objects in space
Mathematical models	Highly formalized description of elements and their relationships represented through the language of mathematics.	Universal language 'Grammar rules' are well-known Relatively compact and simple	Requires a 'technical' knowledge of the language of mathematics to understand.	The model is a formalism; meaning is exogenous to the model and must be assigned to the model by the analyst.
Computational models	Computational models use machines to store data and to rapidly process information mathematically. They are programmed with equations.	Replicable calculations Can quickly change data but keep relationships (structure) intact Graphical user interfaces allow manipulation of complex equations with no technical knowledge of how the entire system works	Data must be quantitative Relationships (structure) must be able to be expressed quantitatively Requires a high degree of technical knowledge to understand how it works	Well suited for problems where the data changes frequently, and/or the relationships among elements are many and complex

33 Justin B. Kinney, Gasaper Tkacik, and Curtis G. Callan, Jr. 2006. "Precise physical models of protein–DNA interaction from high-throughput data", Princeton University, Princeton University(Princeton, NJ, November 8, 2006)

Characteristics / Model Type	Description	Strengths	Weaknesses	Comments
Natural Language	Presented in written or spoken natural language.	No more technical understanding than basic language comprehension is needed to access a natural language model. Handles qualitative data easily Good for stimulating dialogue Can expose implicit assumptions Natural language can be extremely accurate in describing phenomena.	Natural language is very imprecise Overly technical language may require explication outside the model itself. Each iteration of the model will be different…lack of replicability across time and space Natural language often lacks the necessary precision to pass scientific muster.	Useful for problems which are not well-formulated Useful for problems which involve a large amount of qualitative data

Once the presentation medium is chosen, there needs to be a selection of a modeling approach. The modeling approach is the particular type of presentation selected from the entire class of types in a particular presentation format. For example, if the subject matter is social phenomena and the presentation medium is a computational one, the modeling team can choose from such approaches as systems dynamics, agent-based models, networks and the like. *The modeling approach is constrained by the theoretical model at play in the modeling process and the selected presentation medium*. If, for example, the target domain is characterized by the theoretical expert as stocks or quantities or 'buckets' of things that can get bigger or smaller because of outside influences, systems dynamics modeling would be a modeling approach of choice. If he thinks of his target domain as a collection of discrete actors who engage according to specified rule sets, he would select some form of agent-based model. If he is primarily concerned with relationships and social structure, social networks would be appropriate. Thus the modeling approach should be selected by the *theoretical expert* and the *model builder*. The theoretical expert characterizes the target domain, and the model builder selects an approach that is compatible with that characterization. The modeling approach thus serves as a lens to focus the model user's attention on certain parts of the target domain and not others.

We illustrate the relationships among tasks of the modeling process that we have addressed so far (establishment of model purpose, development/choice of theoretical constructs, and selection of presentation medium and modeling approach) and the social roles involved in a modeling team (questioner, user, theoretical expert, data provider, and model builder) in Figure 7. Note that all social roles are activated at some point. We also have included the funder who, particularly in government environments, may drive requirements in areas such as presentation medium and modeling approach.

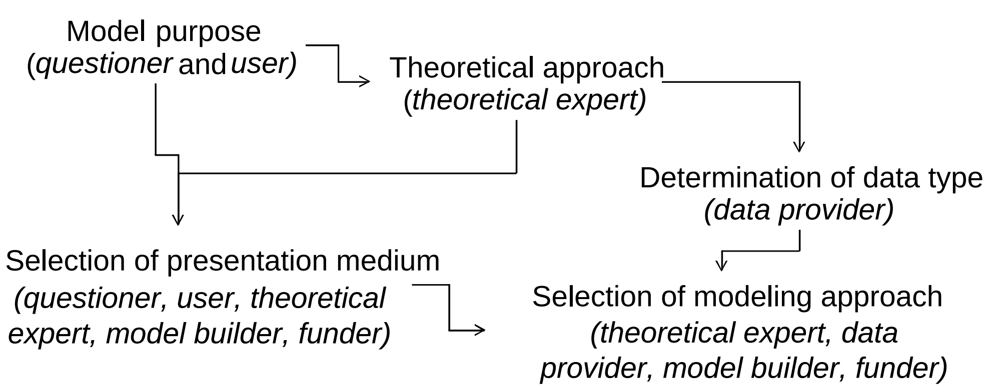

FIGURE 7: PARTICIPANTS IN SELECTED ASPECTS OF MODELING PROCESS

Selection of the modeling approach thus is part of the creative role of models, as it depends upon and embodies theoretical assumptions (the type A model). It is a critical step in the model building process and should be formally explicated and explicitly justified. Addressing this question of choice of modeling approach (also called 'model uncertainty')[34] should be part of the overall argument for the 'goodness' of the model. This, of course, requires an explicit statement of model purpose, parsed as we have here, in terms of the specific problem to be addressed and the characteristics of the use environment. We have found, however, that this argument is rarely made. As a consequence, the modeling approach for many computational social models appears to be a function of the model builder's expertise rather than determined by the requirements of the problem.

As we are concerned with social phenomena, it is not surprising to find that most computational social type B models with which we are concerned are actor-based models.[35] Their basic unit is an actor. An actor can be a person, an inanimate object, or a demographic cohort. Thus, an actor does not necessarily have agency (that is, 'free will' or the ability to make choices). Actor-based models assign various attributes to the actors, and develop logical rules by which the actors interact. These rules are expressed mathematically, which allows them to be manipulated computationally.

The various types of actor-based models are as follows, and are annotated in Table 5:

1. Agent-based models (various types such as queuing models, cellular automata, distributed artificial intelligence, neural networks…)
2. Statistical models
   2.1. Social networks
3. System dynamics

[34] David Draper. 1995 "Assessment and Propagation of Model Uncertainty" *Journal of the Royal Statistical Society, Series B (Methodological)* Vol. 57, No. 1 (2005): 45-97

[35] Carl Hewitt. 1973. A Universal Modular Actor Formalism for Artificial Intelligence

TABLE 5: ACTOR-BASED MODELS

Characteristics / Model Type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
Actor Model – general description[36]	Exist atop a framework of pre-determined laws and relationships (a specific story or explanation about the world, a theory). Actors can be anything from a representation of a social group to an email address Each actor can have a wide range of actions. Actor models are basically exercises in the application of rules – they are inherently logical	Actor models can be predictive. A good actor model will be logically consistent, and thus easily verifiable. Actor models can either examine the nature of the relationships between actors, or the patterns and networks that emerge from a collection of dynamic rule-based actions. They can produce the 'unexpected'	Generally lack dynamics for domains where rules cannot be identified or do not exist Scalability may be a concern Actors can only act in accordance with programmed rules. If rules of interaction are unknown, they must be posited.	Useful for models of phenomena that are composed of discrete entities with known rules of interaction	The underlying type A model and associated theoretical context are always implicit, sometimes to a lesser degree than others. Note that all the underlying assumptions can never made explicit. The largest assumption made in any given actor model is that the modeler has identified the *correct* unit to be treated as an actor. Further, in dynamic actor models determinism (including probabilistic determinism) is assumed, as is the validity of the rules governing actor action.	Models of phenomena defined as collections of individual parts with well-defined rules for interaction. Often used to test causality as rules are generally stated as 'if – then' statements.

[36] Alex Law and Wallace McNeish. 2007. "Contesting the New Irrational Actor Model: A Case Study of Mobile Phone Mast Protest" *Sociology*, Vol. 41, No. 3, 439-456

Characteristics / Model Type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
Agent-based[37,38,39,40]	Agents are a special type of actor which embodies agency: i.e. some way of simulating "decision making" capabilities. They can represent groups or individuals. Agent-based models are a type of computational model for simulation utilizing the individual agent as the unit. They use approaches from complex systems, emergence, Monte Carlo methods, computational sociology, multi-agent systems, and evolutionary	Have the potential to simulate and model infinitely complex target domains. They can simultaneously address static (micro-level) phenomena and dynamic (macro) processes[41] As the agents are determined by 'if-then' statements, the model can play	Must make assumptions contrary to fact, e.g. set initial conditions or the beginning of time. All data must be quantified The model is only as 'good (i.e. useful and predictive)' as its rule set which is only as 'good' as the data, analysis of a particular human collectivity, and	Highly applicable for problems involving human collectivities	Assumes that reductionist approach will be trumped by complexity and emergence, that is that new configurations will emerge from the actions of rule-constrained agents. All human phenomena are quantifiable, either in themselves or through surrogates The set of rules for a	Any problem involving human collectivities acting over time and space

[37] Dan Luo, Longbing Cao, Jiarui Ni, and Li Liu. 2007. "Building Agent Service Oriented Multi-Agent Systems", Agent and Multi-Agent Systems: Technologies and Applications, Published by Springer Berlin / Heidelberg

[38] Matteo Richiardi, Roberto Leombruni, Nicole Saam and Michele Sonnessa. 2006. "Common Protocol for Agent-Based Social Simulation," Journal of Artificial Societies and Social Simulation, vol. 9, no. 1, http://jasss.soc.surrey.ac.uk/9/1/15.html, Published: 31-Jan-2006

[39] Paul Windrum, Giorgio Fagiolo and Alessio Moneta. 2007. "Empirical Validation of Agent-Based Models: Alternatives and Prospects," Journal of Artificial Societies and Social Simulation, vol. 10, no. 2, 8, http://jasss.soc.surrey.ac.uk/10/2/8.html, Published: 31-Mar-2007

[40] K. Smith, H. Brighton, S. Kirby. 2003. "Language Evolution in a Multi-agent Model: the cultural emergence of compositional structure", Language Evolution and Computation Research Unit Theoretical and Applied Linguistics, University of Edinburgh (Edinburgh, UK, 2003)

[41] We use macro and micro in a very broad sense here. The micro level refers to the rules that allow the simulation to run. They do not change over time. The macro is the social phenomenon that emerges from the exercise of the rules. This is dynamic in that each model run will give us a different output.

Characteristics / Model Type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
	programming. Agent-based models consist of dynamically interacting rule-based agents that can create complexity like that which exists in the world. Their rules may be based upon continuous mechanisms. The agents can be intelligent and purposeful, but are not so smart as to reach cognitive closure implied by game theory. Agents are situated in space-time, reside in networks. They tend to use in-sample data, i.e. replicate statistical properties of past data.	out various hypothetical simulations.	underlying theory Agent-based models of the same collectivity structured according to different type A models (theories) will look very different No formal way to compare different models of same phenomena Models are 'one-off' – the theory is multi-organizational but the type B model can be of only one collectivity		given collectivity is knowable A partial set of rules yields relevant behavior	

Characteristics / Model Type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
Statistical models[42] - general description (social networks are a type of statistical model)	A statistical unit is chosen which will be directly observed Multiple observations can be made of the same unit over time (longitudinal research). Observations of a variety of statistical attributes are a common way of studying relationships among the attributes of a single unit.	Patterns in phenomena which might at first appear stochastic or chaotic to the naked eye are highlighted	Inaccessible and often forced to make many assumptions in order to produce meaningful results. Difficult to quantify and communicate uncertainty The unit chosen and any time steps are arbitrary Effects of incompleteness of data on overall model structure is unknown and uncalculable.	Collectivities where units exist over time and data can be collected on a regular basis		Addressing problems related to large collectivities (e.g. populations) that can be observed over time.

[42] Carlos Ordonez, “Building statistical models and scoring with UDFs” International Conference on Management of Data archive Proceedings of the 2007 ACM SIGMOD international conference on Management of data table of contents, ACM Press, (New York, NY, USA 2007) Pages: 1005 - 1016

Characteristics / Model Type	Description	Strengths	Weaknesses	Types of Problems	Implicit Assumptions	Works Well For
System dynamics[43,44]	System dynamics is a single level, single agent, model which attempts to help the user understand the behavior of complex systems over time. It uses stocks and flows as the unit of analysis Incorporates feedback loops	Influence diagrams (first step in modeling process) are easy to construct and intuitively understandable Simple models can be quickly constructed Can easily be used to play 'what if' games Model format is transparent even to unsophisticated user	Stock and flow format requires quantification of all data Can get very complex very quickly	Particularly good for modeling resource consumption/ movement Also good to understand the impact on behavior of various policy scenarios.	The basis of the method is the recognition that the structure of any system is often just as important in determining its behavior as the individual components themselves.	Studying the impact of policies on systems. Studying resource consumption/ movement

[43] Jay W. Forrester, "Origin of System Dynamics" http://www.albany.edu/cpr/sds/DL-IntroSysDyn/origin.htm Accessed 10/13/07

[44] http://www.sportsbusinesssims.com/system.dynamics.htm Accessed 10/13/07

Note the strong theoretical assumptions that are embedded in the definitional statements about agent-based models. These begin with the definition of the actor. Is it a collectivity or an individual? (Our examples here all will focus on models of social phenomena.) If it is a collectivity, by what attributes is that collectivity defined? Kinship? Citizenship? Age? Common experience? If it is by multiple attributes (as most humans experience them), such as teenaged male Sunni Iraqis living in Baghdad, which one is given precedence at some given moment in time and space? How is that one selected? What are the rules of engagement for each of the different dimensions of identity and how do they interact? Note that all these questions are predicated on theories of human interaction. The theory – the type A model – drives what the modeler includes and excludes. That theory is implicit in the type B model – the model of a group of humans at some point in time and space. If the theory does not incorporate global weather patterns, the impact of floods on the life of our teenaged male Sunni Iraqis in Baghdad cannot be calculated. Given its parameters, the type B model may have been accurate – but it did not predict the boys' behavior because the underlying theory (type A model) was incomplete.

This discussion builds Figure 7, which took us from the determination of model purpose through the choice of a presentation medium/language to the selection of the modeling approach, into a more elaborate and complex schema as shown in Figure 8. Figure 8 highlights the areas we just discussed.

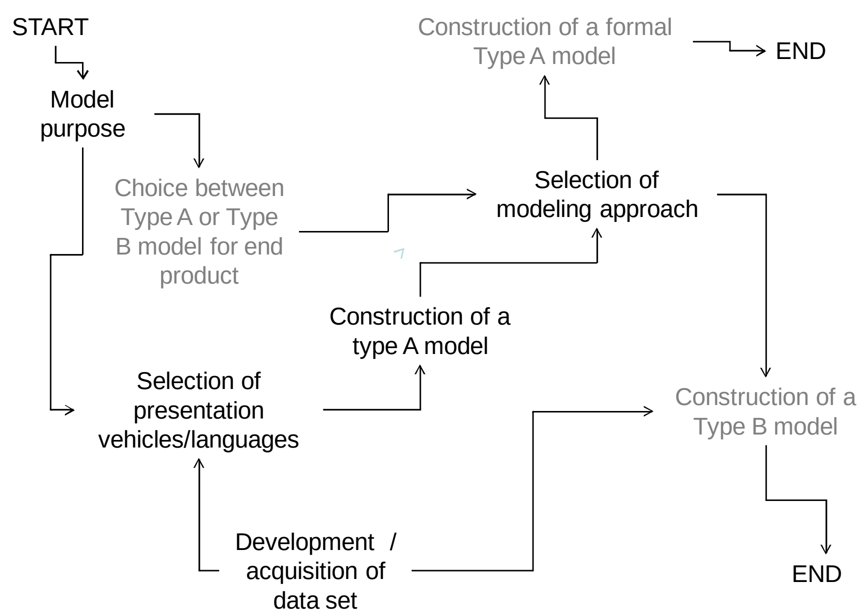

FIGURE 8: THE EXPANDED MODELING PROCESS

DEALING WITH DATA

Much of the data collected about human collectivities is in narrative form (in natural language), or addresses abstract phenomena (such as belief, motivation and affect) that are not observable and hence not quantifiable. Narrative or qualitative data cannot be utilized in computational models which are mathematically based. However, our argument shows that the required data type will be one of the determining factors in the selection of presentation medium or language. What has happened in practice is that the directionality of this arrow has been reversed. The selection of presentation medium has driven the type of data used. If data of the correct type is not available, data that is available is converted to the appropriate type through the use of surrogates. For example, if the use environment (part of what we have called the 'model purpose') requires a computational presentation language, cultural data, which is primarily qualitative and which has historically been collected primarily in narrative form, is converted into data that can be manipulated computationally. 'Religious belief' becomes 'religious behavior,' as in our earlier example, as behavior is observable and so quantifiable while belief is neither. And while most anthropologists will not dispute that (e.g.) religious belief can only be accessed through its behavioral expressions, there are strong reasons to contest the conflation of a single behavior with a very complex cultural phenomenon. In general, the *cost or the impact on the accuracy of the data and so to the isomorphism of the model to 'the real world' of such a conversion is rarely challenged.* The cost must be made explicit and justified in terms of the benefit that the presentation medium provides.

There are areas ripe for research here. The development of computational languages that can easily handle semiotic and other qualitative data would broaden the capabilities of computational methods to work with this type of data without transformation. In so doing, these languages would accommodate the requirements of the user's problem (to understand certain aspects of the human condition) through the use of computers without sacrificing accuracy. We would gain the benefits computers provide, such as the ability to store and manipulate large amounts of data) without incurring significant analytical costs.

VERIFICATION AND VALIDATION

If both model types (types A and B) are always present in a computational presentation, this leads us to some interesting questions regarding verification and validation of the models. We will only address this topic briefly here (the literature on validation, in particular, is voluminous), but it is important. As computational social models are used more widely and for a broader range of applications, questions are arising as to how a user might assess the value of one computational model over another, or, for that matter, the utility of any computational model in the human decision-making process.

The classic way to assess the value of computational models is through verification and validation. Verification refers to the performance of the code itself, while validation generally refers to the degree of isomorphism of the model output with the real world.

Verification is a reasonably well-understood process. Various tools (primarily statistical in nature) have been developed to determine the 'goodness of fit' of the code performance,[45] that is, does the code perform as the model builders intended (is it 'bug-free'). Verification (application of these tools) generally is performed by the model builder.

Validation has much greater variation in definition than verification.[46] In an abstract sense, it has been defined as a determination as to whether a particular computer program is "an accurate representation of the system under study,"[47] whether the simulation is a "good model of the target,"[48] or a legitimate "representation of the actual…system under design or study."[49] The U.S. Department of Defense adds an additional dimension of complexity to the definition by introducing the use environment: "… [validation is] the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."[50] This, of course, was famously captured by George Boc, when he said that "Models, of course, are never true [isomorphic], but fortunately it is only necessary that they be useful."[51]

---

[45] Kleijnen, Jack P. C. 1995 Verification and validation of simulation models. European Journal of Operational Research 82:145-162; Gilbert, Nigel, et al. 1999 Simulation for the Social Scientist. 1st edition. Philadelphia, PA: Open University Press.

[46] Turnley. 2005. op.cit.

[47] Jack P.C. Kleijnen. 1995. Verification and validation of simulation models. European Journal of Operational Research 82:145-162;

[48] Nigel Gilbert et al. 1999. Simulation for the Social Scientist. 1st edition. Philadelphia, PA: Open University Press.

[49] E. J. Williams. 1998 Verification and validation in industrial simulation. Proceedings of the Summer Computer Simulation Conference 57-62.

[50] DOD Dictionary of Military Terms, as amended through 26 August 2008 http://www.dtic.mil/doctrine/jel/doddict/

[51] Quoted in Box, G.E.P. "Some Problems of Statistics and Everyday Lide", *Journal of the American Statistical Association*, Vol. 74, No. 365, 1979, p.2.

These definitions beg the question of exactly what is being 'modeled.' If we apply these more abstract definitions of validation to our typology of A and B models (focusing on computational social models and leaving aside the Department of Defense definition for the moment), we find the following. We would use validation to test our type A model, the theory, against the 'real world.' We would use verification to assess the goodness of fit of the presentation of that theory in computational form (type B model). We arrive at this conclusion as follows. We will present the argument first and then follow it with an example to illustrate its application.

A type A model is a re-presentation of selected elements and relationships from a target domain. That selection is made by analogy. We compare the target domain to some other domain with which we are familiar. We then posit that those relationships are causal. A 'theory' is the expression of that causality. Since the elements in our familiar domain act in a certain manner, we posit that elements in the unfamiliar domain will act in the same way. This is our type A model. This exercise is the provenance of the theoretical expert.

We then give that theory location in time and space by using specific data instead of abstract concepts. This is the basis of our type B model, and clearly brings the data provider into play. We then ask if the output of our type B model gives us the 'same' results as the real world.'

It is in this form that the question is posed in most instances where validation is attempted for computational social models. However, this is actually a two-tiered question. We must first ascertain that our type B model is, in fact, structured the same as the type A. At this point, this actually is an exercise in *verification* not *validation.* We are asking whether or not we have developed and presented the model structure per the template provided by the type A model, the theory. If we have indeed done so, we then ask if the type A analogy itself holds. This is where validation actually takes place. This process is mapped in Figure 9.

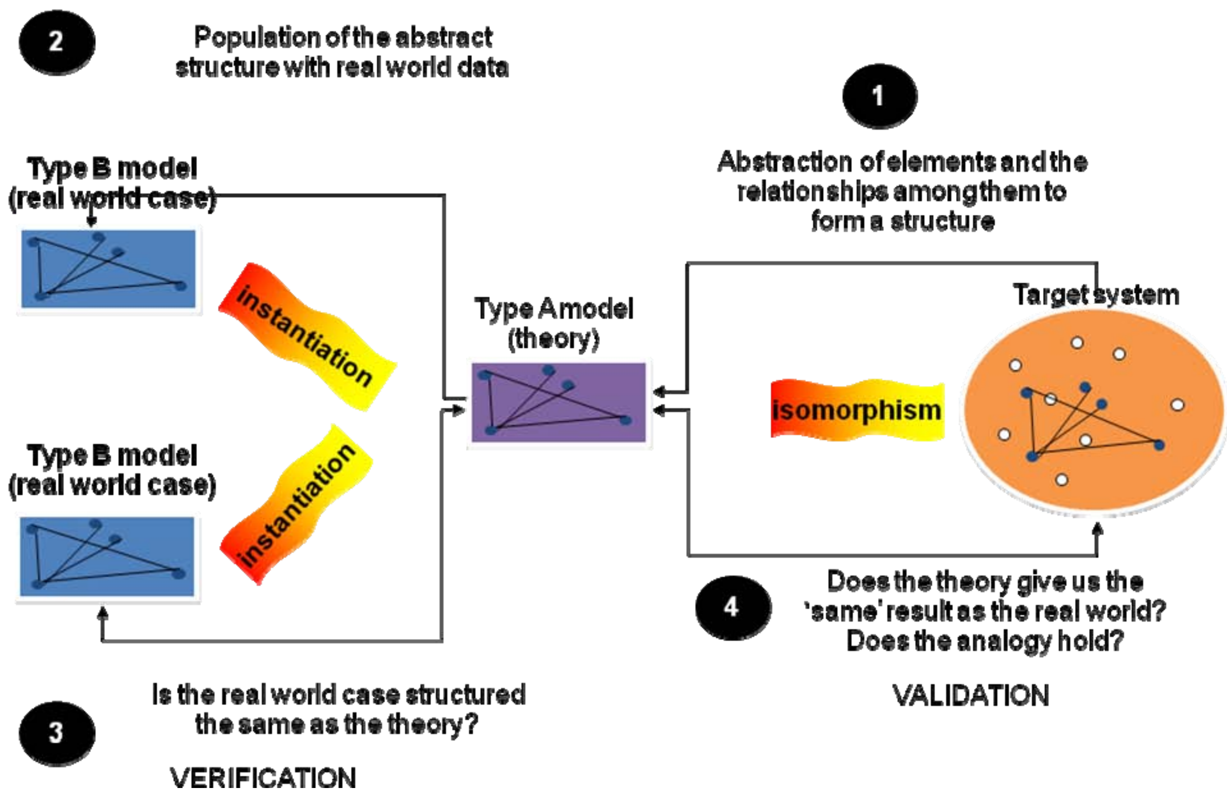

FIGURE 9: VERIFICATION, VALIDATION, AND TYPE A AND B MODELS

Type B models can be 'objective' insofar as they are verifiable, i.e. the outputs are the necessary consequences of the assumptions in the model structure, and therefore are objectively true, given the parameters of the model. Thus the tie of the type B model back to the world (the validation track) is actually verification (i.e. a check for internal consistency) of the type B model followed by a validation of the type A model.

If the model does not 'look like' the 'real world' when it is run, three possibilities can be explored. First, the code builder can check the 'goodness' of his data (in effect, challenge the data provider). There often are questions about data completeness (for example in social network models, where it is not clear what impact incomplete data has on the integrity of the network). Second, there may be questions about the fidelity of data that has been converted from qualitative to quantitative in order to be computationally manipulated. As we noted earlier, this is a question that is rarely asked but is of potentially significant import. Finally, the code builder can go back to the underlying theoretical mode and challenge the theoretical expert. If (for example) the model was constructed to determine social identity by neighborhood (geography) and the coding holds and the data set is good, the fault may lie with the initiating assumption as captured in the theory (type A model). Geography may *not* be a strong behavioral motivator or constraint. It may be trumped by kinship, or language use, or some other factor. In short, the coding language in the type B model was correct and complete, and the model was verified; the wrong human dimension was modeled: the theoretical premises captured in the type A model were incorrect, and the model is invalidated.

An example might make this clearer. Think of human beings. Our theory, or structure, is that humans are not just random collections of organisms but are organized into defined collectivities. We posit that kin relationships are an important part of how people define themselves as a member of collectivities, as is religious affiliation and language. We also posit some known relationship between religious affiliation and kin group membership, but no relationship between religious affiliation and language. We now have created a type A model. We have abstracted religious affiliation, kinship relationships and language as relevant elements. By the same token, we have implicitly if not explicitly posited that many other elements such as height, number of toes and place of origin are not relevant in self-identification as member of collectivities. We also have posited certain relationships among the elements we have selected.

We now go to some particular part of the world and collect data to populate our model. We collect data on language, kinship, and religious affiliation. We do NOT collect data on an individual's place of origin, height, or number of toes. We 'run' our model, that is, we make assignments of individuals to groups based on the relationships we have posited among the elements. We then compare our groups to groups in the 'real world,' or the target system. If we find that our groups do not match those in the target system, we have discovered that either we have selected the wrong elements, the relationships among those elements are incorrect or both. In short, we have falsified the type A model. The type B model, that particular instance of the type A model, cannot be falsified. It can only be checked for consistency with the analogy in the type A model. We can only check to be sure that we have correct and complete data.

Validation of a computational social model thus actually focuses on the 'goodness' of the theory – how well the theory explains the 'real world.' And it is here that the Department of Defense use dimension of validation comes into play. We have argued that 'purpose' is a strong constraint upon the configuration of a type A model. That purpose includes a definition of the use environment, including the problem the model must answer and the social location of the user. It also constrains the theoretical approach and therefore the modeling approach. (For example, if I am interested in who is communicating with whom, I am likely to look at theories of social structure and relationship, and not at ideological development. I am likely to select a social network approach, rather than a systems dynamics or agent-based approach.) Therefore, an important part of the validation process is to ascertain the model purpose. This reinforces the argument we made earlier in this discussion for the importance of clarity of purpose in the modeling process. Once the type A model is populated with data from the 'real world,' the question shifts to *verification*…how well does the output from the type B model match the world as constructed and presented in the type A model.

The validation process is a means to use models to help 'prove' theories, to demonstrate that a given type A model (a given theory) can describe all possible instances of a phenomenon. We are a long way from such conclusive proof in the social sciences for a variety of reasons, including the complexity of the target phenomena (socio-cultural environments) and the ethical barriers to the conduct of experiments on human populations. As a consequence, the modeling team must explicitly state why it chose the theoretical approach it did. Alternatively, the team could construct multiple models of the same phenomena based on different theoretical approaches, either searching for the one with the greatest utility given the model purpose or using elements from all.

We are arguing that one validates a theory by applying it to a particular instance in real time and space, verifying that the application is correctly constructed, and then validating the underlying theory. The theory is captured and expressed in the model elements, structure, and process. If the theory is implicit (i.e. not formally stated in a formulation document) or is simply a collection of heuristics, it will be extremely difficult to validate the model. This strengthens our argument for the explicit delimitation of the role of theoretical expert in the modeling process, and the importance of a strong, stated and justified theoretic base for the model.

ALTERNATIVES TO VALIDATION

Validation of physical models is achieved by comparing the output of the model to the real world. Exercising computational models has the same function as experimentation in the scientific process, and therefore, a similar role relative to the development of theory. However, just as it is impossible to run most types of experiments on human populations (for a variety of ethical and practical reasons which are beyond the immediate scope of this paper), so too is it impossible to manipulate human populations in ways that would allow rigorous validation of computational models.

Validation provides an objective assessment of one aspect of the 'goodness' of a computational model. In fact, all it tells the user is that the model replicates (some portion of) the real world. However, if the model purpose is defined in some way such that there is another measure of goodness, validation in the classic sense may become moot.

Suppose, for example, the purpose of the model is to help the user see novel or unexpected possibilities in some social space. An appropriately *verified* agent-based model may make significant contributions in this area. An agent-based model (for example) typically is exercised by doing multiple runs and providing the user with a 'possibility space.' This space consists of all the outputs that can be generated from the interaction of the same set of agents over some period of time, acting in accordance with the same set of rules and the same initial conditions. In this scenario, only one of the possibilities computationally generated will actually play out (or, conceivably, none of them will and the actual future will be a possibility not yet generated computationally). Furthermore, it is likely that among these generated possibilities is one (or many) that had not occurred to the model user, and which may cause him to see scenarios (not yet) generated computationally or even hitherto unimagined relationships or elements in the target domain. If the user's confidence in the theoretic base for the computational model has been established by substantiating the particular construct of the rules for interaction by references to research literature, his confidence in the data established by the credentials of the data source, and if his confidence in the way in which the model is exercised is established through traditional verification techniques, he is likely to give credence to the imagined futures. In this case, the end game for the modeling team is not validation (comparison against some part of the 'real world') but increasing the confidence of the user in the structure of the computational model. The output is not used for predictive purposes but to help the user gain insight and develop creative solutions.[52] Changing the model purpose changes the tools needed for model legitimacy.

A second solution to the 'validation problem' for computational social models also moves us outside the scientific realm, but this time into that realm of craft. We will illustrate this approach with an example from the engineering world.

The example we will use is the modeling used for vehicle electronic control module (ECM) programming which controls fuel injection. In a fuel injection system, the programmer essentially makes a multi-dimensional map of air flow rates through the engine at given speed and throttle levels. This is based on an "ideal" volumetric efficiency (VE), which represents how much air "should" be moving through the engine (a cylinder) at a given rpm/throttle according physical laws.

If all premises were correct, all VE maps would read 100%. In fact, a given map for a given engine will be far above or below this mark. To solve the problem, the mechanic invalidates the premises from the physical laws and manually 'tunes' the engine to find the correct VE percentage values. Other variables also come into play to tune the engine correctly, further impacting the fuel injection model. The result is a well-tuned engine, but a fuel injection model that neither conforms to the laws of physics nor represents the actual air flow through that particular engine.

In our engine example, the initial computational model must be valid (it must incorporate laws of physics which have been proven to be true). It is a type A model. However, the final computational model of the particular engine (arguable a type B model) is not valid in the strict sense of the term. However, it is *useful* as it provides a touch point for the manual tuning of the engine. Here a successful process is heavily dependent upon the implicit knowledge of the mechanic, knowledge often called 'craft' rather than scientific knowledge,[53] which builds upon the base provided by the validated (type A) model. Once again, the type B computational model is not predictive but still functions as a decision support tool.[54]

In the two examples we gave here, the measure of 'goodness' of the computational model did not depend upon its predictive capability. In one case, its goodness depended upon its ability to help generate insights and stimulate a creative process. In the other, it was used as a touch point for the application of additional knowledge. Given the tremendous difficulties of validating (in a classic sense) computational social models, it will be useful to rethink their purpose and so devise other

[52] One could now argue that the computational model is being used to support a creative (humanistic) process, not a scientific process. But any process that requires human judgment arguably has humanistic elements. We argue, though that in this example the computational model still being used to inform expert judgment, it has no less import or legitimacy for moving outside the scientific domain.

[53] Stephen R. Barley and Julian E. Orr, editors. *Between Craft and Science: Technical Work in U.S. Settings*. Ithaca, New York: Cornell University Press, 1997.

[54] The authors are grateful to various technical personnel at Sandia National Laboratories who stimulated this type of thinking about this problem in conversations about their own work.

measures of goodness. This would be a full research project in itself, so we do no more than identify the need here.[55]

[55] Such a project could look at other fields that computational manipulate non-observables (such as computational linguistics), cannot experiment (such as astronomy or nuclear weapons physics), or which deal with complex, dynamic multi-dimensional systems (such as systems ecology).

SUMMARY OF THE MODELING PROCESS

We now re-look at our schematic of the modeling process and the exercise of the various roles in that process. To help with the discussion, we present the entire modeling process in Figure 10.

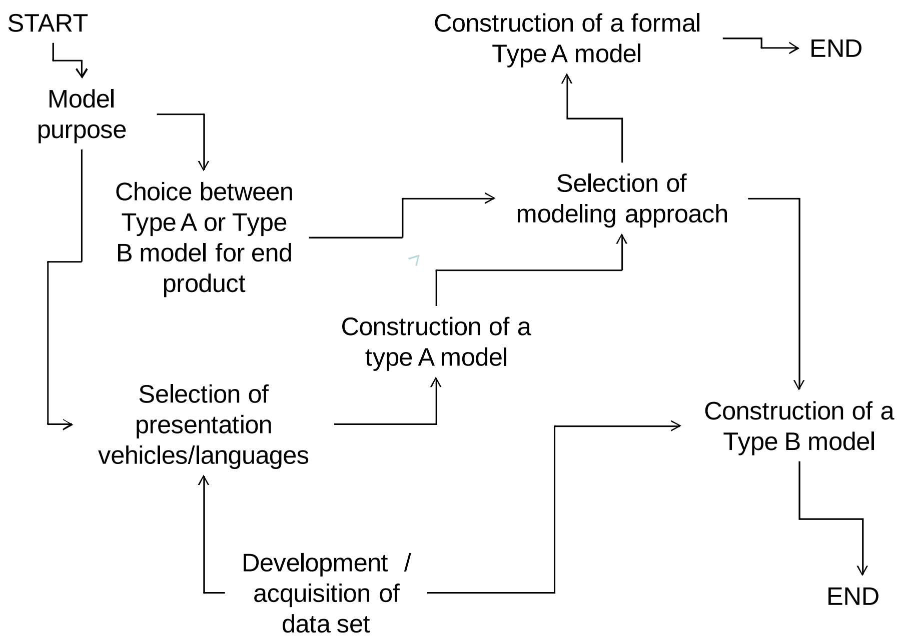

FIGURE 10: THE MODELING PROCESS

In Figure 11, we show the process with the various social roles added.

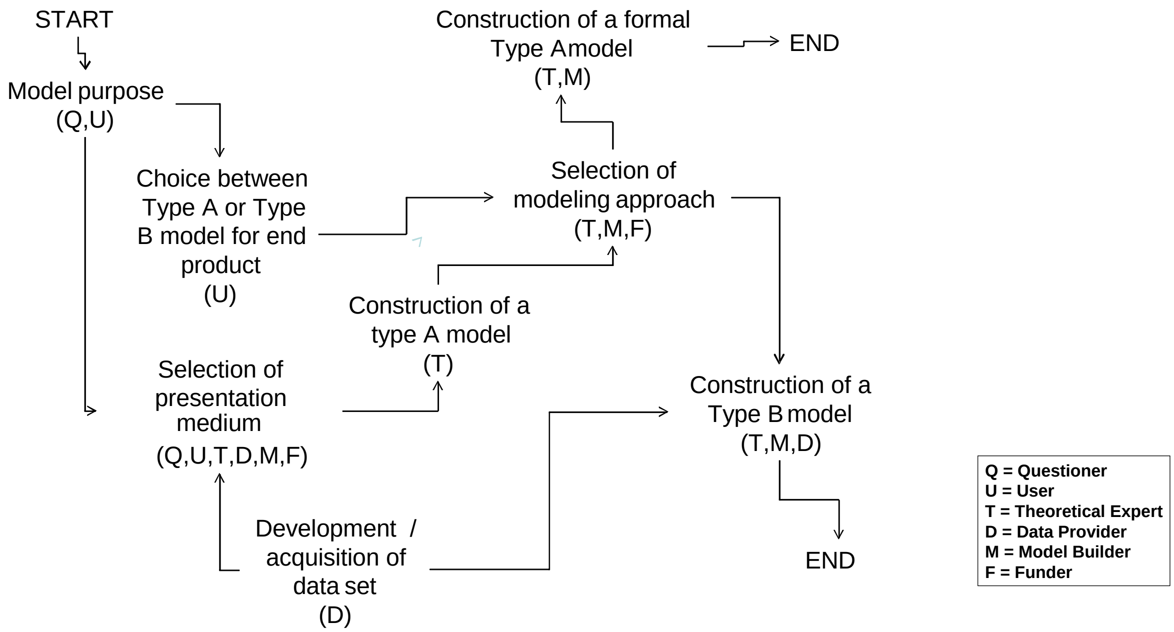

FIGURE 11: MODELING PROCESS AND ASSOCIATED SOCIAL ROLES

We have deconstructed the model construction process as follows:

- *The questioner* defines model purpose. This will include determination of the following. What is the use to which the information will be put (which is often defined as a question)? Will that require the manipulation of very large data sets, data sets with special types of data (e.g. images, video), and frequent recalculation or reassessment of information? Are there other constraints the question places on data and/or structure?

- Who is the *user?* What will his physical environment look like? (Will the user be in a jeep in the middle of Afghanistan? In a room with sophisticated display technology and state-of-the-art communication connections to other individuals, sites, and data repositories? In a village in Iraq with no electricity and questionable satellite or other communication connectivity?) What is the sophistication and/social location of the user? (Is he highly computer literate? An unsophisticated user? Has never touched a computer?) What is the decision or information utilization tempo? (Is it a strategic environment with long time horizons, or a fast-moving tactical environment?)

- On what kind of *data* will the solution to the problem depend? Is it quantifiable? If not, are there surrogates which are arguably suitable if necessary?

- Which *theories of human interaction* will drive and constrain model structure? Why these and not others?

- Which *presentation medium or communication language* is most appropriate? The answers to the questions about theory, model purpose and user will determine which type of presentation medium or communication language is most appropriate. Is the available data in a format that can be manipulated by the preferred presentation medium? If not, what is the cost of converting it?

- What is an appropriate *modeling approach*? This will be determined by the theoretical approach (the type A model) and by the selected presentation medium.
- Does the problem require a *model of a particular instance*? If so, we are working with a type B model. If not, then we can formalize our type A model and end.

SUMMARY AND PATH FORWARD

Computational social models have gained significant prominence with analysts in the national security arena as our geopolitical environment has changed since 9/11. With such prominence comes a responsibility to understand their limitations as well as their contributions, and to develop and use them appropriately. This discussion is an effort to contribute to that understanding.

SUMMARY

We have argued that there are five key roles that are always at play in the development and application of any type of model: the questioner, the user, the theoretical expert, the data provider, and the model builder. Of these five, perhaps the most neglected in the construction of computational social models is that of the theoretical expert.

Social theory is contested in a way that theory about the physical world and (although to a lesser degree) the biological realm is not. Since a theory is statement about a subset or subsystem of the elements and relationships we have determined is important for understanding how the world works, construction or application of a theory determines which part of the world we believe is important. Understanding the nature of the relationship between the theory and the world—a relationship (and others) we have characterized as an analogical one—is important.

The chosen social theory drives data collection and constrains the choice of modeling approach. If we believe that ties of blood and marriage drive social relationships and direct the flow of social resources, we will collect data on kinship relations but not on the number of toes people have. In this case, we will choose a modeling approach that shows structure and relationships. However, there also is another body of theory which tells us that social connections are based on common ideologies: belief systems and the behavior they drive are more important than kinship ties. If this is the case, different types of data are relevant as we develop an understanding of human behavior. Here we will choose an approach that allows us to incorporate the rules by which social actors make behavioral choices. And so on.

Theory is conceptualized and captured in what we called a type A model (e.g. ties of blood and marriage are important). We can test that theory (i.e. determine its predictive value) by inserting data about a real place. If a Weirdistani may only marry someone from a particular kinship group, and if Weirdistanis traces their membership in a kinship group through their father, we can construct a model of patrilineal kinship in Weirdistan. If we exercise the model and it gives us results that are unexpected in terms of what we see on the ground in Weirdistan, we have three possibilities. First, we may have an error in the code or have entered data incorrectly. We need to verify the model. If the model is verified, we can assume that our theory or type A model is wrong or incomplete. Either kinship is not the determining factor for marriage partner choice, or it is *a* factor and others will either override or confound it. Since validation along classic lines is difficult for computational social models, we can either tweak the existing model until it matches closely enough to what we have (the 'craft' approach to model development), or we can address the theoretical component and challenge the notion of kinship as a basis for marriage partner choice. The downside to the first approach (the 'craft' approach) is that there is illustrative but not explanatory power in the model. It thus becomes difficult to transfer any understanding we gain in

Weirdistan to any other time or place. The downside of the second approach (challenging the theory) usually lies in the exigencies of time. Such an approach generally would require a reconstruction of the entire computational model as key objects and relationships are redefined (or, in the most extreme case, the team makes a fundamental change in modeling approach). In a tactical and operational world where high consequence decisions are made daily, there often is not the time available for significant reworks.

This brought us to the question of validation. Computational social models cannot be validated in the classic sense of the word, that is, they cannot be compared in a rigorous fashion to some portion of the 'real world.' Ethical and other considerations preclude the manipulation of human populations to these types of end. There thus is no accepted measure of 'goodness' of these models, as validity is the standard measure used to assess computational tools such as these. We have argued that if the purpose of the models is other than prediction, the question of validity becomes moot. Computational social models can be used to inform judgment by stimulating creativity and generating insights. Granted, these are more subjective measures of goodness than a test for validity, but they may be sufficient. More work needs to be done in this area to ascertain how other fields, such as systems ecology, astronomy and nuclear weapons physics where 'validity testing' is difficult, assess and utilize computational modeling.

Absence of recognition of the importance of theory in a computational social model clearly is an issue in this emerging field. The costs of converting data into a format that can be computationally manipulated also are of importance to the value of the model and also are generally neglected. A large part of cultural data is qualitative in nature. The current state of computational science requires that this qualitative data be converted into a quantitative format. There is a cost involved in this transformation. That cost—or even the fact of the transformation—is rarely stated much less evaluated. There is a methods question here as to how that cost will be calculated. There also is a research opportunity for the computer science field to develop advanced techniques for more effective manipulation of qualitative data. While advances are being made in fields such as computational linguistics, current state of the art has the tool (the computer) dictating the shape of the object (the data). Rather than turning screws into nails, perhaps we should invest in the invention of a screwdriver.

PATH FORWARD

Our discussion of the modeling process has illustrated several areas where we believe computational social modeling has fallen short. Some of these concerns can be addressed through further research. In other cases, a requirement for additional rigor in the modeling process and in the presentation (communication) of the model structure, functioning and supporting assumptions would address the concerns.

Absence of a clear, shared understanding of purpose. A clear *and shared* understanding of the model's purpose is critical for successful utilization of the model. The model's purpose, defined as we have done here to include both the problem and the use environment, constrains the presentation medium chosen and the type of data required. In fact, the model's purpose is often *not* clearly stated at the beginning of a model description, or it is not stated with enough rigor to reassure the listener that all are working with the same definition, and to provide justification for some choices that are made further along in the modeling process. This has at least two consequences of note. First, it can lead to computational social models being utilized for decisions beyond their design scope. There is danger in this as the design scope, as we have discussed, is extremely specialized and situation specific. This, of course, raises some interesting higher-level questions about particularity and universality and underscores the need for broad-based theories of social interaction and cultural interpretation. Secondly, if the modeling team is unclear as to purpose (or has a different understanding of purpose than the questioner and/or user), the team may develop a model that answers a different question or which is inappropriate for the intended use environment. The user may need results more or less detailed than those generated by the model, for example, or may need the results quicker or the output in a format other than that designed in.

Recommendation: A formal part of the construction of any model should be the development, recording and clear communication of a statement of purpose. Dialog among all parties of the modeling process is key to effective crafting of a statement of purpose. Without such dialog, ultimately captured in a statement of purpose, the questioner and user may find that, indeed, they got what they asked for – but it was not what they needed.

An implicit and unsupported theoretical base. We have argued that every model of a particular instance (what we have called a type B model) contains elements from type A models. That is, every model is structured by theory. That theory determines which elements and relationships the type B model selects out of all possible elements and relationships. It also is the basis for validation. This calls for strong participation by at least one theoretical expert in the modeling process, particularly in social sciences where theory is highly contested. The theoretical expert should be involved in the selection of the presentation medium or language, in the selection of the modeling approach, and should be primarily responsible for development of the type A model. In fact, many computational social models are constructed with little or no involvement by social science theorists, and are built around heuristics developed by the model builders. These heuristics are implicit, but still have the same explanatory power relative to the type B or actor-based model as does a formal model based on the best of intellectual traditions. We would argue strongly for a formal description of the social theoretical underpinnings of any computational social model in its presentation.

Recommendation: There needs to be a standard of excellence established such that no model is considered complete without a full and formal documentation and justification of the social theory driving and constraining the model structure.

Data-related issues. The data themselves often are problematic in computational social models. Often the set is incomplete, particularly if the target population is difficult to observe or survey. Our knowledge of how to extrapolate from incomplete sets of social data is far from perfect. Data also may be inaccurate. There currently are no good ways to account for these uncertainties in the modeling process or in its output. As importantly, much of the critical data is non-observable (motivation, intent, beliefs…), or is collected in narrative or natural language form. In order for this data to be computationally manipulated, it must be converted to a form the computer can handle. While this data can be converted to quantitative representations, there is some cost to that conversion: a tradeoff of accuracy (isomorphism) for precision and ease of manipulation. This should be an explicit part of the assessment of data quality for any computational social model. In fact, it generally is absent.

Recommendation: To be judged complete, model documentation must state where data surrogates are utilized. In addition, research should be conducted in two areas. First, there should be some reasonably standardized method of assessing the costs of using such surrogates. Second (and perhaps more importantly), computational capabilities should be extended to enable computers to effectively manipulate qualitative data. This probably will require some significant breakthrough(s) in computational approaches.

Absence of justification of modeling approach. It often appears that the modeling approach is a function of the expertise of the modeler rather than a consequence of the theory driving the model structure. The probability that this is the case may be intensified if the role of the theoretical expert is not overt and explicit.

Recommendation: Model uncertainty should be addressed in the model's supporting documentation. This will force the modeling team to explicitly justify their choice of modeling approach.

Over-emphasis on validation. Computational social models cannot be validated with the same rigor and according to the same tenants as computational models of physical phenomena. Therefore, they cannot be used as predictive tools. The expectations and purposes of these computational social models need to be revised in accordance with the value the models do bring to the table—their ability to contribute in significant ways to informed judgment by providing insight and stimulating creative thinking.

Recommendation: Research into the ways in which fields such as astronomy and nuclear weapons physics validate models of subjects on which experiments cannot be performed, combined with exploration of the ways in which other fields such as systems ecology deal with models addressing domains of varying levels of complexity with different temporal dynamics could offer interesting and useful insights into a productive path forward in this area. Additional work on analogic and other non-deductive methods of learning could suggest ways to position computational social models so they are perceived to be valuable without requiring that they be forced into an inappropriate validation structure.

While these issues and concerns we have raised are significant, they can be overcome through more rigor in the modeling process and presentation. Models of any sort can be powerful creative tools, helping us see the world in new ways, and to manipulate large data sets over time and space. Computational social models have a lot to offer the field of social science in general and national security analysis in particular. They can play a creative role in theory generation and provide insight to decision-makers. Developing a rigorous understanding of the modeling process as a whole and the roles at play in the process will help ensure that appropriate individuals are recruited to fill those roles, appropriate theory is generated and applied, and that the model itself is well-built.

BIBLIOGRAPHY

1. Aber, John D. "Why Don't We Believe the Models?" *Bulletin of Ecological Society of America* (1997): 232-233.

2. Axtell, Robert. *Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences*. Washington DC: The Brookings Institution, 2000.

3. Barley, Stephen R., and Julian E. Orr, editors. *Between Craft and Science: Technical Work in U.S. Settings*. Ithaca, New York: Cornell University Press, 1997.

4. Brown, Theodore L *Making Truth: Metaphor in Science*. Urbana-Champaign, Illinois: University of Illinois Press, 2003.

5. Craik, Kenneth. *The Nature of Explanation*. London, U.K.: Cambridge University Press, 1967.

6. Draper, David. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society, Series B (Methodological)* 57, no. 1 (2005): 45-97.

7. Forrester, Jay W. "Origin of System Dynamics." Web page, [accessed 13 October 2007]. Available at http://www.albany.edu/cpr/sds/DL-IntroSysDyn/origin.htm.

8. Frigg, Roman. "Models and Representations: Why Structures Are Not Enough." *Center for Philosophy and Natural Science, Measurement in Physics and Economics Technical Paper 25/2* London, Britain: London School of Economics, 2002.

9. Gilbert, Nigel. "Agent-Based Models." *Series: Quantitative Applications in the Social Sciences*. Thousand Oaks, California: Sage Publications, Inc., 2008.

10. Gilbert, Nigel, and Klaus G. Troitzsch. *Simulation for the Social Scientist*. first ed. Philadelphia, Pennsylvania: Open University Press, 1999.

11. Hesse, Mary B., Author. *Models and Analogies in Science*. Notre Dame, Indiana: University of Notre Dame Press, 1966.

12. Hewitt, Carl. "A Universal Modular Actor Formalism for Artificial Intelligence."1973.

13. Ilgen, Daniel R., and Charles L. Hulin, Editors. *Computational Modeling of Behavior in Organizations: The Third Scientific Discipline*. Washington, D.C.: American Psychological Association, 2000.

14. Kinney, Justin B., Gasaper Tkacik, and Curtis G. Jr. Callan. "Precise Physical Models of Protein DNA Interaction From High-Throughput Data." Princeton New Jersey: Princeton University, 2006.

15. Kleijnen, Jack P. C. "Verification and Validation of Simulation Models." *European Journal of Operational Research* 82 (1995): 145-162.

16. Kuhn, Thomas S. *The Structure of Scientific Revolutions* Chicago Illinois : University of Chicago Press, 1970.

17. Küppers, Günter, and Johannes Lenhard. "Validation of Simulation: Patterns in the Social and Natural Sciences." *Journal of Artificial Societies and Social Simulation* 8, no. 4 (2005)

18. Lakoff, George and Mark Johnson. *Metaphors We Live By*. Chicago, IL: University of Chicago Press, 2003.

19. Lansing, Stephen J. ""Artificial Societies" and the Social Sciences." *Artificial Life* 8 (2002): 279-292.

20. Law, Alex, and Wallace McNeish. "Contesting the New Irrational Actor Model: A Case Study of Mobile Phone Mast Protest." *Sociology* 41, no. 3 (2007): 439-456.

21. Lawley, James, and Penny Tompkins. *Metaphors in Mind, Transformation Through Symbolic Modelling* The Developing Company Press, 2000.

22. Levi-Strauss, Claude. *Tristes Tropiques*. Translators John and Doreen Weightman New York: Pocket Books, 1977.

23. Lieberson, Stanley and Freda B. Lynn. "Barking Up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science." *Annual Review of Sociology* 28 (2002): 1-19.

24. Lou, Dan, Longbing Cao, Jiarui Ni, and Li Liu. "Building Agent Service Oriented Multi-Agent Systems." *Agent and Multi-Agent Systems: Technologies and Applications* Berlin/Heidelberg : Springer, 2007.

25. Macy, Michael W., and Robert Willer. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28 (2002): 143-166.

26. Marks, Robert E. "Validation and Functional Complexity." Australian Graduate School of Management University of New South Wales Sydney NSW Australia: University of New South Wales, 2002.

27. McKelvey, Bill. "Model-Centered Organization Science Epistemology." *Companion to Organizations.* J. A. C. Baum, editor: Thousand Oaks, CA: Sage, 2002: 752–780.

28. McKelvey, Bill. "Quasi-Natural Organization Science ." *Organization Science* 8, no. 4 (1997): 352-380 .

29. McPherson, Miller, Lynn Smith-Livin, and James M. Cook. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27 (2001).

30. Merriam-Webster Online Dictionary. "Model." Web page, [accessed May 2008]. Available at

http://www.merriam-webster.com/dictionary/model.

31. Merton, Robert K. *Social Theory and Social Structure*. Second ed. New York, NY: Free Press, 1957.

32. Moore, Keefe E. *Alternative Approaches to Validation; Policy and Capability Studies* Dstl, WP13453. Porton Down, Salisbury, Wiltshire, UK: Defense Science and Technology Laboratory UK, 2005.

33. Morgan, Mary S. "Imagination and Imaging in Model Building." *Philosophy of Science* 71 (2004): 752-766.

34. Morgan, Mary S. and Margaret Morrison. "Learning From Models." *Models As Mediators; Perspectives on Natural and Social Science.* Mary S. Morgan, and Margaret Morrison, editors Cambridge, United Kingdom: Cambridge University Press, 1999: 347-388.

35. Morrison, Margaret, and Mary S. Morgan. "Introduction." *Models As Mediators; Perspectives on Natural and Social Science.* Mary S. Morgan, and Margaret Morrison, editors. Cambridge, UK: Cambridge University Press, 1999: 1-9

36. *Morrison, Margaret. "Models As Mediating Instruments." Models As Mediators; Perspectives on Natural and Social Science. Mary S. Morgan, and Margaret Morrison, editors. Cambridge, UK: Cambridge University Press, 1999: 10-37.*

37. National Research Council. *Behavioral Modeling and Simulation: From Individuals to Societies* Board on Behavioral Cognitive and Sensory Sciences Division of Behavioral and Social Sciences and Education Committee on Organizational Modeling. Washington, DC: The National Academies Press, 2008.

38. Ordonez. Carlos. "Building Statistical Models and Scoring With UDFs." International Conference on Management of Data Archives, Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data . New York, NY: ACM Press, 2007: 1005-1016.

39. Oreskes, Naomi. "Evaluation (Not Validation) of Quantitative Models." *Environmental Health Perspectives* no. 106, suppl. 6 (1998): 1453-1460.

40. Prietula, Michael J., Kathleen M. Carley, and Les Gasser, editors. *Simulating Organizations: Computational Models of Institutions and Groups*. 1st ed. Cambridge, Massachusetts: The MIT Press, 1998.

41. Reeves, Carol, Author. *The Language of Science*. New York, New York: Routledge, 2005.

42. Resnyansky, Lucy. "Conceptualization of Terrorism in Modeling Tools: Critical Reflexive Approach." *Prometheus* 24, no. 4 (2006).

43. Defence Science and Technology Organization, Command and Control Division, Australia *Integration of Social Sciences in Terrorism Modeling: Issues, Problems and Recommendations*.

Defence Science and Technology Organization, Command and Control Division, Edinburgh, Australia,2007

44. Richiardi, Matteo, Roberto Leombruni, Nicole Saam, and Michele Sonnessa. "Common Protocol for Agent-Based Social Simulation." *Journal of Artificial Societies and Social Simulation* 9, no. 1 (2006).

45. Saltelli, Andrea. "The Critique of Modeling and Sensitivity Analysis in the Scientific Discourse. An Overview of Good Practices." Washington, DC: TAUC, 2006.

46. Smith, K., H. Brighton, and S. Kirby. *Language Evolution in a Multi-Agent Model: the Cultural Emergence of Compositional Structure*. Language Evolution and Computation Research Unit, Theoretical and Applied Linguistics Edinburgh, UK: University of Edinburgh, 2003.

47. Summers, J. E., R. F. Gragg, and R. J. Soukup. "Topography Measurement of Scale-Model Representations of the Rough Ocean Bottom by Touch-Trigger Probe and Its Implications for Spectral Characterization." *Oceans* (2006): 1-6.

48. Tannenbaum, Lawrence V. "And So We Model: The Ineffective Use of Mathematical Models in Ecological Risk Assessments (Letter to the Editor)." *Integrated Environmental Assessment and Management* 3 (2007): 473-475.

49. Turnley, Jessica G.. 2005. Validation issues in computational social simulations. Paper presented at 3rd UCLA Lake Arrowhead Conference on Human Complex Systems May 2005. http://www.hcs.ucla.edu/lake-arrowhead-2005/HCS2005_JessicaTurnley2.pdf

50. Udehn, Lars. "The Changing Face of Methodological Individualism." *Annual Review of Sociology* 28 (2002): 479-507.

51. Wasserman, Stanley, and Katherine Faust. *Social Network Analysis: Methods and Applications*. New York, New York: Cambridge University Press, 1999.

52. Weick, Karl E., *Sensemaking in Organizations*. Thousand Oaks, California: Sage Publications, 1995.

53. "What is System Dynamics? - A Closer Look." Web page, [accessed 13 October 2007]. Available at http://www.sportsbusinesssims.com/system.dynamics.htm.

54. Williams, E. J. *Verification and Validation in Industrial Simulation*. *Proceedings of the Summer Computer Simulation Conference* (1998): 57-62.

55. Windrum, Pail, Giorgio Fagiolo, and Alessio Moneta. "Empirical Validation of Agent-Based Models: Alternatives and Prospects." *Journal of Artificial Societies and Social Simulation* 10, no. 2 (2007).