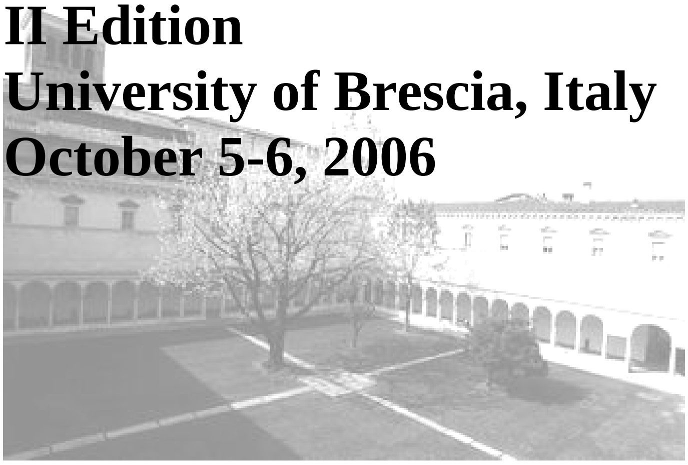
EPOS 2006

Epistemological Perspectives on Simulation

II Edition
University of Brescia, Italy
October 5-6, 2006



Narrative Scenarios, Mediating Formalisms, and the Agent-Based Simulation of Land Use Change

Nicholas M. Gotts and J. Gary Polhill*

The paper argues that the kinds of system studied using agent-based simulation are intuitively, and to a considerable extent scientifically, understood through natural language *narrative scenarios*, and that finding systematic and well-founded ways to relate such scenarios to simulation models, and in particular to their outputs, is important in both scientific and policy-related applications of agent-based simulation. A projected approach to the constellation of problems this raises – which derive from the gulf between the semantics of natural and programming languages - is outlined, involving the use of mediating formalisms: ontologies and specialised formalisms for qualitative representation and reasoning. Examples are derived primarily from ongoing work on the simulation of land use change.

* Macaulay Institute:{n.gotts,g.polhill}@macaulay.ac.uk. Nicholas M. Gotts is the corresponding author.

1. Introduction

The paper argues that simulation in general, and agent-based simulation in particular, can benefit from the development and use of what we will call *mediating formalisms*, to link natural language texts on the one hand, with the program code and the input-output behaviour of computational simulation models on the other. These formalisms should have strong qualitative aspects, be computer-readable, and, where possible, support computationally tractable reasoning. The paper concentrates on the potential role of such formalisms in relation to the potential uses of *narrative scenarios* in two ongoing research projects building spatially explicit agent-based models of land use change, and particularly the role of social networks among farmers in mediating such change: a large-scale project on "Protection and Enhancement of Landscapes and Rural Communities" in Scotland, and the CAVES (Complexity, Agents, Volatility, Evidence and Scale) project (http://cfpm.org/caves/). These projects are not described here, but the ideas outlined have emerged from them, and examples from this domain are used. A narrative scenario, in the sense used here, is a description of the past, or a possible future, with a strong temporal component: a story about what has happened or might happen. Such scenarios may be produced by groups of experts or "stakeholders", then used as an input to simulation modelling; or they may be produced from simulation output, then compared with descriptions of real-world sequences of events, and/or presented to experts or stakeholders for judgements of plausibility.

The structure of the paper is as follows: section 2 provides motivation, while section 3 outlines possible mediating formalisms; section 4 considers how to get from natural language narrative scenarios, via mediating formalisms, to simulation models, and more birefly, how to travel in the opposite direction; section 5 is a very brief conclusion. It should be stressed that this paper sketches an ambitious programme of work currently at an early stage.

2. Motivation: Narrative Scenarios and Agent-Based Simulation

The kinds of system we study using social simulation are intuitively (and to a considerable extent scientifically) understood using narratives: natural-language stories about what has happened, might happen, or can be imagined to happen, almost always including attributions of causality as an important aspect. Such narratives can be divided into a number of classes, of which the following are the most relevant here:

- *Historical narratives*: accounts of what has actually happened over some period in the past, within some geographical area and domain of human activity, or to some person or group of people. We will be concerned here with narratives which are intended by their authors to be accurate – not to include false or misleading statements. Even if the author(s) of a historical narrative strive for accuracy, it may not be achieved; and any such narrative is bound to be selective in terms of what is included and what left out.

- *Possible future narratives*: narratives describing what might happen in future. Such narratives may be intended simply for entertainment, but they may also be produced in order to help participants in their production, and others, prepare for future contingencies, or to justify current policies or argue for particular future ones.
- *Counterfactual narratives* or "alternative histories" . Again, such narratives may be intended purely for entertainment; but they may also be intended to illuminate the causal structure of what actually happened (e.g. Schmalberger 1998).
- *Simulation narratives* derived from simulations or games. Agent-based models, and computer games such as those used in military strategy training, produce sequences of events which can either be considered as occurring within a computational context – as steps in the implementation of an interactive program – or treated as the basis of possible future or counterfactual narratives about the real world.

Historical narratives are an important aspect of what might be called historical social sciences, which include economic history, aspects of social and political history and political science, archeology, paleoanthropology, and historical geography among others. We are interested here in approaches which aspire to be scientific, or at least to objectivity (there are also important in approaches to history and social research which do not so aspire, but these are not considered). The aims of the historical social sciences include the following, to all of which narrative scenarios are centrally relevant:

- Reconstructing events and sequences of events. What constitutes an "event", and how we know whether an event occurred, varies from case to case. While "James McDonald sold Danesbridge Farm to John Robertson for £250,000 on 14th August 1997"[1] is a straightforward case, and the ways we might go about verifying or falsifying it are clear (seeking documentary evidence, interviewing the participants and other witnesses), "McDonald became convinced there was no future in farming during the mid-1990s", "The price of farmland in Scotland rose by 35% between 1990 and 2000" and "Many Scottish farmers became convinced there was no future in farming around the time of the CAP reform of 2003" are, for different reasons, much less straightforward. In the first of the three, the event concerned is an apparently private one for which we have to rely on McDonald's memory – although we might seek evidence of statements he made or actions he took during the period concerned – and its temporal location is vague. In the

[1] The example is an invented one.

second, the “event” is a summary description of a large set of events (sales of farmland); we might question whether it is an “event” at all, and if it is, what information we need about individual sales to judge whether it is true: for example, what if records of some sales are missing, or if it is uncertain whether some parcels of land sold count as “farmland”? The third case combines the difficulties arising in the first two, and adds additional vagueness in its use of “Many”. Turning from single events to sequences of events, narrative scenarios become indispensable to understanding and to further investigation: without the temporal and causal structure that a narrative provides, all we have is a set of unconnected happenings.

- Explaining particular events. Explanation in the historical social sciences rarely takes the form, frequently found in the physical sciences, of specifying a set of initial conditions and a presumed natural law which ensures that given those initial conditions, the event to be explained must occur (there may be rather trivial exceptions: e.g. explaining the death of King Charles I of England and Scotland, given the initial conditions that he was a human being and that his head was cut off). Rather, an (implicit or explicit) range of possibilities is considered, and causal factors are specified for which it is claimed that, had they been different, the event which did occur would have been impossible or less likely, and some alternative(s) certain or more likely. For example, we might explain Charles’s loss of the Civil War (as opposed to his winning it, a negotiated settlement being reached, or a stalemate occurring) in terms of his personal qualities, those of his enemies, the balance of ideological and/or class forces in England in the 1640s, innovations in weaponry and military tactics, the interaction between English, Scottish and Irish politics – or more plausibly, some combination of these. Whether a particular explanation of an event is judged likely to be true frequently depends on the theoretical (and often, political) stance of those making the judgment. However, *any* such explanation must refer to events preceding the one to be explained.

- Discovering regularities across time and space. Two of the “events” referred to above could also be described as regularities: a rise in the price of farmland across Scotland, and the conclusions of “many” farmers following CAP reform. Other regularities, or patterns of events, are more complex, taking the form of correlations between variables, e.g. between agricultural subsidy levels and the price of farmland; or of spatio-temporal patterns of change, such as the spread of agricultural innovations (Feder and Umali 1993). Some cases of correlations between variables can be investigated without involving narrative scenarios: those which occur at one time but across space, for instance. However, describing co-variation across time at a single spatial location, or co-variation patterns involving both space and time, necessarily involves a narrative element.

- Explaining regularities across time and space. Sometimes the explanation of a regularity may take something like a law-and-initial-conditions form. For example, when the average price of a commodity rises, demand for it generally falls, while supply (perhaps after a lag) rises. Specific instances of such patterns of events (which are themselves regularities, as they necessarily involve multiple attempts to buy and sell) may be explained by reference to this economic "law". However, it is worth noting that such "laws" do not always hold: for example, a price rise may be taken as a signal that further rises will occur, boosting demand. Frequently, what appears to be involved in explaining social or historical regularities is the specification of a mechanism or principle which is claimed to underlie instances of the regularity, and which will produce such instances *if nothing else interferes*. This can also be said of most laws in the physical sciences – but there, the experimental method is widely used in attempts to ensure that this implicit condition is met. The mechanism or principle referred to may or may not be part of a broader theory of social or historical processes.

- Explaining possibilities: how could entities/events of type X exist/happen? A central theme in recent social (and biological) science is the explanation of altruistic behaviour (see Gotts, Polhill and Law 2003 for a review): given the obvious advantages of selfishness (in economic and genetic terms), why does anyone ever behave altruistically – that is, so as to benefit someone else at their own expense? There are in fact a number of plausible candidate explanations, with controversy continuing about which of them contribute.

The last three classes of aim listed above in particular indicate the extent to which the historical social sciences depend on the comparative method: finding events or groups of events which share important features, but differ in one or more crucial respects; and which illuminate regularities across space and time, and patterns of causal influence. If we wish to investigate the phenomenon of rising Scottish farmland prices in the 1990s, for example, we already have a class of events with many features in common. We may proceed to divide them into subclasses (by size of farm, location of farm, date of sale main agricultural products, age of farmer), and also to compare the change in Scottish farmland prices in the 1990s with changes in England or Wales over the same period, and with changes in Scotland over the 1980s and 1970s, in order to explain the observed regularity, fit it into wider patterns, and redescribe it in more illuminating ways, picking out the key factors from the mass of detail. While maps, diagrams and tables can all play important parts in presenting the results of such investigations, none of these will be intelligible without a connecting narrative scenario.

In addition to asking whether a historical narrative scenario is accurate – in its account of the course of events, and in causal attribution – we can also ask whether it is *adequate*, in the sense of mentioning the most important events and causal connections: an account of the battle of Waterloo which focused solely on the movement of a single soldier – even Napoleon – would clearly be inadequate. Adequacy must of course be assessed relative to the narrative's length, and function (a biography of Napoleon would reasonably concentrate more on his movements than a general history). Given a length and function, the adequacy of two accurate narratives of the same sequence of events might be compared by asking whether either could be a true description only of a proper subset of those the other could truly describe: adequacy thus defines a partial order on true historical narrative scenarios.

Turning to possible future scenarios, these are far more relevant to policy development than to scientific investigation; and questions about whether they are accurate cannot be answered until the time they refer to, and is in general irrelevant: they can be of considerable use even if the events they describe never happen. However, judgments of their *plausibility* are intrinsic to their use in policy development: only if it is judged that the sequence of events described in a possible future narrative scenario could occur, and that sequences of events resembling it in crucial respects are at least reasonably likely to occur, can it perform the function of helping policy professionals or stakeholders prepare for the future. Judgments of plausibility may be made by asking a range of experts or stakeholders, different from those producing the scenario, whether they consider it plausible; and if not, what parts or aspects of it are implausible, and why; or by examining it in the light of particular theories of social and historical processes.

Counterfactual narratives, perhaps surprisingly, are used in historical social sciences, notably in political and military history (Schmalberger 1998), epidemiology (Kay, Prüss and Corvalan 2000) and macroeconomics (Cooper 2004); but many of the same questions concerning plausibility, and the same range of possibilities for assessing this, arise as for possible future scenarios. However, we also note that counterfactual scenarios have an important role in assessing the attributions of causality in historical narrative scenarios: if a causal attribution is valid, then changing a factor to which an important causal role is attributed in bringing about some event should lead plausibly to a scenario in which that event does not occur.

The fourth class of narrative scenario mentioned above – those derived from computer simulations or games – can have important roles in assessing and improving the accuracy and adequacy of historical narrative scenarios, and in assessing the plausibility of possible future and counterfactual narrative scenarios.

With regard to historical narrative scenarios, while evidence from the real world will always be the final arbiter of descriptions of the course of events and of regularities, and of proposed causal explanations of those events and regularities, simulations have already been used both to test the adequacy of proposed explanations (Lansing and Kremer 1994), and to direct the search for new evidence (Dean et al 1999). We will in this paper sketch a systematic approach which fully recognises the role of historical narrative scenarios in historical social sciences, and makes maximum use of the properties of simulation models. It contains two largely independent elements:

- Taking an existing historical narrative scenario, attempting to build a simulation model which can produce simulation narratives as similar as possible to the historical narrative from theoretically and/or empirically grounded causal mechanisms, while minimising the number and complexity of additional assumptions required. This element is described in some detail in section 4.

- Using an existing simulation model with a collection of two or more parameter sets which can be interpreted in terms of real-world differences in initial conditions. Provided different members of this collection produce outputs which differ systematically in some respect, aspects of the model can then be tested by finding real-world examples of the different kinds of initial conditions, and checking whether real-world outcomes differ in ways corresponding to the differences between simulation outputs. For every such test the model passes, confidence in its verisimilitude should increase. This approach can be adopted without using narrative scenarios, but doing so could greatly strengthen it: because simulation models produce sequences of events, comparison with real-world sequences provides a rich source of information for assessing the model.

With regard to future and counterfactual scenarios, approaches corresponding to the two outlined for historical narratives can be used for different purposes: to test the plausibility of an existing scenario, and to generate new scenarios, respectively.

3. Agent-Based Simulations and Mediating Formalisms

Computer simulation models are, in general, not readily accessible. Even if the code is available, and well-commented, it is difficult and time-consuming for anyone other than the programmer to understand it. If design documents are not available, it is also necessary to reverse-engineer from the program to the design, to link the code with the high-level description of the model given in any source text (such as a journal paper). Even if design documents are available, the source code may not necessarily implement the design as expected, as illustrated by Edmonds and Hales's (2003) exploration of Riolo et al.'s (2001) model.

There are several contexts in which it is desirable to link such a model, its input parameters, or its output, to natural language texts, and this raises additional problems. Issues arising when these texts are academic publications, or interview transcripts, are explored in Polhill and Ziervogel (2006), and Polhill and Gotts (2006), and here we consider issues in relation to narrative scenarios, but the source of the difficulties can be described quite generally: computer programs are written in formal languages, and expressions in natural and formal languages acquire meanings in very different ways. If expressions in a formal language are assigned any meaning at all, this is done by specifying a formal semantics, within which elementary terms are given precisely specified meanings, and complex expressions' meanings are defined using rules for combining the meanings of these elementary terms. In the case of programming languages, the formal semantics will refer to mathematical objects such as partial functions from input to output (denotational semantics: Stoy 1977) or sequences of computational steps (operational semantics: Plotkin 1981) – in any case, the terms of the language refer to nothing beyond the computational domain. Natural language expressions, by contrast, do not have formal semantics, acquiring their meanings from their use in real-world contexts. Expressions within natural languages do not generally have precise meanings, although precision can be increased if desired (if I say "I saw a tall man near the park", my interlocutor may ask "How tall?" or "How near?" for example). Furthermore, natural language is far more expressive than any formal language. The claim made here is that intermediate languages, formal but designed to refer to and support reasoning about real-world things and processes, can help bridge the natural language – programming language gap. In this section, we outline possible intermediate formalisms.

Within computer science, an ontology has been defined as "a formal, explicit specification of a shared conceptualisation" (Gruber 1993). Ontologies formulate relations between the meanings of a set of terms, combining taxonomies of concepts with information about relations that may hold between entities belonging to specified elements of the taxonomy (e.g. "woman" and "man" might be immediate subconcepts of "human being", with additional information specifying possible biological and familial relationships between instances of these concepts). Relations, as well as concepts, can be given a taxonomy: so "father-of" and "mother-of" would both be subrelations (specialisations) of "parent-of". Christley et al. (2004) and Polhill and Gotts (2006) have already argued that ontologies can be used to address issues caused by the ad-hoc way in which agent-based models are programmed. This paper argues that they can also serve to link simulation models and their output to narrative scenarios.

Useful work with ontologies requires a formalism in which to express them: OWL (Antoniou and van Harmelen 2004) appears to be the most widely used formalism, is supported by the semantic web community, is compatible with some of the most useful ontology-related software available (notably Protégé (http://protege.stanford.edu/)) and has a sound logical basis in the description logic SHIQ (Baader., Horrocks and Sattler 2004) the formal properties of which are well-understood. Using automated reasoning software, an OWL ontology can be checked for consistency, satisfiability of and equivalences among concepts can be inferred, and since OWL also allows the representation of particular instances, such instances can be inferred to be examples of one or more appropriately defined concepts.

How do we know what a simulation model, a part of that model, or its input and output, represent? In general, the model is described in natural language, possibly accompanied by diagrams, and this description is accompanied by a description of the real world entity or situation (or if the model is less specific, the type of entity or situation), that the model is intended to represent. This again is generally couched in some combination of natural language and diagrams; and in fact the description of the model, and of what it is intended to represent, may not be distinguished. There will always be many relevant aspects of the real world that are not represented in the model, and conversely, many components of the software implementing the simulation model that are not intended to represent aspects of the real world, but are necessary in order to produce a working program. To the extent that it is specified exactly which aspects of the real world are represented in the model, we can say that a *conceptual model* has been defined. To the extent that it is specified exactly which components of the model represent these aspects of the real world, we can say that the relationship of this conceptual model to the simulation model has been pinned down.

We contend that both the conceptual model, and its relationship to the simulation model, can be greatly clarified by the use of ontologies. We stress that this does *not* require the existence of an ontology common to the social simulation community (let alone to social scientists in general): the point is for the authors of a particular simulation model or set of models to share *with* wider communities a precise specification of what they intend their model (and its inputs and outputs) to represent, and which parts of the computer code concerned are intended to have representational significance. In fact, a number of ontologies are needed: our ideas on this are still evolving, but figure 1 illustrates one possible setup, showing four types of ontology:

- It is easiest to start at the right-hand side of the figure. Here we show a scenario ontology, represented by the darkest and frontmost oval (with other scenario ontologies indicated behind it). The concepts and relations in a scenario ontology represent the types of entity (and individual entities) that exist in a part of the real world – or of a possible world as envisaged in a possible future or counterfactual narrative scenario. To avoid complications, we shall assume in this section that it is part of the real world that is represented. Note that the scenario may, but need not, be described by a natural language narrative of the kind discussed above – or indeed, multiple such narratives.

- To the left of the figure, we show a model ontology, again with others of the same kind indicated behind it. The concepts and relations in a model ontology represent the types of *software* entity (and individual entities) that exist in a simulation model: instances of the concepts in a model ontology will be pieces of code or of data, depending on the concept.

- At the top of the figure is the domain structure ontology. The concepts and relations in this are intended to capture what is common to the structure of a set of real-world scenarios and a set of models. Each concept in a scenario ontology will be a subconcept (specialisation) of some concept in the domain structure ontology, as will each concept in a model ontology. Each model ontology or scenario ontology will "import" the domain structure ontology (copy its structure and contents), then add its own subconcepts and subrelations.

- The bottom of the figure contains the representation ontology. This imports one scenario ontology and one model ontology, and has the sole purpose of defining the relationship between the two, linking real-world concepts and instances with those representing them in the model. It will include two additional concepts, heading the taxonomies of real-world and software entities.

In order to give a more concrete idea of what constructing ontologies involves, we include here a draft of the upper layers of a domain structure ontology for use in the land use domain – that is, in relating real-world land use scenarios to agent-based models of land use change. The concepts in the ontology are described in real-world terms, but are intended to cover both real-world entities, and the software entities corresponding to them in an agent-based model. The lowest-level entries are *examples* of (relatively) low-level *concepts* (*not* instances). The subconcepts below any specific concept are not taken to be exhaustive.

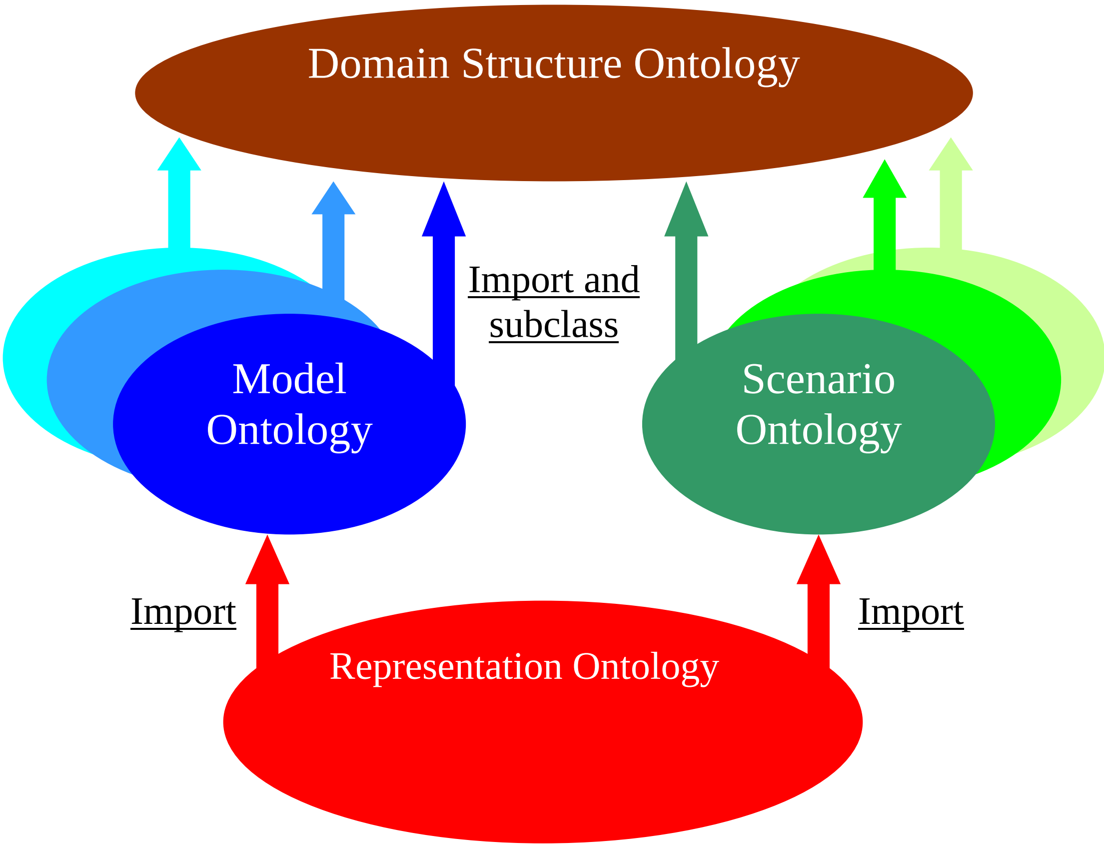


Figure1: Ontology relationships.

Particular [The top-level node: anything that cannot itself have instances.]

- **Endurant** [Entities that "are 'in time', they are 'wholly present' (all their proper parts are present) at any time of their existence." (Masolo, Borgo et al 2003). They are "things" rather than "processes", and are contrasted with "Perdurants" – see below.]
  - **PhysicalEndurant**
    - **AmountOfMatter** [For example, a tonne of grain, a litre of water.]
    - **PhysicalThing**
      - **PhysicalObject** [A physical object is something you can pick up and throw – if you are the right size.]
        - **LivingThing**
          - **NonHumanOrganism**
            - **DomesticAnimal**
              - Sheep, Cow…
            - **CropPlant**
              - MaizePlant,TomatoPlant …
          - **HumanBeing**
            - Adult, FemaleHumanBeing…
        - **NonLivingPhysicalObject**
          - Tractor, Fence…
      - **FieldOfCrop** [Taken to be a conglomeration of PhysicalObjects and AmountsOfMatter.]
        - FieldOfMaize, FieldOfTomatoes
      - **HerdOfAnimals** [A collection of DomesticAnimals. Note that a herd can remain "the same herd" even while individual animals come and go.]
        - HerdOfSheep, HerdOfCows…
    - **Feature** [A "Feature" is dependent on a specific PhysicalThing, to which it "belongs".]
      - GapInFence, SkinOfCow…
  - **NonPhysicalEndurant**
    - **MentalThing**
      - Memory, Attitude…
    - **SocialThing**
      - **ImpersonalSocialThing**
        - Law, Currency…
      - **SocialFormation**
        - EthnicGroup, Class…
        - **SocialNetwork**
          - KinshipNetwork, FriendshipNetwork…
        - **Organisation**

FarmersUnion, Corporation…

SocialRole

FormalSocialRole

Spouse, Landlord…

InformalSocialRole

Friend, RoleModel…

Agent [This subsumes HumanBeing and Organisation, producing the only examples of multiple inheritance in this concept hierarchy. In other domains, non-human animals and other kinds of social formation than organisations could also be agents.]

Perdurant [States, events, activities, processes – things which have proper parts in different temporal locations. All Perdurants involve Endurants as "participants". All Endurants participate in Perdurants.]

AgentivePerdurant [A Perdurant that involves intentionality on the part of at least one Agent.]

CourseOfAction

LandUse [That is, apply a land use to a specific LandParcel or LandParcels.], Irrigate…

OneOffAction

Sell, Buy, Ask, Tell…

NonAgentivePerdurant

OnGoingOccurrence

IllHealthBout, Infestation…

OneOffOccurrence

PriceChange, Death…

Location [Masolo, Borgo et al. (2003) class these (calling them "Regions") as abstract. They are related to the real world in a different way: Endurants and Perdurants "occupy" Locations, and Locations are (at least in the Spatial and Temporal cases) necessarily defined in relation to Endurants and/or Perdurants. Masolo, Borgo et al (2003) in fact use a more complex way of representing endurant/perdurant relationships with locations, involving an intervening category of "Qualities".]

SpatialLocation

Province,LandParcel,LandHolding…

TemporalLocation

Year, January

ConceptualLocation [A possible place to put locations in "conceptual spaces" (Gärdenfors 2000) – a way of handling object-attribute-value information (although Protégé allows this to be done using "Properties"). Monetary value, weight, colour are possible examples]

Abstract

NetworkThing [both PhysicalThings and SocialNetworks have a network structure. However, they can change their topology while remaining the same network – which a network considered as a mathematical structure cannot. It therefore seems best to regard the current mathematical structure of a real-world Network as a relationship it has with that mathematical structure.]

Network

Tree, DirectedNetwork, UndirectedNetwork

Node

Link

DirectedLink, UndirectedLink, LabelledLink

Clique

Procedure [A CourseOfAction may involve following a Procedure (as written down, or encoded in memory – and we need to distinguish between the two.]

LandUseProcedure, VeterinaryProcedure…

Ontologies can be used to describe processes and spatio-temporal relationships, but to do so with the richness necessary to capture the structure of human-produced narrative scenarios, or the output of agent-based simulation models (particularly where these are spatially explicit), additional, specialised formalisms are likely to be required. The most promising include James Allen's temporal interval calculus (Allen and Kautz 1985), the Region Connection Calculus (RCC) (Cohn et al 1997), and Qualitative Differential Equations (QDE) (Kuipers 2001). These have in common that they are primarily qualitative in nature, an important advantage in mediating between natural language narratives (which almost always have central qualitative aspects) and simulation output.

Allen's temporal interval calculus is based on a set of 13 possible qualitative relations which two temporal intervals can have. These are illustrated in figure 2. Clearly, any two continuous stretches of time must have one, and only one, of these relations; but partial information about which of them holds can be dealt with by specifying a subset of the 13 to which the true relation is asserted to belong. RCC is a closely related set of topological spatial relations, ilustrated in figure 3. Again, any two regions (which can be of any dimensionality, provided both are of the same dimensionality) must have one and only one of these relations; and in this case, the regions can even consist of multiple pieces. The names of the eight relations as given in figure 3 are abbreviations standing for equal (EQ), externally connected (EC), disconnected (DC), partial overlap (PO), tangential proper part (TPP) and non-tangential proper part (NTPP), plus inverses of TPP and NTPP (TPPi and NTTPi). Although RCC, like Allen's calculus, was originally formulated to represent commonsense reasoning, without a specific model in mind, Gotts (1996) shows that it can be given an interpretation in terms of conventional point-set topology, with regions being interpreted as the non-empty regular closed sets of a topological space (basically, this means areas that include their own boundaries and do not have any weird geometric properties such as isolated points subtracted from their interiors), and two regions A and B being connected (~DC(A,B)), if and only if they share at least one point. Two regions are EC if they share only boundary points, and the other relations can be given similarly straightforward definitions.

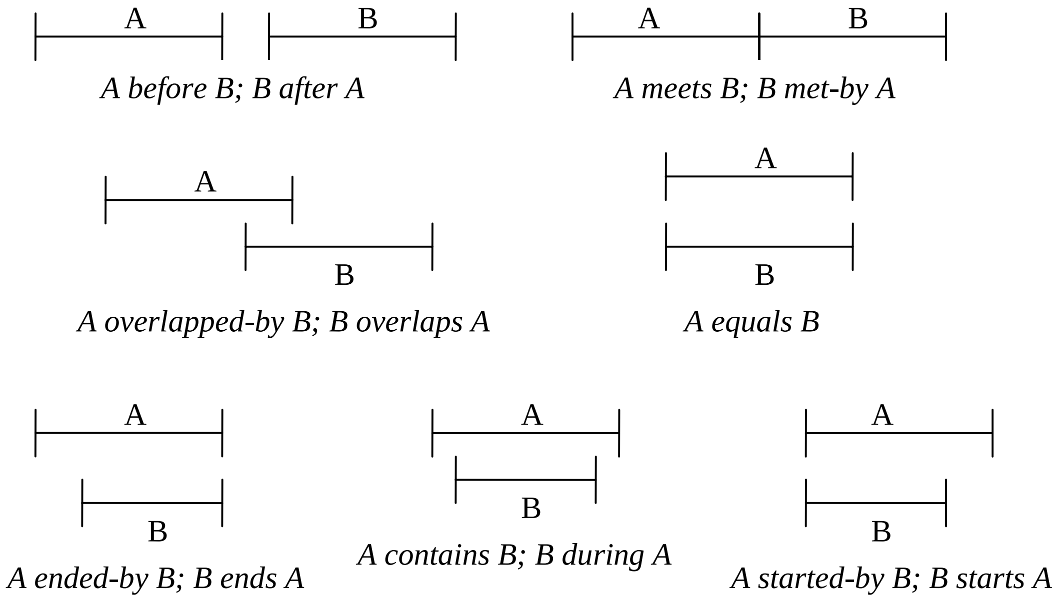


Figure 2: Allen’s 13 qualitative temporal interval relations.

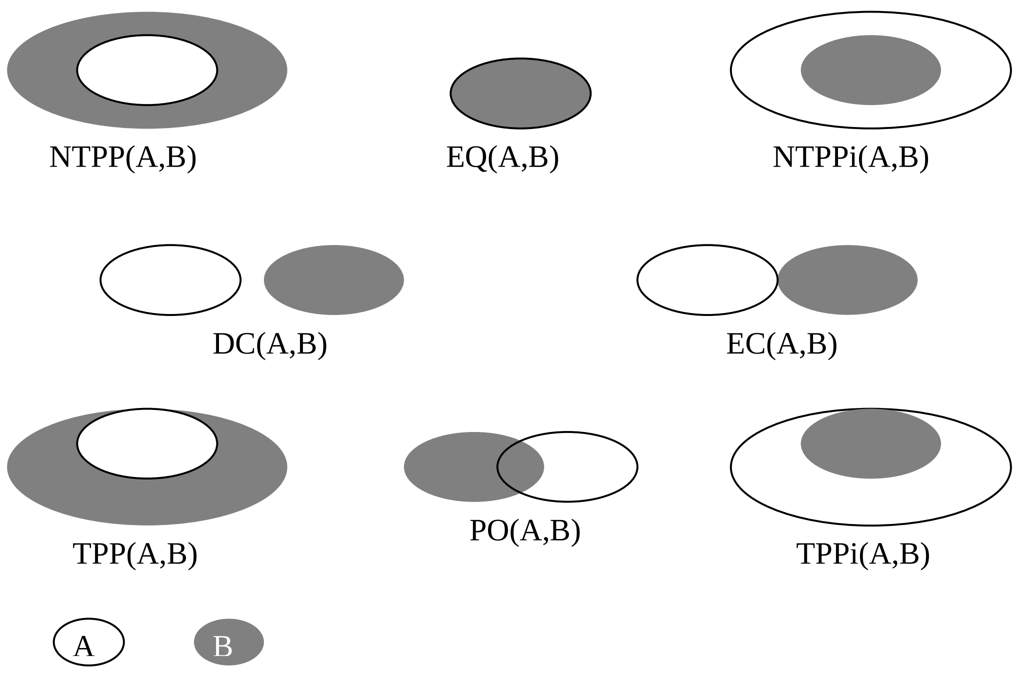


Figure 3: RCC’s eight basic topological relations.

Both Allen's calculus and RCC have been extensively investigated with regard to their computational properties. For RCC, the full first-order language is undecideable (Grzegorczyk 1951, Gotts 1996) – that is, if we allow quantification over variables, there is no algorithm to determine whether an arbitrary formula is valid; and it would be surprising if the situation were different with Allen's calculus, although its first-order language does not appear to have been investigated explicitly. For both, the "constraint language" is decidable – that is, given any finite set of relations among regions, it can be determined in finite time whether it could be satisfied; but the problem is NP-hard – assuming the truth of an unproven but almost universally accepted mathematical conjecture, the difficulty of the problem grows exponentially with the number of relations considered (Vilain, Kautz and van Beek 1990, Bennett 1994, 1996). For both calculi, there have also been investigations finding computationally tractable subsets of this class of problems, where difficulty can be shown to increase with some polynomial function of problem size (Vilain, Kautz and van Beek 1990, Bennett 1994, 1996, Renz and Nebel 1999). Bennett et al (2002) show that RCC and Allen's calculus can be combined to express spatial and temporal relations simultaneously without losing decidability, and Gerevini and Nebel (2002) show that it is possible to add the requirement that changes in spatial regions be continuous over time, or that regions should remain the same size, without losing this property.

While Allen's calculus and RCC both leave large gaps in what can be said about spatial and temporal properties and relations (they have no metric component, and RCC as presented above says nothing about shape, although there are extensions allowing the convexity of a region to be asserted (Davis, Gotts and Cohn 1999) and for the existence of regions without definite boundaries (Cohn et al 1997)), what they can express is highly relevant to questions of causality: in the everyday world, a cause must not come after its effect, and the two must be spatio-temporally connected. Questions of qualitative temporal relationship are also central to the description of plans and procedures, and concerns about boundaries and the spatial continuity of parcels of land are of particular concern to land managers, as well as to those concerned with the ecological implications of land use change. Given this, and the extensive literature about the expressivity and computational properties of the two formalisms, we intend to investigate combining them with ontologies in representing narrative scenarios. A possible approach is discussed in the next section.

Also of interest in this connection is the QDE (qualitative differential equation) formalism (Kuipers 2001). In a QDE representation of a system, each numerical variable is assigned a *quantity space*: a finite, totally ordered set of qualitatively important "landmark values". In a land use context, taking rainfall over the growing season as an example, the landmark values might be those (not necessarily specified exactly) necessary to make various crops viable. Variables can be related by algebraic constraints (e.g., the return from a crop is the multiple of its yield with the price per unit weight), or by differential ones (e.g. the relation of the rate of pollutant inflow to a closed body of water to the quantity of that pollutant in the lake); but it may also be specified that increasing one variable will increase (or decrease) another, without further specifying the form of the dependence: these are called functional constraints. Transition conditions may be attached to a set of QDE constraints, specifying when they will cease to apply (e.g. increasing the number of sheep in a field will increase their total rate of weight gain, but only up to the point where they eat the grass faster than it can grow). As with Allen's calculus and RCC, we return to this formalism in the next section.

4. From Narrative Scenario to Simulation… and Back

We describe here a proposed development of the proposed "story and simulation" approach to scenario development (European Environment Agency 2001):

> "The storyline describes in story form how relevant events unfold in the future, while the model calculations complement the storyline by presenting numerical estimates of future environmental indicators and helping to maintain the consistency of the storyline."

A small set of storylines, each based on different assumptions, is developed by a "scenario panel" of experts and/or stakeholders, usually in a workshop held for that purpose. A modelling team then creates a simulation to match each storyline; the additional (quantitative) detail and any caveats about consistency are fed back to the scenario panel, this process being repeated as necessary. However, the process of getting from storyline to model is not described in any detail, nor is the kind of feedback given. We suggest the following sequence of steps in this process, beginning with a set of possible future natural language narrative scenarios, an initial domain structure ontology, a set of qualitative formalisms such as those described in the preceding section, and possibly a set of existing scenario ontologies. All the steps listed require further decomposition, in the course of which they may turn out to require revision. We will assume here that the domain structure ontology includes at least the concepts in the taxonomy in section 3:

1. Identify concepts in the natural language narrative scenarios which might be relevant to simulation modelling, and instances of those concepts. Natural language processing software such as GATE (Cunningham, Maynard et al 2002), which can pick out words and phrases using syntactic and semantic criteria, may well be useful here and in later stages, but our judgment is that automation of this and later stages is a long way from feasibility.
    a. Identify the Agents (HumanBeings or SocialFormations) in the narratives.
    b. Identify the non-agentive Endurants (Physical and NonPhysical).

c. Identify the AgentivePerdurants, along with the Endurants involved in each.

d. Identify the NonAgentivePerdurants, along with the Endurants involved in each.

e. For each Perdurant, identify any Locations mentioned in connection with it. There may be SpatialLocations and/or TemporalLocations specifying where and when the Perdurant happened; there may also be candidates for treatment as conceptualLocations.

2. Identify relations between instances of the concepts identified in step 1, and properties of those instances. Of particular importance will be relations between the Locations identified in step 1e: existing systematisations of such relations, such as those described in the preceding section, should be part of the domain structure ontology.

3. For each concept identified in step 1, find the most specific concept in the domain structure ontology of which it can be considered a subconcept.

4. Where more than one of the concepts identified in step 1 has the same immediate superconcept in the domain structure ontology, consider whether any of that set of concepts should be grouped under more specific, intermediate concepts. If so, add these to the scenario ontology.

5. Carry out analogues of steps 3-4 for relations and properties.

6. On the basis of commonsense knowledge, stakeholder knowledge and/or existing theoretical and empirical literature, consider what additional real-world concepts, properties and relations need to be added to the scenario ontology.

7. Construct a formal description of *each* narrative scenario, in terms of the scenario ontology and the set of qualitative formalisms described in section 3. This formal description would take the form a labelled, directed graph. Details are still under consideration, but provisionally:

a. There would be a node for each of the specific Endurants, Perdurants and Locations mentioned in that narrative scenario or inferred by combining it with background knowledge, labelled with the concepts of which they are instances.

b. Each Perdurant would have links to all the Endurants involved in it, with a label on the edge identifying the role played by that Endurant.

c. Each Perdurant would also have a link to the TemporalLocation it occupied, and to one or more SpatialLocations (one of these would be the spatial union of all the SpatialLocations involved; further details remain to be decided).

d. Instances of relations between Endurants, between Perdurants, between Endurants or Perdurants and Abstracts, and relations between an Endurant and a Perdurant other than that of participation (e.g., a HumanBeing *learning* of a Perdurant) would also have nodes, with edges linking them to each of the concept-instances involved in the relation-instance, and an edge linking it to the TemporalLocation during which it held (one such node would represent the "universal" TemporalLocation, indicating that a Perdurant or relation linked to it continued or held throughout the time covered by the scenario).

e. Properties of Endurants and Perdurants would also have nodes, with a link to one Endurant or Perdurant, and to a set of TemporalLocations, the latter labelled with the values holding during that TemporalLocation (these could be taken from a QDE-type quantity space in the case of numerical properties).

f. SpatialLocations could have RCC-labelled links with each other, TemporalLocations interval-relation links with each other.

g. Pairs of property-nodes could also be linked to nodes representing QDE algebraic, differential or functional constraints.

8. Decide on a subset of the concepts, relations and properties identified which will be represented in the simulation model to be constructed, and determine how they are to be represented.

9. On the basis of step 8, construct a model ontology. Note that a concept in the scenario ontology is likely to have relations and properties which the model ontology concept for the class of software entity representing that real world entity does not have, and vice versa. Polhill and Gotts (2006) includes some discussion of the information to be encoded in the concepts of a model ontology.

10. For each pair of corresponding concepts (relations, properties) in the scenario and model ontology, consider whether a new concept (relation, property) subsuming both (and nothing else) should be added to the domain structure ontology. This should be done if and only if two concepts (relations, properties) share structural relationships with other items in their respective ontologies which do not hold for their current common superconcept in the domain structure ontology.

11. Build the model. This process may simply modify an existing model, may require a new model to be built within an existing modelling system (again, see Polhill and Gotts (2006) for discussion of modelling systems), or may require an entirely new modelling system.

12. Experiment with the model, exploring its parameter space to discover whether simulation runs giving rise to simulation scenarios similar to the members of the original set of narrative scenarios, and if so, how easily, and where the greatest difficulties lie.

13. Feed back information from step 12 to the experts/stakeholders (along with a description of the simulation model), and if there were indeed significant difficulties in producing simulation narratives similar to the original narrative scenarios, ask whether they can adjust these accordingly.

Given a simulation model based on an ontology, it should be possible to arrange for output from a simulation run to take the form of a directed, labelled graph such as that described in step 8 above. Matching such structures against each other should also be possible, using adaptations of existing graph-matching algorithms such as that described in Feng, Goldstone and Menkov (2004), which is designed for aligning conceptual systems, and allows for post-matching adjustment by the user. This would allow systematic comparison with the output from a different run of the same model; with outputs from different models, and with narrative scenario structures produced from natural language narratives, and perhaps other sources such as time series. We intend to explore how a natural language narrative scenario could be created from such a labelled, directed graph – specifically, whether there are existing natural language generating algorithms that could be adapted to partially automate the process.

5. Conclusions

The paper argues that narrative scenarios play crucial roles in both intuitive and scientific understanding of the kinds of system agent-based social simulation models. Since agent-based models naturally produce outputs with a quasi-temporal structure, ways must be found to match such outputs against natural language narrative scenarios, if such models are to fulfill their potential in historical social sciences and in policy development. However, the transparency which both scientific and policy-related applications require of agent-based simulation modelling in areas such as land use change, cannot be achieved by any combination of program code and natural language description alone. It has been argued that ontologies, combined with existing qualitative formalisms designed to express spatial, temporal and dynamical relationships and associated with useful results concerning expressivity and computational complexity, could be of great benefit in this regard, and specifically in enabling simulation model outputs to be linked to and compared with natural language narrative scenarios.

Acknowledgements

The work for this paper was funded by the Scottish Executive Environment and Rural Affairs Department, and by the European Commission as part of NEST Pathfinder project 12186 (http://cfpm.org/caves/).

References

Allen, J.F. and Kautz, H. (1985) A model of naive temporal reasoning, pp.251-268 in J. Hobbs, J. and Moore, R. (eds), *Formal Theories of the Commonsense World*, Norwood, NJ., 1985, pp. 251-268

Antoniou, G. and van Harmelen, F. (2004) Web Ontology Language: OWL. Ch.4 in Staab, S. and Studer, R. (eds) *Handbook on Ontologies.* Springer

Baader, F., Horrocks, I. and Sattler, U. (2004) Description logics.Ch.1 in Staab, S. and Studer, R. (eds) *Handbook on Ontologies*. Springer

Bennett, B. (1994) Spatial reasoning with propositional logics. In Doyle, J., Sandewall, E. and Torasso, P. (eds) Principles of Knowledge Representation and Reasoning: Proceedings of the 4th International conference (KR94), pp.51-62

Bennett, B. (1996) Modal logics for qualitative spatial reasoning. *Bulletin of the Interest Group in Pure and Applied Logic* (IGPL) 4(1), 23-45.

Bennett, B., Cohn, A.G., Wolter, F. and Zakharyaschev, M. (2002). Multi-dimensional modal logic as a framework for spatio-temporal reasoning. Applied Intelligence 17(3), 239-251

Christley, S., Xiang, X., & Madey, G. (2004) Ontology for agent-based modeling and simulation. In Macal, C. M., Sallach, D., & North, M. J. (eds.) *Proceedings of the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence, co-sponsored by Argonne National Laboratory and The University of Chicago, October 7-9*. http://www.agent2005.anl.gov/Agent2004.pdf

Cohn, A.G., Bennett, B., Gooday, J. and Gotts, N. M. (1997) RCC: a calculus for region based qualitative spatial reasoning. *Geoinformatica* 1, 275-316

Cooper, A. (2004) The impact of interest rates and the housing market on the UK economy *Economic Outlook* 28, 10-18

Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V. (2002) GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*. Philadelphia, July 2002

Davis, E., Gotts, N.M. and Cohn, A.G. (1999) Constraint networks of topological relations and complexity. *Constraints* 4, 241-280

Dean, J.S., Gumerman, G.J., Epstein, J.M., Axtell, R.L., Swedlund, A.C., Parker, M.T. and McCarroll, S. (1999) Understanding Anasazi culture change through agent-based modelling. In Kohler, T. and Gumerman, G. (eds) *Dynamics in Human and Primate Societies*, pp.179-205. Oxford University Press

Edmonds, B., & Hales, D. Replication, replication and replication: Some hard lessons from model alignment. *Journal of Artificial Societies and Social Simulation* **6** (4). http://jasss.soc.surrey.ac.uk/6/4/11.html

European Environment Agency (2001) *Scenarios as Tools for International Environmental Assessments*, Environmental issue report No. 24, Luxembourg Office for Official Publications of the European Communities

Feder, G. and Umali, D.L. (1993) The adoption of agricultural innovations: a review. *Technological Forecasting and Social Change* 43, 215-239

Feng, Y., Goldstone, R. and Menkov, V. (2004) A graph matching algorithm and its application to conceptual system translation. International Journal on Artificial Intelligence Tools 14(1-2), 77-99

Gärdenfors, P. (2000) *Conceptual Spaces: The Geometry of Thought* MIT Press

Gerevini, A. and Nebel, B. (2002) Qualitative Spatio-Temporal Reasoning with RCC-8 and Allen's Interval Calculus: Computational Complexity, *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI 2002)*, pp. 312-316

Gotts, N.M. (1996) *An axiomatic approach to topology for spatial information systems.* Technical report 96.25, School of Computer Studies, University of Leeds.

Gruber T. R. (1993) A translation approach to portable ontology specifications, *Knowledge Acquisition* **5**, 199-220

Grzegorczyk, A. (1951) Undecidability of some topological theories. Fundamenta Mathematicae 38, 137-152

Kay, D., Prüss, A. and Corvalan, C. (2000) Methodology for assessment of environmental burden of disease: Report on the ISEE session on environmental burden of disease, Buffalo, 22 August 2000. WHO/SDE/WSH/00.7, available online at http://www.who.int/docstore/peh/burden/wsh00-7/Methodan1.htm

Kuipers, B. (2001) Qualitative Simulation, in Meyers, R.A. (editor-in-chief) *Encyclopedia of Physical Science and Technology, Third Edition*, pp.287–300.

Lansing, J.S. and Kremer, J.N. (1994) Emergent properties of Balinese water temple networks: coadaptation on a rugged fitness landscape. In Langton, C.G. (ed) Artificial life III, 201-223, Addison-Wesley

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. and Schneider, L. (2003)
The WonderWeb Library of Foundational Ontologies Final Report WonderWeb Deliverable D18, version 1.0, IST Programme of the Commission of the European Communities project number IST-2001-33052

Plotkin, G.D. (1981) *A Structural Approach to Operational Semantics*, Tech. Rep. DAIMI FN-19, Computer Science Department, Aarhus University, Aarhus, Denmark

Polhill, J.G. and Gotts, N.M. (2006) A new approach to modelling frameworks, Paper to be presented at the *First World Conference on Social Simulation (WCSS-06)*, Kyoto, Japan, 21-25 August 2006

Polhill, J.G. and Ziervogel, G. (2006) Using ontologies with case studies: an end-user perspective on OWL, *Second National Conference on e-Social Science (NceSS2006)*, Manchester, 29-30 June 2006

Randell, D.A., Cui, Z. and Cohn, A.G. (1992) A spatial logic based on regions and connection, pp.165-176 in Nebel, B., Swartout, W. and Rich, C. (eds) *Principles of Knowledge Representation and Reasoning: Proceedings, Third International Conference*, Cambridge, Massachusetts, Morgan Kaufman

Renz, J. and Nebel, B. (1999) On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the Region Connection Calculus, *Artificial Intelligence* 108:69-123

Riolo, R. L., Cohen, M. D., & Axelrod, R. (2001) Evolution of cooperation without reciprocity. *Nature* **411**, 441-443

Schmalberger, T. (1998) *Dangerous Liasons: A Theory of Threat Relationships in International Politics*. PhD thesis, available online at http://www.usc.edu/dept/LAS/ir/cews/html_pages/Dangerous_Liaisons/

Stoy, J.E. (1977) *Denotational Semantics: The Scott-Strachey Approach to Programming Language Semantics*, MIT Press, Cambridge, Massachusetts