Chapter 4

Agent-based modelling and economic theory

"When you explain a 'why', you have to be in some framework that you allow something to be true. Otherwise you are perpetually asking why."

Richard Feynman[1]

"How would you physicists like it if you had to survey a bunch of molecules to find out what they planned to do, only to have most of them change their minds anyway, and the government restructure the laws of physics because of some opinion poll?"

'Gaz'[2]

4.1 Introduction to chapter

This chapter introduces the concepts behind **ABM**. It also puts them in the context of ideas about modelling, concentrating particularly on those that have been developed in economics. The reason for this approach is to understand Richard Feynman's quote above: explanation requires knowing what framework the modeller is currently allowing to be true. This is an

[1]In interview on 'Fun to Imagine,' BBC, 1983.

[2]Comment at http://tamino.wordpress.com/2010/03/11/not-a-random-walk/comment-40731

argument that goes back at least to Hume, who imagined a peasant and an artisan wondering why a clock had stopped. The peasant manages no better than 'it does not go right'; the artisan knows more of the inner workings, and suspects the springs or the pendulum (Hume 1739 p.132 in Hoover 2001 p.11). But while it is intuitively obvious, trying to work through its implications for model choices is less so. It is a tricky problem and as Friedman notes:

> "There is no magic formula for wringing knowledge about complicated problems from stubborn facts. No method is proof against incompetent application." (Friedman 1953b p.613)

When discussing economic modelling, examples from the natural sciences present themselves as obvious thought experiments. This chapter uses plenty of them, and looks at arguments made by economists that do the same. A lot is written on the validity of such parallels; critiques of the utility-maximising foundations of neoclassical economics often go no further than pointing out its roots in Newtonian physics as self-evident proof of its absurdity (see section 4.3). But the quote above gets to the heart of the issue. Regularities in human behaviour do exist: how people react to cost changes is central to this thesis. But parallels with physics can only go so far. To pre-empt the conclusions, the argument for modelling people as predictable, reactive objects (perhaps with added noise) must be lightly made. It is unlikely that there will ever be the same robust link between levels of explanation in human behaviour as there are between atomic kinetics and gas theory (see section 4.4.8). Vriend quotes Lucas: economists are "programming robot imitations of people, and there are real limits on what you can get out of that" (Klamer 1984 p.49 in Vriend 1994 p.31).

The structure of the chapter is as follows:

- An explanation of the basics of **ABM**, introducing **Object-Oriented Programming (OOP)** and discussing how it has come to inform model building
- A brief overview of the **ABM** literature, setting the scene for the more specific argument that follows
- 'Mapping the model', a section looking at how theorists from both economics and **ABM** points of view have understood how models should be built. The goal is to think through how to distinguish a good from a bad 'mapping'.
- 'Production and utility in an agent context' grounds the previous discussion by asking how agents with utility and production functions can be built.

4.2 What is agent-based modelling?

In **ABM**, the 'agents' are distinct code objects, programmed to interact with their environment and each other. **ABM** developed in tandem with **OOP**. Though **OOP** has a history going back to the sixties, it was not until the nineties that computing power and programming languages like Java developed enough to mark a 'paradigm shift' in programming (Robinson and Sharp 2009 p.211). This shift provided the soil for **ABM** to flourish.

Writing in 2000, O'Sullivan and Haklay noted that **ABM** was "a rapidly developing field that is already well beyond the scope of any limited survey" (O'Sullivan and Haklay 2000 p.1410), a sentiment shared by one of the few authors brave enough to attempt an **ABM** textbook (Wooldridge 2009 p.xix). **ABM**'s use ranges from the most abstract artificial life to 'autonomous' agents earning their keep controlling real-world infrastructure. As a field, **ABM** still seems as fluid and as expansive as it did ten years ago: little in the way of consolidation has taken place, despite pleas for standardisation and various groups arguing their particular framework offers a one-stop-shop for all things agent-based.

Wooldridge defines objects as "computational entities that encapsulate some state, are able to perform actions, or *methods*, on this state, and communicate by message passing" (Wooldridge 2009 p.28). Objects are created from classes; a real-world metaphor would be that classes are the blueprint and objects the physical form. Thus, a model may have a single 'Firm' class but many 'Firms' created from that blueprint, with their own internal state.

This focus on objects' 'encapsulation' - that 'objects have control over their own internal state' - requires the programmer to think of them as separate entities, and to define relations between those entities very clearly. (The procedural element of programming has not gone away, however, and is still of vital importance, particularly with regard to timings; see section 5.6.3.) A popular online java tutorial (1995) has a simple illustration of encapsulation: if a rider of a bicycle attempts to change gear, the bicycle should have a 'method' that ensures it cannot exceed its gear number. The rider is denied the ability to *force* a gear-change above those available. In coding practice, this mean providing methods to raise and lower gears: the rider may 'request' a gear change, but the bicycle 'decides' if it can be done.

Many of the coding innovations now associated with **OOP** that found their way into **ABM** have a prehistory in procedural programming. Some of these have become emblematic of the agent modelling ethos. For example, Thomas Schelling's segregation model, first presented in 1969 (Schelling 1969) and later described in 'Micromotives and Macrobehaviour' (Schelling, 1978) is often seen as the 'Eve' of agent models (see, for example, Schelling's closing essay (2006) in volume two of the **ACE** handbook), and is used by Krugman as a compact illustration of how a good economic explanation links micro to macro results (Krugman 1996, p.15). Craig Reynold's 'Boids', while originating from a desire to animate flocking, talked of agents as objects (Reynolds 1987). Schelling initially used only pencil and paper (Schelling 2006); Reynolds used an **OOP** extension to LISP.

These models pre-empted the later vital importance of encapsulation. In conjunction with a set of supporting **OOP** coding concepts, this has created a very powerful set of tools for defining how objects interact. This is not just of historical interest: it is hard to overstate the importance of the **OOP** paradigm's approach to the subsequent shape of agent modelling theory and practice. Section 4.4 examines this issue in depth. Before that, the next section gives an overview of **ABM**.

4.3 ABM overview

4.3.1 Agent-based computational economics (ACE)

This thesis uses agents to look for spatial economic outcomes from actors with heterogenous locations. There is very little in the way of 'complex' behaviour, and where it does exist (for instance in the interaction of price-setting and production), it is not the focus of analysis. This is not to deny the uses of **CAS** theory. Evolutionary ideas are particularly important for developing an understanding of how economic growth actually functions (Martin and Sunley 2007, and see section 4.5.2), as well as the evolution of diverse market structures, rather than just postulating a 'Market' (see e.g. Mirowski 2007a discussing 'marketomata'). But a perceived need to keep **ABM** and complexity umbilically linked cuts off some of the more mundane - yet crucial - uses of heterogeneity. (Section 4.4.8 looks at an example of an 'argument from complexity' that illustrates this point.)

Otter *et al.*'s agent paper presenting an economic geography model, for instance, argues it is possible to use "complexity as underlying explanatory variable" (Otter *et al*. 2001 p.1). If 'complexity' can be used this way, so can 'equations' or 'statistics'. It is suggested that "regional economics and geography" are apparently "not sufficient to explain the complex spatial patterns, such as clusters and sprawl, that we encounter" (Otter *et al*. 2001 p.1). Their model is in some sense believed to be 'validated' if it manifests some element considered complex. In another example, Boschma and Martin ask,"in what sense can complexity theory notions (or metaphors), such as the emergence, self-organisation, criticality and so on, be used to conceptualize the economic landscape?" (Boschma and Martin 2007 p.541) This is perhaps a natural consequence of some of the theories of explanation that have grown along with **CAS** theory; see section 4.4.7 on Epstein.

As O'Sullivan and Haklay say, **ACE** is driven by "a view of the 'economy as an evolving complex system' promoted by the Santa Fe Institute", accompanied by a "widespread disillusion with neoclassical equilibrium economics" (O'Sullivan and Haklay 2000 p.1412). Disillusion is a gentle term: the reality for some researchers is a rather more 'year zero' feel, seeing **CAS** as "a pioneering break from a moribund Newtonian worldview" (Manson 2001 p.412). Mirowski, for example, argues (paraphrased by Blaug) that "the whole of neo-classical economics ever since the **marginal revolution** has been an attempt to create an economics that emulates all the essential features of of nineteenth-century physics." (Blaug 1997 p.284)

As Blaug notes, this is rather strong; there was no conscious attempt to emulate physics. Grauwin also points out that the economic interpretation is very different from the physical: "scientific fields assume distinct points of view for defining the 'normal' or 'equilibrium' aggregated state": physics uses entropy where economics has agents finding Nash equilibria and, as they note, "the two approaches lead to radically different outcomes" (Grauwin *et al*. 2009 p.20622). Nevertheless, from a **CAS** perspective, the faults with 'traditional' economics are seen to be self-evident. What Conlisk calls the "strange sacrifices required for the 'ritual purity' of optimisation-only models" (Conlisk 1996 p.686) are considered relics. While understandable at a time when no computers were available, the feeling is - as Mirowski says - economists should "kick the habit of their physics envy and join the 21st century by rethinking the importance of computation and evolution in the way that they approach markets" (Mirowski 2007b p.359). Edmonds goes even further, suggesting that numerical representation itself is questionable. For example, he criticises numerical representation of variety on the basis that not all the dynamics associated with variety (such as evolutionary dynamics) can be described numerically (Edmonds 2004 p.5).

As section 3.2 discussed, **GE** is descended from this 'Newtonian' lineage. Unsurprisingly, then, there is very little overlap between the research agendas of **GE** and **ACE**. **ACE** as a distinct sub-discipline of **ABM** takes it name from the broader school of computational economics. Despite arguments that **ACE** is perfect for re-examining the Marshallian roots of economics (Leijonhufvud 2006), it has tended to follow the 'Santa Fe' research agenda. One editorial contains a good checklist of the concepts of this agenda: "complexity, evolution, auto-organisation... emergence... bounded rationality, inductive reasoning" (Consiglio 2007 p.vi).

The economic coordination problem in particular lends itself to **ABM**'s default focus on interacting agents. Indeed, Tesfatsion defines **ACE** as "the computational study of economies modelled as dynamic systems of interacting agents". The natural question for **ACE** theorists thinking about markets is then: "who does the job of the so-called market adjustment?" (Posada *et al*. 2007 p.102). This requires the modeller to "analyze explicitly how agents interact with each other" (Kirman and Vriend 2001 p.460). Howitt, for one, sees this requirement as a key strength of the approach: "one of the virtues of the **ACE** approach to economics... is that it forces one to make explicit the mechanisms through which individual actions are coordinated, for better or worse" (Howitt 2006 pp.1068). Agents, it is hoped, force theorists to crack open the problem and look at the dynamics inside.

Hayek's short essay, 'the use of knowledge in society' (Hayek 1945 p.529), is something of a talisman for this way of thinking. In particular, his strong aversion to macro-level assumptions fits **ACE** perfectly. He insisted that "we must show how a solution is produced by the interactions of people each of whom possesses only partial knowledge" (Hayek 1945 p.530) - articulating both an assumption that the 'solution' arises from micro-level interactions, and that it must come from 'boundedly rational' actors. Epstein's 'generativist's question' (discussed in-depth in section 4.4.7) - "how could the decentralised local interactions of heterogeneous autonomous agents generate the given regularity?" (Epstein 2006 p.5) - echoes Hayek's take. Some theorists (Vriend 2002 p.2; Miller and Page 2007) can imagine that, had Hayek only been able to access present technology, he would surely have embraced it, and complexity theory along with it.

This Austrian tilt to **ACE** is further reason for the lack of overlap with the kinds of questions **GE** asks. A good example of how this manifests itself is the criticism levelled at the concept of equilibrium. As Blaug notes, “the desideratum of any economic theory is the delineation of an equilibrium end-state” (Blaug 2009 p.224). These end-states are analysed in the same way that basic physical models of stationary systems assume all forces balance to zero and are then able to deduce force values required for that stationary state. In-keeping with the attacks on its Newtonian underpinning, Colander and Rothschild argue:

> “the self-correcting ‘stability’ vision cultivated by economic pedagogy is problematic in several respects. First and foremost, it is simply wrong: stability is not the norm in complex systems” (Colander and Rothschild 2010 p.286).

Critiques of the equilibrium assumption predate **ABM**, of course. Holub made the case that any new framework would need to declare a ‘year zero’:

> “The long tradition of equilibrium thinking in economics has lead to an unrivalled, consistent structure of thought... an anti-equilibrium attempt cannot build further on this structure, not even on the ruins of equilibrium theory (should it succeed in toppling this construction), it must on the contrary seek a new site for its own, a new thought structure, in other words, a new central idea” (Holub 1977 p.395).

Many seem to believe **ABM** is exactly this new central idea. The concept of the ‘Walrasian auctioneer’, often a target of anti-equilibrium critics, is a good example of this in practice. It is described as an actual mechanism, though in Walras’ own work it is used only as a thought experiment to explain the equilibrium outcome (Blaug 1997 p.555-6). Many **ABM** theorists take issue with a centralised adjudicator capable of mediating a process of ‘groping’ towards a set of market-clearing prices, and knowing the moment when they are all correct[3]. The auctioneer concept is a good summary of everything **ACE** opposes as implausible. Once such convenient (but, it is argued, impossible) assumptions are shown to be unjustifiable, a Pandora’s box is opened:

> “The modeller must now come to grips with challenging issues such as asymmetric information, strategic interaction, expectation formation on the basis of limited information, mutual learning, social norms, transaction costs, externalities, market power, predation, collusion, and the possibility of coordination failure.” (Tesfatsion 2006 p.836)

[3]See e.g. Tesfatsion 2006 p.834, where it is argued that “equilibrium values... are determined by market clearing conditions imposed through the Walrasian Auctioneer pricing mechanism; they are not determined by actions of consumers, firms, or any other agency supposed to actually reside within the economy”; see also Vriend (1991); Posada *et al*. 2007.

Each of these ‘challenging issues’ does in fact garner plenty of attention from analytic economics: Stiglitz got his Nobel prize for his work on asymmetric information going back to the seventies (Rothschild and Stiglitz 1976), game theory deals with strategy (and ‘predation and collusion’ are forms of strategy), and Krugman himself discusses expectation and the role of history as a key element of his geographical work (Krugman 1991a). As regards agglomeration economies, the importance of externalities has been discussed in chapter 2.

ACE models often share one vital feature with the equilibrium methods they reject, however: a market clearing point. This often means that auction-style models are used where markets reach a defined end-point via a bidding process - through many interactions, perhaps, but still terminating at the end of trade. An example would be Kirman and Vriend’s fish market model, with an interdependent morning and afternoon’s trading (Kirman and Vriend 2001 pp.467). The question of what happens if markets do not ‘clear’ was asked by Hicks, and Isard understood its implications for trade across space. If all actors are making market decisions on the head of a pin, an auction-style clearing market presents itself as the natural way to think about it. But as soon as actors have heterogenous location, that becomes impossible: ‘clearing market’ conditions do not hold. As Leijonhufvud puts it, quoting Hicks -

> “In a normal, ongoing market, transactors are not all brought together in a single location and at the same time. Without centralisation and synchronisation, the supply-equals-demand condition ‘cannot be used to determine price, in Walras’ or Marshall’s manner’.” (Leijonhufvud 2006 p.1633; see also Ladley and Bullock 2007 p.83-4)

While easier to ignore this problem if geography is not an element of the model, one cannot consider spatially varied trade as ‘one market’ (Isard 1956 p.43). From an **ACE** point of view, Dibble (2006 p.1516) notes these effects of space on standard market-clearing assumptions. Ladley and Bullock also point out (2008 p.296) that market actors may be segregated by space, able only to interact with a given subset of others, and that spatial segregation forms natural networks as actors’ ranges of interaction overlap. Hamill and Gilbert apply this idea ‘in reverse’: they use actors with randomised location in an R^2 space, and a fixed radii around each actor determines their social network connections (Hamill and Gilbert 2009, 2010).

It is the ‘non-clearing’ part of this problem that the thesis model attempts to deal with, rather than the network element. Section 3.5.1 already made the argument that transport networks can be replaced with a Euclidean proxy when discussing distance, though only by acknowledging the severe limitations for discussing change this imposes. Ladley and Bullock note Wilhite making a related point: that if network change is slow compared to activity on the network, its structure becomes more important, not less (Ladley and Bullock 2008 p.299 referring to Wilhite 2006). This thesis cannot answer this issue fully as it works within the Euclidean assumptions from section 3.5.1, where it was argued that the success of Euclidean stand-ins is dependent on both the spatial and temporal scale under examination, as well as whether network dynamics are part of the research question. So the model cannot ask, as Epstein does, how "the endogenous connectivity - the topology - of a social network affects its performance as a distributed computational device, one that... computes price equilibria, or converges to (computes) social norms, or converges to spatial settlement patterns such as cities?" (Epstein 2006 p.17-18).

4.3.2 Spatial versus economic agents

A Venn diagram of **ACE** and spatial **ABM** would perhaps reveal just as little overlap as between **ACE** and **GE**. Torrens' recent review of '**ABM** in the Spatial Sciences' (2010) illustrates the point. One economic example picked up on is not actually about spatial modelling: Torrens cites Farmer and Foley's Nature piece who argue that the economy is complex, **ABM** can 'do' complexity where equilibrium models cannot, representing something closer to reality; therefore, **ABM** is better:

> "Agent-based models potentially present a way to model the financial economy as a complex system, as Keynes attempted to do, while taking human adaptation and learning into account, as Lucas advocated. Such models allow for the creation of a kind of virtual universe, in which many players can act in complex - and realistic - ways" (Farmer and Foley 2009 p.685-6).

They suggest an effort akin to general climate modelling for the economy is in order, coupling the various social science disciplines in the same way climate models might couple ocean and atmosphere feedback. Their emphasis on building a 'virtual universe' is also present in many of the spatial models Torrens surveys (this idea is the subject of section 4.4.4). The cellular automata models surveyed by Torrens in particular have this quality. Crowd and 'swarm' models loom large, leveraging the interactive properties of agents (a point noted by O'Sullivan and Haklay, 2000 p.1411).

In terms of spatial *economic* models, Torrens cites his own work (among others) - in one example, a technically rich implementation of a combined cellular-automata agent hybrid model (Torrens and Benenson 2005). There appears to be no driving research agenda, however, beyond proving the technical feasibility of doing so, though the ability of disaggregated models to produce 'emergence and self-organisation' is mentioned (p.396).

Agent models using explicitly spatial economic ideas that would seem familiar to a **GE** theorist are few. Sasaki and Box use **ABM** to attempt to 'verify' classic spatial analytics like the von Thunen model (Sasaki and Box 2003) by showing how the spatial equilibrium result can come about through individual-level action but, aside from referring to von Thunen himself, it sticks to the **ABM** literature. The work of Christopher Fowler stands out as a theoretically grounded agent model with the explicit aim of examining the **core model**; this is discussed in-depth in the next section.

4.3.3 Fowler

Christopher Fowler has attempted to recreate the **core model** in agent form. This section discusses the two papers where he does this. In some key aspects, his goal is the same as this thesis. Specifically, he keeps the focus on the simple interactions:

> “the effects of economies of scale, the role of preferences for variety in driving trade between regions, and the push/pull relationship between these two factors in shaping patterns of agglomeration and dispersal in economic activity based on transportation costs” (Fowler 2011 p.2).

The title of his first paper - ‘Taking **GE** out of Equilibrium’ - indicates what Fowler initially thought an **ABM** approach should be good at: “the inability of the deductive models to describe the movement of a system between equilibria represents a major drawback of these models” (Fowler 2007 p.267). An agent approach, he argues, is a good method for achieving this. In his first attempt, he set himself a specific goal: to “re-create as exactly as possible the relationships in the economic [core] model” (ibid p.266) in agent form, meaning that “to the extent possible, the equations used to express the analytic model have been maintained.” (p.272) His second paper made a few extra choices and aimed to “explore the capacity of an economic system to identify a stable Nash Equilibrium without it being enforced by assumption.” (Fowler 2011 p.14)

Fowler’s direction of travel, then, is very much from the analytic core model towards agent modelling. The main conclusion from his first attempt was that -

> “although the equations of the analytic model can be mimicked in a way that is sufficient for a simulation to run, such a simulation cannot be made logically defensible without significantly altering the relationships among workers, firms and cities posited in the analytic model.” (p.282)

In initially binding himself to the assumptions of the analytic core model, Fowler was stuck with one of the more thorny abstractions: there are no firms explicitly defined as separate actors. The equilibrium number of firms - a key feature of of the Dixit-Stiglitz **monopolistic competition** approach, as discussed in section 2.2.2 - is purely a consequence of consumer demand combined with **love of variety**. As he puts it, “firms appear and disappear in cities based on the full employment of the workforce even to the point of following the lead of workers who move from place to place in order to benefit from increased real wages.” (Fowler 2011 p.2)

Fowler’s first pass at an agent version of the core model, then, lacking any method for linking firm and consumer behaviour even implicitly, could not produce equilibrium firm levels. Fowler notes that “as a result, the model fails to move towards any of the equilibrium conditions predicted by its analytic counterpart.” (p.266) As he notes, the full ‘general’ conditions present a fairly daunting proposition for the agent modeller:

> "the amount a firm can offer in wages is dependent on the amount it can produce and the price it receives for its goods. These quantities are affected, in turn, by consumer's wages, which depend on which firm employs them (and whether or not they have found employment at all.)" (Fowler 2007 p.277)

In this situation, how can actors, unable to coordinate through any central mechanism, produce stable spatial economic outcomes, especially given the extra new problems of an 'absence of rational expectations' (p.275) stemming from actors' intrinsic inability to predict the actions of others? (This 'true uncertainty' is discussed in section 5.6.5.) He also points out the uncertainty that firms face: they need "some mechanism with which they could predict an appropriate production level for themselves and estimate the levels chosen by their competition in an environment where sufficient labour supply is not guaranteed. The delicate balance among the equations of the analytic model does not allow for the error and uncertainty that are necessary parts of this sort of prediction, and so a new set of relationships needs to be specified" (p.282).

In a statement he later appeared to regret[4], Fowler goes on to say, "as complex as this specification sounds, it is actually relatively simple in an agent-based framework where bounded rationality, learning by doing and other types of decision-making all have substantial supporting bodies of literature" (ibid p.283).

There are two particular points to pick up on. First, to return to the need for explicit 'firm' actors: it could be argued that the original **core model** *did* have firms - as part of an argument built on the **monopolistic competition** model. The fact that it did not have distinct firm *agents* is what Fowler is criticising: he is arguing for what the next section calls a closer 'descriptive mapping' - and, perhaps, that it is self-evident that any model without distinct 'firm' objects is invalid. Certainly, it appears to be what he sees as the most unjustifiable simplification: "geographical economics does itself a disservice by ignoring the labour market dynamics that are arguably the most important set of relationships driving the movement of labour and capital in the real world" (Fowler 2011 p.7).

The second point is the flipside of this: of course, Fowler must make his own simplifying assumptions. For example, he uses what could be described as a 'localisation externality' simplification in the later paper's model. Three actions are taken: workers calculate their individual utility, an average utility level for each region is then worked out, and finally one randomly selected worker compares their individual level with the average between regions before deciding where will give them the better utility. The externality here is the regional average: each worker is allowed to 'know' this value for all cities. This a necessary simplification for his purposes, but nevertheless not a dynamic that emerges from agent interaction.

[4] "Economists are largely willing to leave the exploration of the labour market to other models. Given the complexity of the model necessary to replace this assumption, this researcher, at least, has grown increasingly sympathetic to the allure of such a potent simplifying assumption" (Fowler 2011 p.7).

Again, then, how does one tell the good simplifications from the bad? Which merely (in Hayek's phrase) "assume the problem away and... disregard everything that is important and significant in the real world" (Hayek 1945 p.530; see also Miller and Page 2007 p.87). The next section deals with this question.

4.4 Mapping the model

4.4.1 'Descriptive' versus 'functional' mapping

Miller and Page (2007 p.36) argue that one can construct a 'homomorphism' between model and reality, an exact equivalence between identified real-world and model structures, but that too broad a homomorphism would be "at the cost of lowering the model's resolution and value" (p.40). As with arguments that models should aim for isomorphism[5], these terms have particular mathematical meanings that encourage the idea that object structures should very explicitly map onto reality, and be transformable in the same way. They are too strict: the looser idea of 'mapping' that Miller and Page start with can be a more useful way to think about building models.

As Scott says, models and maps share a common purpose, both being "designed to summarise precisely those aspects of a complex world that are of immediate interest to the map maker and to ignore the rest" (Scott 1998 p.87). Baumol and Blinder conclude that modelling means choosing the best map - and that will depend on the purpose (Baumol and Blinder 2005 p.12). Miller and Page point out it is intuitively obvious why a more realistic map may not be a better one (Miller and Page 2007 pp.36). Krugman takes this idea further: in the earliest Colonial explorations, map data consisted of verbal reports - sometimes apocryphal, spatially inaccurate, but still very useful ('six days south of the end of the desert you encounter a vast river flowing from east to west'.) However, as more formal mapping took place, "the improvement in the art of mapmaking raised the standard for what was considered valid data". For a time, much useful information was lost (Krugman 1993). In the end, more accurate maps resulted - but in the transition to formalism and rigor, descriptive information gained may temporarily be lost. This is a map version of the streetlight effect: "the methodology of economics creates blind spots. We just don't see what we can't formalise" (Krugman 2008).

One of the supposed strengths of **ABM** is precisely that it promises to combine both descriptive accuracy and formalism, thus avoiding the streetlight effect. 'Descriptive mapping' can appear a natural extension of the structure of **OOP**, which makes defining a close link between object structures and real-world counterparts an obvious thing to attempt. This is true both for objects themselves and the structural relations that **OOP** uses. For example, Tesfatsion uses **OOP**'s distinctions between public, private and protected methods to define

[5] e.g. "Isomorphism is a relation between mathematical structures. If there is a function that maps each element of one structure onto each element of another the structures are isomorphic" (Downes 1992 p.147). The term is used by, among others, Epstein to describe model-to-reality mappings (Epstein 2006 p.24-5).

private behaviours and to allow agents to "communicate with each other through their public and protected methods" (Tesfatsion 2006 p.837).

This section critiques this kind of 'descriptive mapping' by looking at the reasoning of **ABM** theorists and more traditional economic thinkers, in particular Milton Friedman. The argument developed is that while **ABM**'s flexibility makes it appear feasible - and even desirable - to attempt 'more realistic' mappings, the focus should be more on 'functional mapping', where the function is closely tied to the purpose of the model.

Wooldridge notes from an early commenter a "tendency to think of objects as 'actors' and endow them with human-like intentions and abilities." (Inc. 1993 p.7 in Wooldridge 2009 p.28; see also Franklin and Graesser 1996). As regards structure, Tesfatsion argues that -

> "encapsulation into agents is done in an attempt to achieve a more transparent and realistic representation of real-world systems involving multiple distributed entities with limited information and computational capabilities." (ibid p.838)

It is an entirely understandable and intuitive belief: the closer the match between code and real-world - the more 'realistic' - the better. But in what way should real-world systems be represented? It is obviously true that any effective model must in some way represent the dynamic it wants to examine; what Craik (1967 p.51) calls a 'relation structure' (see Stafford 2009 p.3) and Suarez *et al.* (2003 p.225) a 'mapping of source to target'. But to what extent must this representation actually map directly? How important is 'realism?' The goal is, as Hoover puts it, that -

> " the idealised model capture the essence of the causal structure or underlying mechanism at work... Models are not, of their nature, cleanly idealised; they must involve particular properties, whose only function is to make them operable or realisable in a manipulable form." (Hoover 2010 p.346)

Two examples of actual *physical* models illustrate the point graphically. Craik argues that models "need not resemble the real object pictorially; Kelvin's tide-predictor, which consists of a number of pulleys on levers, does not resemble a tide in appearance, but it works in the same way in certain essential respects" (Craik 1967 p.51). MONIAC, the Newlyn-Phillips hydraulic model of the economy, makes the same point. It was built mainly as an educational tool, using water to represent money-flows and various levers and wheels to control flows. Some argue it actually influenced some Keynesians (see Wood 1994 p.249). There were counter-arguments about the 'misleading reduction of economics to hydraulics' (Shackle 1983 p.189 in Wood 1994 p.249). Newyln himself was well aware of this limitation: "once the model has served its purpose... the student will need to return to the literature for the complications and refinements - hydraulics is no substitute for economics" (Newlyn 1950 p.119).

This is a perfect illustration for the upcoming section: one needs to know the purpose of the model before attacking it as misleading. The next section starts to unpick this issue by examining the place of both simplicity and realism in models.

4.4.2 What role for simplicity?

The underlying motivation for descriptive mapping is, perhaps, that a perceived closer match to reality is its own 'validation'. Conversely, this offers an easy line of attack for anyone unhappy with particular simplifying assumptions: they are unrealistic. But how can perceived lack of realism be used to judge a model, given that all models are simpler than reality? How to distinguish good from bad simplifications?

'Occam's Razor' is the idea that, given two theories that may explain some phenomenon, the simpler one is likely to be the better explanation. The centrality of complexity for **ABM** has tarnished the idea of simplicity, though the two should not really be opposed to each other. A particularly severe example of this sees **ABM** theorist Bruce Edmonds attacking Occam's Razor because 'simplicity is not truth-indicative' (Edmonds 2007) He is very blunt about what simplicity should mean to modelling:

> "if I am right, model selection 'for the sake of simplicity' is either: simply laziness; is really due to pragmatic reasons such as cost or the limitations of the modeller; or is really a relabelling of more sound reasons due to special circumstances or limited data. Thus appeals to it should be recognised as either spurious, dishonest or unclear and hence be abandoned." (Edmonds 2007 p.78)

There are grounds for simplicity that go beyond laziness or dishonesty on the part of the modeller, however. Occam's Razor is not an 'appeal' to simplicity, and Edmonds is right: simplicity by itself, is *not* truth-indicative at all. Occam's Razor is a shortcut for finding successful theoretical needles in among the haystack of competing ideas. Friedman nicely outlines the role it has played in the physical sciences: "the theorist starts with some set of observed and related facts, as full and comprehensive as possible" (Friedman 1953b p.282-3). There are, he argues, then an infinite number of theories consistent with the facts. Some 'arbitrary' method is needed to choose between them - such as Occam's Razor. One could just as arbitrarily choose the more complex theory, but it so happens that simple explanations have done better in the physical sciences.

Plenty of the time, however, simple assumptions do not get tested in any meaningful way. These can sometimes be seen in the wild, appearing as 'heroic assumptions'[6] in the literature. An heroic assumption is characterised by two things: being highly unrealistic, but opening up new avenues for analysis. As such, they have always been an easy target for critics, and **ABM** theorists have certainly taken aim at them.

Is there any way of telling what sort of assumption is valid? What distinguishes 'heroic' from 'useful' from 'silly'? The next section looks at this question.

[6] One reviewer of the English translation of Weber's 'Theory of the Location of Industries' (Weber 1909b) in 1930 warns, "in approaching this book the reader must be prepared to meet a very abstract treatment and some very heroic assumptions" (Fetter 1930 p.233). There is a more recent example from Brakman *et al.*, who see the same type of heroism underpinning the success of the **Dixit Stiglitz model** (Brakman *et al*. 2009 p.93).

4.4.3 What role for realism?

Moss and Edmonds' paper on 'good social science' argues that:

> "The essential feature of software agents devised for purposes of social simulation is that they should be validated as good descriptions of the behaviour and social interaction of real individuals or collections of individuals" (Moss and Edmonds 2005 p.10).

They want a social science that "coheres with directly observable evidence in as many ways as possible" (ibid p.5). In seeking this coherence, they say that "evidence and observation have priority over theory" and "when evidence and theory disagree the theory is changed" (ibid p.4). Their goal here is to firmly fix the causal arrow from reality to theory and reject any approach that points in the other direction. As they put it:

> "There are many such cases in the natural sciences where observation and experimentation lead to conceptualisation. We know of no such cases in the core of mainstream economics or sociology, where the conceptualisation has tended to come first" (Moss and Edmonds 2005 p.10).

In this view, valid theory must flow from reality by a process of accreted, validated induction. Theirs is only a subtle tilt in emphasis away from theory-building, but its impact on what counts as 'valid' agent modelling is huge. It is suggested that Einstein's theories, built up from Maxwell's and a number of other theorists, embody the approach they espouse, being "driven by experiment and observation of natural phenomena" (ibid p.3). However, Einstein's view of the matter seems to have been rather different:

> "Physics constitute a logical system of thought which is in a state of evolution, whose basis cannot be distilled, as it were, from experience by an inductive method, but can only be arrived at by free invention. The justification (truth content) of the system rests in the verification of the derived propositions by sense experiences. The skeptic will say: 'it may well be true that this system of equations is reasonable from a logical standpoint. But it does not prove that it corresponds to nature'. You are right, dear skeptic. Experience alone can decide on truth." (Quoted in Kaldor 1972 p.1239.)

The point here is not that 'free invention' has free reign - 'experience alone' still determines the truth-content of theory - but that reality does not automatically supply the descriptive elements of theory.

The search for an understanding of gas theory (discussed below in depth) shows much the same role for 'free invention' being thrown against the wall of reality. The many scientists involved went down different experimental routes, depending on prior assumptions about the nature of matter going back to Aristotle, and especially about atomism and whether a vacuum was possible. As Webster notes, for instance, Boyle got ideas about the elasticity of air from Descartes, who suggested “air was analogous to a pile of wool fleeces” (Webster 1965 p.445). Ultimately, reality was the arbiter of which theories were corroborated - and notice that reality provided the metaphor in this case - but induction never had the unique priority Moss and Edmonds want to give it.

Three issues flow from this. The first is whether social modelling *needs* to meet the same criteria as physical modelling. The second, related, issue is what it should take to reject any particular theory as beyond the social-scientific pale. Any conclusive answer is beyond this thesis, but the above should at least give rather more slack to the use of Einsteinian ‘free invention’ in model experimentation. Third is that, as regards realism, a straightforward ‘coherence’ test of the sort quoted above would appear to be problematic.

In economic models, probably the most frequent subject of scorn is the idea that people are *homo economicus*: infinitely rational utility-maximising machines. This serves as a good subject for thinking through the role of realism. Sociological critiques of *homo economicus* abound: Hollis (1975; 1994 pp.52) takes on ’rational economic man’, and for Granovetter and Swedberg, the starting point of economic sociology is precisely that “while interests are central to any explanation of economic activities, a purely interest-driven model is unacceptably distorted” (Granovetter and Swedberg 2001 p.9).

As mentioned above, a model is by definition simpler than the system it aims to represent. So what counts - in Granovetter and Swedberg’s terms - as an acceptable distortion, and what is unacceptable? Moss and Edmonds’ take is again a good representation of the starting point for many agent modellers. Talking about financial prediction (or lack of), they attack utility-based models:

> “The standard, naïve response... follows Friedman’s classic claim that the descriptive accuracy of assumptions is irrelevant and all that counts is predictive accuracy... its *ceteris paribus* conditions fly in the face of common observation, common sense and experimental evidence.” (Moss and Edmonds 2005 p.9)

Milton Friedman’s argument - called the ‘F-twist’ by Samuelson (see Blaug 1992 pp.91 for an overview) - is indeed a classic critique of the ‘basic confusion between descriptive accuracy and analytical relevance’. Friedman says that “a theory cannot be tested by the ‘realism’ of its ‘assumptions’ ” (Friedman 1953a p.23) Note, however, that he is not quite arguing assumptions are irrelevant. This idea has become something of a caricature - to the point where Moss and Edmonds can claim the argument is now ‘naïve’. Friedman makes an interesting case, however, and it is a very useful way into picking apart how models should map to real world.

Friedman spends some time criticising “necessarily unsuccessful attempts to construct theories on the basis of categories intended to be fully descriptive” (ibid p.34). A model may succeed in ‘descriptive accuracy’, but what about analytical relevance? It is true that Friedman focuses on prediction as the ultimate arbiter:

> "Complete 'realism' is clearly unattainable, and the question whether a theory is realistic 'enough' can be settled only by seeing whether it yields predictions that are good enough for the purpose in hand or that are better than predictions from alternative theories." (Friedman 1953a p.41)

A vital question is then: what is meant exactly by prediction? This is dealt with in section 4.4.5. Before that, however, some context for Friedman's thinking. His argument is built on keeping clear blue water between a self-contained theoretical structure (like Einstein's take on physics as a 'logical system of thought') and the problems it is used to analyse. For Friedman, theory in and of itself is nothing more than a tautologous filing system that, while internally consistent, by itself has no 'substantive content' (Friedman 1953a p.7):

> "The objective is to construct a language that will be most fruitful in both clarifying thought and facilitating the discovery of substantive propositions." (Friedman 1962 p.8)

Internal to any particular system, it does not make sense to single out any one element as an unrealistic assumption. They cannot, by themselves, be used to accept or reject a theory, because 'everything depends on the problem' (Friedman 1953a p.7). That includes whether or not something is an assumption: Friedman concludes, much as Blaug does, that "the logical distinction between 'assumptions' and 'implications' disappears in a perfectly axiomatized theory" (Blaug 1992 p.143). The two are distinguished from each other only by the particular question under examination. This does not mean, however, that Friedman thinks 'assumptions are irrelevant'. As he says -

> "if this were all there is to it, it would be hard to explain the extensive use of the concept and the strong tendency that we all have to speak of the assumptions of a theory and to compare the assumptions of alternative theories. There is too much smoke for there to be no fire." (Friedman 1953a p.41)

What counts as a 'crucial' assumption will depend on the problem at hand, and is something that Friedman thinks is beyond the scope of any simple methodology to determine (ibid p.25). There are, in fact, situations where assumptions "can be used to get some indirect evidence on the acceptability of the hypothesis in so far as the assumptions can themselves be regarded as implications of the hypothesis" (ibid p.28). This will, of course, depend on the hypothesis being proposed.

In a simple thought experiment, he suggests that one can state 'leaves seek to maximise the sunlight they receive'. This is, of course, an egregious simplification of the processes a tree goes through to achieve that maximisation, but it is nevertheless "more compact and at the same time no less comprehensive" than a list of particular rules would be (Friedman 1953a p.24). Friedman goes on to a human thought experiment: how one would go about modelling billiard players? A successful model - that is, one able to make good predictions of game outcomes, on average - might assume that they "knew the complicated mathematical formulas that would give the optimum directions of travel, could estimate accurately by eye the angles, etc... could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas" (ibid. p.21). Conlisk asks (discussing life cycle decisions):

> "But what of a beginner taking the first shot, in poor light, on a badly warped and randomly moving table, with assorted friends and relatives guiding the cue stick?" (Conlisk 1996 p.684)

Then the assumptions would be poor ones for this situation: the model would not work. A relevant thought experiment of Friedman's in this case goes as follows. The equation $s = 1/2gt^2$ is a good representation of the way bodies fall under gravity (where s is distance, t time and g a gravity constant). An assumption is that this happens in a vacuum: air resistance is left out. This leads to several examples where the formula fails: for feathers, or for objects dropped from thirty thousand feet. The point is -

> "... under a wide range of circumstances, bodies that fall in the atmosphere behave *as if* they were falling in a vacuum. In the language so common in economics this would be rapidly translated into: the formula assumes a vacuum. Yet it clearly does no such thing." (Friedman 1953a p.18)

There exist only certain conditions where this simple model works, but that does not nullify its use in those conditions, and requires understanding the sort of error one might expect.

In sum, Friedman's argument is run throughout with a strong vein of 'it depends on the problem' and 'nothing is set in stone'. Focusing purely on isolated assumptions makes it easy to dismiss (for example) *homo economicus* as obviously wrong, but this is an error, stemming from a misunderstanding of the type of model-building Friedman argues to be useful. Later, section 4.5.1 examines the issue of utility in more depth, and looks at how the difference between assumptions and implications depends on what is being asked. Next, section 4.4.4 looks at two model styles - 'virtual worlds' and 'engines of analysis' - and argues they closely resemble the distinction between descriptive and functional mapping.

4.4.4 Virtual worlds versus engines of analysis

Di Paulo *et al.* contrast two positions on what simulations should be: "maximally faithful replicas" versus "thought experiments: unrealistic fantasies which nevertheless shed light on our theories of reality" (Di Paolo *et al*. 2000 p.4). This argument in **ABM** has a strong parallel in economic comparisons between two nineteenth century theorists, Walras and Marshall. Friedman was a defender of Marshall's approach, saying consistently that he "took the world as it is;

he sought to construct an 'engine' to analyse it, not a photographic reproduction of it" (Friedman 1953a p.35). Walras, in comparison, built what Blaug calls a "a peculiar vision of a sort of 'realistic utopia' " (Blaug 1997 p.569). He was the first theorist to construct a full 'general equilibrium' model, in an attempt to work out how all elements of the economy managed to interact to produce an apparently stable outcome. (His use of the concept of tatônnement or 'groping' towards clearing prices has already been mentioned in section 4.3.1.) This approach became the foundation for arguments that such equilibria were also optimal. As Blaug notes, over time Walras displayed "an increasing tendency... to fit the world to the model rather than the model to the world" (Blaug 1997 p.569) - something carried on by further refinements to general equilibrium over time. Cast into modern language, Walras could be said to have created a 'virtual world'.

A model's status as 'virtual world' or 'engine' is not an intrinsic property of the model itself, though via an accretion of theorists' choices, one or the other can become set in. As Friedman put it -

> "... by slow and gradual steps, the role assigned to economic theory has altered in the course of time until today we assign a substantially different role to theory than Marshall did. We curtsy to Marshall, but we walk with Walras." (Friedman 1953a p.89)

Building virtual worlds is sometimes an entirely reasonable goal. Weather models are an obvious example: the goal of meteorology is precisely to create as close a mapping as possible between virtual and real, where the virtual model can be played out faster than reality to make predictions about the future. But it is a mistake to think this is always a laudable, or even feasible, goal for human systems.

ABM is perhaps more structurally suited to virtual world building, insofar as it can be used to make simulacra that can appear to be worlds *in silico*, but are these really different from the sort of virtual world Walras made? In **Artificial Intelligence (AI)** research, the 'strong versus weak a-life' argument sees some saying virtual worlds can indeed be considered as more than just 'simulations of living systems': actually 'realisations of living systems'. Miller wonders whether, if the strong case were true, "we would have to add a sixth kingdom of life to the current five... and databases of biological phylogenies would have to be updated every time a new Ph.D thesis in a-life was written" (Miller 1995 p.21). It is hard to imagine this argument even arising in the same form for purely equation-based models - but **ABM** does not represent quite the break from purely mathematical methods that some claim.

While computers are iterative by default (the system clock imposes a discrete structure), this does not necessarily mean that agent models must be discrete, or indeed that they must be algorithmic: approximations to continuous and algebraic[7] functions are possible. **ABM**, how-

[7]Symbolic computing is capable of producing good analytic results: for example, where iterative approaches to integration like Runge Kutta functions can only approximate, symbolic computing can generally do as well as analytics in producing exact results where they exist.

ever, has tended to integrate discrete time-steps into its approach, and methods are employed to design agent models with appropriate timing.

It is true that **ABM** is algorithmic rather than equation-based[8]. While equation-based elements may be present (as they are in the thesis models) they are used by agents as part of their discrete decision-making. In a comparison of equation-based versus **ABM** methods for supply-chain management, Van Dyke Parunak *et al.* highlight the fundamental differences. Equation-based methods using differential equations start with observables; in contrast, the **ABM** modeller -

> " - begins by representing the behaviours of each individual, then turns them loose to interact. Direct relationships among the observables are an output of the process, not its input." (Van Dyke Parunak *et al*. 1998 p.10)

Again, this sounds like a virtual world: the 'direct relationships among observables' can only be examined through a pseudo-empirical observation of the model as it plays out. In reality, there is less difference between equation-based and algorithmic approaches. Krugman's own work illustrates one aspect of this: it took him many years to work out what the full details and consequences of his modelling approach were:

> "Why exactly I spent a decade between showing how the interaction of transport costs and **increasing returns** at the level of the plant could lead to the 'home market effect' and realizing that the techniques developed there led naturally to simple models of regional divergence remains a mystery to me." (Krugman 1999)

Part of the answer is that equation-based modelling can be just as opaque as **ABM**, and that working out the implications of the interactions is equally challenging. As Di Paulo *et al.* note: "in general, and qualified claims of the superiority of one style of modelling over another are not compelling." (Di Paolo *et al*. 2000 p.3) Ultimately, however, the difference between virtual worlds and engines of analysis does not come down to their structure, but to the way they are interpreted by modellers. The one defining characteristic of 'virtual worlds' is that they claim to explain something about this world purely through their own existence. It is taken to be self-evident that virtual and real world dynamics are identical in some key aspects. Arguments for model 'realism' lend weight to this interpretation: virtual and real world dynamics are more alike.

A useful way to put some meat on the bones of this discussion is to examine an especially popular take on agent modelling, the idea of 'generative' social science. Before that, the next section outlines three types of prediction: these will be used in the following sections to help shed light on the difference between 'engines of analysis' and 'virtual worlds'.

[8]Epstein argues this distinction is logically false; see Epstein 2006 p.27. Epstein argues all agent models must logically have an equation-based counterpart. In practice, however, this logical equivalence is less important than the way algorithmic and **OOP** approaches affect the choices modellers can make.

4.4.5 Three types of prediction

Three types of prediction can be defined. The first is forecasting: making a claim about something that will occur at a future point. The second is closely related, and uses precisely the same methods: in modern terms, it is 'back-casting'. As Friedman puts it, prediction -

> " - need not be forecasts of future events; they may be about phenomena that have occurred but observations on which have not yet been made or are not known to the person making the prediction." (Friedman 1953a p.9)

The third is what Betz calls 'ontological prediction' (Betz 2006). This is a claim about a phenomenon that has always existed (that has occurred, is occuring now, and will in the future) but was not known about or looked for. Einstein's work on the relationship of gravity and space, for example, describes something that always happened. The theory led Eddington to carry out his photographic test confirming that light 'bent' near the sun as it followed a straight line in gravity-warped space. (See Almassi 2009 for a recent discussion of this example.) Betz also notes the prediction of Neptune's existence from Newton's laws, which showed other planetary bodies' orbits to be incorrect without another planet to explain it.

The discovery of tectonic plates is a more down-to-Earth example, and one used by Epstein (2008). This example highlights the clear relation between the different types of prediction: understanding tectonic plates places a boundary around possible future outcomes that were unknown before. Earthquakes will mostly happen on or near lines of tectonic plates, and understanding of tsunamis can also be built on that knowledge (Thompson and Derr 2009, who critique Epstein's approach).

'Ontological predictions', then, can lead to discoveries, if the predictions are good, as well as provide methods for placing bounds around forecasts. Putting it in 'streetlight effect' terms, they shine a light into previously dark areas. Krugman's use of the **monopolistic competition** model, it could be argued, is an economic example of this. If a theory can facilitate the 'discovery of substantive propositions', these may take the form of ontological predictions.

4.4.6 Applied versus theory

Webber, writing in the eighties, identified some confusion from mixing up the purpose of models. He distinguishes between applied and theoretic (or 'scientific') models (Webber 1984). Theoretic models are "designed to increase our understanding... and to integrate theoretical and empirical research." Applied models, in contrast, are tools used in the day-to-day running of organisations and institutions, and in longer-term forecasting and planning. They -

> " - regard the form of society as given and are produced to make that society operate more profitably; since scientific research examines the conditions under which the society operates and the manner in which it is changing, such research ought not to regard the form of society as constant." (p.149)

Webber argues this comes to down to ‘a difference between the need to understand and the need to prescribe’ (p.151). Traditionally in economics, this is the distinction between positive and normative analyses (Friedman 1953a).

Questions from an applied or theoretic perspective are completely different. Transport economics is a good example of an applied approach; as Button says, it takes a “given land-use pattern and looks at methods of providing efficient transport services within this constraint” (Button 2010 p.51). He also notes the same ‘applied’ ethos, taking existing structures as given in the short-term, characterises supply chain analysis and operations research (ibid p.327). Compare this to Haggett’s search for a ‘comprehensive model of route development’ discussed in section 3.5.1: a transport economist may need to consider, for example, how the risk from fuel cost changes could be mitigated through infrastructure design, but that is quite different to asking, how would the spatial economy as a whole be affected by these cost changes in the long term?

ABM, as with any form of social modelling, has feet in both camps, but the boundaries between them are more blurred than were the urban models Webber examined. Some do identify a similar distinction; for instance, Brown and Xie (2006 p.941) talk of two ‘modes’: instrumental and representational. Instrumental agents are carrying out some real-world task, whether collecting information on the internet or optimising container port flows, whereas representational agents are doing just that: representing some object. But these do not seem to count as either ‘applied’ or ‘theoretic’: either can be used in applied settings, and either could have a role in operations-related work. Optimisation of container ports is a typical case. The gap between models of good flow and reality is shortened; the model not only represents the target system, it is uploaded directly the port’s computer systems, in effect making the model reality (Steenken *et al*. 2004). If models are maps, these maps are reflexive: it not only describes, but is used to act upon the world (Scott 1998 p.87).

The need to model systems that require regulation is not new; Conant and Ashby (1970), for instance made an early argument for this. One might reply that agent-based operations models are qualitatively different, because cognitively more sophisticated. Regardless of whether this is true (and some cyberneticians would probably disagree; see e.g. Beer 1994), there are other reasons why the distinction between applied and theoretic models has become blurred in **ABM** that tie back to the type of mapping a model is meant to achieve.

This mirroring of real and virtual systems is underpinned by arguments that, as Epstein says, “certain social systems, such as trade networks, are essentially computational architectures.” (Epstein 2006 p.16) Epstein can thus develop a theory that computations carried out *in silico* can be considered as a version of the physical computation under study. Similar lines of thought about theoretic agent modelling have created quite a tangle as regards the model-reality relationship. For instance, an argument has developed that such models can actually be “synthetic sources of empirical data” (Di Paolo *et al*. 2000, who criticise this approach). In combination with the idea of ‘emergence’, this has been followed through to its logical conclusion: what Epstein calls ‘generative social science’, in which a model that generates a particular phenomenon can be said to explain it.

4.4.7 Generation versus explanation

Joshua Epstein’s work on ‘growing artificial societies’ (Epstein 1996) first presented his notion that *in silico* emergence could be a form of explanation in and of itself. The motto became, ‘if you didn’t grow it, you didn’t explain it’ (Epstein 2006 p.xii). Epstein calls this ‘generative explanation’ and styles himself has a ’generativist.’ He sees computational models as ‘a new scientific instrument’ (ibid. p.xv) able to generate explanations about the social world almost without reference to physical reality. This stems from the idea that both ‘worlds’ are in some sense producing the same dynamic. As Epstein puts it, “the agent-based approach invites the interpretation of society as a distributed computational device, and in turn the interpretation of social dynamics as a type of computation” (ibid. p.11). Epstein (perhaps accidentally) gives a perfect illustration of the blurring between virtual and real worlds this argument leads to:

> “No-one would fault a ‘theoremless’ laboratory biologist for claiming to understand population dynamics in beetles when he reports a regularity observed over a large number of experiments. But when agent-based modellers show such results - indeed, far more robust ones - there’s a demand for equations and proofs.” (Epstein 2006 p.28)

They might fault them, however, if they claimed their experiments had in fact explained the behaviour of beetles in the wild. More than that, there are plausible reasons for thinking live beetles might be a better model than an artificial life version, even in a laboratory setting. No-one would fault them for using either digital or live models to explore the problem - but this is a much less grandiose goal than explanation. Yet for Epstein, any model ‘sufficient to generate’ a target macroscopic phenomena has succeeded in explaining it. This may be in part hyperbole, since it appears to be contradicted by statements elsewhere (see below), but it is still striking.

Epstein picks on the central dogma of neoclassical economics, discussed in section 4.3.1, as a basic illustration of his argument: demonstrating the existence of equilibria says nothing about the process of reaching it. Epstein says of equilibrium -

> “To the generativist, this is unsatisfactory; to explain a pattern, it does *not* suffice to demonstrate that - under this ensemble of strictures - *if* society is placed in the pattern, no (rational) individual would unilaterally depart. Rather, one must show how the population of boundedly rational (i.e. cognitively plausible) and heterogeneous agents... could actually arrived at the pattern on time scales of interest” (ibid. pxiii).

The generative approach, then, sounds very Hayekian: "it is irrelevant that equilibrium can be computed by an economist external to the system... The entire issue is whether it can be attained - generated - through decentralised local interactions of heterogenous boundedly rational actors" (ibid. p.27).

Schelling's segregation model, discussed by Epstein as a classic model of emergence, illustrates the problem of privileging this kind of 'generation'. As a virtual world, one makes the conclusion Epstein does: it 'explains' segregation, and the model's use stops there, 'brushed under the carpet of emergence' (Di Paolo *et al*. 2000 p.8). As an engine of analysis, it presents a hypothesis: a possible dynamic factor, among others, that may explain something about real-world segregation. Putting aside the hidden assumption - that all social macro phenomena must be the result of a process of interaction between micro-elements - it would need to be put into the context of a fuller theory, that might include (for example) the racial impact of transport infrastructure (e.g. the policies of Robert Moses, see Winner 1980), regeneration policy (Gaffikin and Morrissey 2011) or the effect of estate and letting agents' steering (Phillips and Karn 1992).

A last example gets to the nub of the issue with the generative approach. Epstein considers the 'confusion between explanation and description'. He looks at an example from his 'Sugarscape' models (Epstein 1996) where agents generated a sine-like oscillation of population change over time, and asks, "could you not get that same curve from some low-dimensional differential equation, and if so, why do you need the agent model?" He suggests an equation to describe the model, just a simple, stable sine-based oscillation. He then asks (italics in original) -

> "now, what is the explanatory significance of that descriptively accurate result? It depends on one's criteria for explanation. If we are generativists, the question is: how could the spatially decentralised interactions of heterogenous autonomous agents generate that macroscopic regularity? If that is one's question, then the mere formula $P(t) = A + B \cdot sin(Ct)$ is *devoid of explanatory power despite its descriptive accuracy.* The choice of agents versus equations always hinges on the objectives of the analysis" (Epstein 2006 p.28-9).

Epstein is absolutely right: it does depend on one's criteria for explanation. His example takes the story back to Feynman's quote at the start of the chapter: whether something counts as an explanation depends on what level of explanation is required. The most vital point is that nothing is implied by accepting one particular level of explanation; certainly, it does not mean the theorist has failed to understand that deeper levels of explanation may exist. (The next section examines this issue in detail.)

Epstein appears to be completely in agreement that a model may represent many possible theories about the world, and need testing appropriately, and that the 'generativist' is merely subscribing to one particular 'criteria for explanation'. Yet his central argument is still that 'growing it explains it'. Why this far-reaching claim, rather than perhaps more modestly proposing his models as a way to develop hypotheses? The only way it makes sense is if the model is, in some sense, an empirical world of its own.

The story in this section has been about **ABM**'s tendency toward building of virtual worlds, and treating them as sources of empirical data. As Di Paulo *et al.* say, however, "simply treating non obvious patterns or entities as 'emergent' is not an explanation at all, but rather the statement of a problem" (Di Paolo *et al*. 2000 p.8). Epstein's mistake, then, is to water down the concept of 'explanation' so much that it cannot be distinguished from 'hypothesis'. Epstein seems to tacitly acknowledge the distinction when he argues that, if more than one candidate 'microspecification' is found that could potentially be applied to the problem, "as in any other science, one must do more work, figuring out which of the microspecifications is most tenable empirically," which may involve the need for new experiment or data collection (p.9). It is perhaps an issue of semantics then, but the impact on the outcome of models is not trivial.

4.4.8 What level of explanation?

The theory of the ideal gas law often arises in the literature as a useful thought experiment for examining the difference between 'levels' of explanation, and is particularly well-suited to thinking through **ABM** issues. Flake puts it well:

> "Collections of gas molecules behave in very predictable ways. Knowing only the temperature and pressure of a gas tells you enough about the whole ensemble of molecules that you can effectively ignore what the individual molecules are doing." (Flake 1998 p.134)

So, atoms are 'not like' their macroscopic behaviour might suggest, but this does nothing to change the validity of describing collections of atoms in terms of temperature and pressure. There are theories for understanding the behaviour of gases at both the aggregate and atomic level, and each is useful in the correct context. Heating a gas while keeping the volume constant will lead to an increase in pressure.

The thought experiment reveals a number of different issues in separate areas of economic theory. The first relates to **ACE**'s understanding of complexity. Take Tesfatsion's definition of complexity for example. Two requirements are given: the system is "composed of interacting units" and "exhibits *emergent* properties, that is, properties arising from the interactions of the units that are not properties of the individual units themselves" (Tesfatsion 2006 p.836). By this definition, temperature and pressure exhibit complexity, and yet clearly they can be described using simple known equations. As regards the thesis models, section 6.5.2 has interacting agents producing an emergent result, but this can also be described using summary statistics.

Secondly, as Hoover suggests, many economists are unhappy with having several explanatory levels, each qualitatively different from the other. He argues some believe that 'aggregates are nothing else but summary statistics reflecting individual behaviour'. In comparison -

> "... those who believe that the ideal gas laws reduce to statistical mechanics do not claim that the ideal gas laws should be abandoned for practical purposes." (Hoover 2010 p.331)

Hoover charts the change in economics from a Marshallian emphasis on 'individual *and social* action' (emphasis added) to the micro-economic focus on 'human behaviour as a relationship between ends and scarce means which have alternative uses' (Robbins 1935 p.16 quoted in Hoover 2001 p.108). Many economists, like Hayek, think aggregate properties have no real existence - a 'fallacy of misplaced concreteness'. From this point of view, aggregate or emergent properties cannot be understood in the straightforward way that temperature and pressure can in gases:

> "Hayek thus argues that aggregates exist, but derivatively rather than fundamentally, and that... they do not exist objectively (i.e. unconstituted by the representations of theory." (Hoover 2001 p.108)

The argument is not unique to economics: physicists also tussle over whether certain emergent properties 'are not reducible without loss to the behaviour of the particles that constitute their substance' (ibid p.112). Yet economists often have a more troubled relationship with the connection between explanatory levels. In his 'Self-Organising Economy' (Krugman 1996, p.15), Krugman asks "what constitutes an 'explanation' from the point of view of economists?" The title of Thomas Schelling's 'Micromotives and Macrobehaviour' (1978) is, for Krugman, a compact answer to that question: a good economic explanation shows how micromotives link to macro results. A common argument from **ABM**-sympathetic economists is that, while the goal of linking micromotive and macrobehaviour is laudable, mainstream economists have built only fallacious arguments.

What arguments are these? Returning to Flake, he makes a second point: "notice that the properties of temperature and pressure cannot be attributed to a single gas molecule but only to collections of molecules" (Flake 1998 p.134). Making this sort of attribution is known as the 'fallacy of division'. The reverse - assigning properties to macro-level entities because its components possess them - is a fallacy of composition. In all these cases, the fallacy is not that such claims are *a priori* unjustifiable, but rather in presuming them to be true without evidence to support that presumption. Howitt believes that -

> "these twin fallacies play an even bigger role in a macroeconomist's education than they did a generation ago; the difference is that instead of being taught as pitfalls to be avoided they are now presented as paradigms to be emulated." (Howitt 2006 pp.1069).

A particular target for critics is the use of the '**representative agent**' approach: a single agent, with one utility function, stands in for all agents. Kirman (1992) has made a strong critique of the **representative agent**, claiming it is nothing more than 'pseudo-microfoundations' (p.125). It is, he argues, a necessary lynchpin for claiming that equilibria are unique, and thus comparable. One of his most effective points is that many economic problems can actually be understood much more simply if the economy is treated as many separate agents. As he notes, "erratic individual demand behavior may give very smooth aggregate demand behavior, if individuals are different enough" (p.129). An intuitive example is stability of crop prices over a season's production, as farmers individually look to maximise profit: by selling when they judge they will get the best price, and avoiding over-supply points where prices will be lower, the market is smoothed. More formally, Kirman points out that a 'representative' consumer with non-convex preferences[9] would see demand "jump from one bundle to another at certain prices" (ibid). In contrast, many heterogenous actors with non-convex preferences may well produce a smooth overall market response in aggregate[10].

This fallacy, Conlisk notes, might well be an "ironic misspecification problem" (Conlisk 1996 p.677) that suggests (in Heiner's words) -

> " a reversal of the explanation assumed in standard economics: the factors that standard theory places in the error term are in fact what is producing behavioural regularities, while optimizing will tend to produce sophisticated deviations from these patterns. Hence, the observed regularities that economics has tried to explain on the basis of optimization would disappear if agents could actually maximize" (Heiner 1983 p.586).

In conclusion, there are arguments over just about every element of 'levels' of explanation. What seems to separate them all is a view of the nature of the connection between levels. Some, like Hayek, dismiss macroscopic explanations completely, while one of his critics suggests he fails to analyse the connection between levels adequately: he ''merely invokes the magic words *the price system* without examining its entrails. It is as if correctly sensing the importance of sunlight for life on earth, we were to merely worship the sun rather than study astronomy or photosynthesis" (Desai 1994, p.47).

On the other hand, many economists are, allegedly, "scandalised to discover how cavalier physicists are in making conjectures that lack any fundamental justification" (Ball 2007 p.647 paraphrasing Miller and Page 2007), preferring theories with what they perceive as solid micro-foundations. This is very evident in the way gravity models are treated by economists. Whereas gravity model use in trade cost flows has proved to be very effective (e.g Carrere *et al.* can use a gravity model to conclude that "distance impedes trade by 37% more since 1990 that it did from 1870 to 1969"; Carrre *et al.* 2009 p.6), Fujita *et al.* (2001) can lament the 'limitations of regional science', noting that "the general sense of loose ends left hanging prevented it from

[9]The useful property of convex preferences is that a budget line drawn through them will only ever provide a single unique point of consumption. In contrast, other sorts of 'not well-behaved' preferences may mean demand 'jumping' as income or prices change. See e.g. Varian 2006 p.77.

[10]The model in section 6.7 has oscillatory behaviour: it may be the case that heterogenous agents with non-convex preferences would help to smooth out those market shifts.

becoming a well-integrated part of mainstream economics." (p.33) However, they also point out that it *did* become a 'toolbox for practical analysis' used to guide policy, despite its purported lack of a 'rigorous framework'. Brakman *et al.* are equally happy to reject 'macro' theories like market potential and gravity models on the same basis. These are theories, they argue, that -

> "... try to come to grips with a spatial regularity but that lack a convincing economic-theoretical foundation. In contrast to (neoclassical) economic theory, there is a tendency to merely give a representation using, for example, simple equations, of the regularity without a connection to a model of the underlying individual behaviour by economic agents." (Brakman *et al*. 2009 p.48)

Yet they acknowledge that a gravity equation can accurately capture the drop-off of trade flows between regions; one study of Germany they mention manages an r-squared value of 0.915. This sort of result explains why such models end up in planners' toolboxes.

It also suggests a question: why can't gravity models be just as appropriate a description *at their level* as temperature and pressure are for gases? Wilson, for one, consciously acknowledges the parallels between them (Wilson 2000 p.151). In the development of gas theory, physicists attacked the problem at all levels and, as argued, no approach *a priori* nullified the other. Newton's original theory of gravity illustrates that one level of explanation can be entirely successful - indeed, in this case, is the exemplar of a scientifically robust, generalised theory - while lacking 'microfoundations'. As Newton said: "I have not been able to discover the causes of those properties... and I frame no hypothesis" (quoted in Silver 2000 p.44). The search for gravity's microfoundations carries on to this day at CERN.

Given all that confusion, the next section makes some attempt to see how the connections between levels of explanation might be important for **ABM**.

4.4.9 Connecting levels of explanation

The previous sections have discussed that different levels of explanation exist. But how does one 'get between levels'? This section looks at some examples. Below, the nature of firms' decision-making is discussed. Before that, the focus remains on thinking through how the microscopic level of actor interaction connects to macro outcomes.

O'Sullivan and Haklay are absolutely right to point out the default state of **ABM** as one of 'methodological individualism' (O'Sullivan and Haklay 2000 p.1413): it is concerned almost exclusively with how atomistic actors produce macro-level regularities. This often "unacknowledged assumption" (ibid) is a natural consequence of **OOP** guiding the modeller to individual objects.

There is, clearly, a connection between individual behaviour and aggregate outcome, but how can it be theorised? When building agents, what cognitive resources are required to connect them? The problem can be thought about as two poles: *metis* versus zero-intelligence. The former is Hayek's position. For example, he argues any kind of socialist economy is *a priori* impossible precisely because complex, embedded human knowledge makes up the microscopic components of economic activity: his argument is "not so much that a socialist economy could not transmit the necessary data, but rather that it could not generate it to begin with" (Chamberlain 1998). The knowledge involved, according to this argument, cannot be summarised or modelled.

Scott calls this kind of knowledge *metis*: situated knowledge, of the kind a tug captain has who knows how to pilot through one specific harbour. Metis is a form of knowledge which "represents a wide array of practical skills and acquired intelligence in responding to a constantly changing natural and human environment" (Scott 1998 p.313). The metis argument implies that exchange of information between humans cannot be reduced to 'information' in a computer. As Hodgson says -

> "the information held and transmitted in the form of a symbol is thus embedded in a network of interconnected meanings, related to and produced by social structures. Genetic or computer information does not have this quality; it is at most indexical. In contrast, human information is structured and cultural; it is entwined with institutions." (Hodgson 1996 p.253)

Putting aside that genetic information is obviously entwined with its entire past evolutionary history of interacting in given environments, the point is the same: if the metis argument is right, it implies not that models of human behaviour must fail, but that generating valid macroscopic results from modelling at the actor level would require cognitively sophisticated objects able to develop their own 'metis' - for example, their own particular, contextual understanding, to be summed up in a choice of price.

Relatedly, but not quite the same argument, is that (as Tesfatsion describes Penrose) "there is something fundamentally non-computational about human thought, something that intrinsically prevents the algorithmic representation of human cognitive and social behaviors" (Tesfatsion 2006 p.844). Tesfatsion brings up Franklins' 'first **AI** debate'; as she frames it, the problem is this:

> "in any purely mathematical model, including any **ACE** model in which agents do not have access to 'true' random numbers, the actions of an agent are ultimately determined by the conditions of the agent's world at the time of the agent's conception. A fundamental issue... is whether or not the same holds true for humans" (Tesfatsion 2006 p.844).

The question for the first **AI** debate is, 'can we, or can we not, expect computers to think in the sense that humans do?' (Franklin 1997 p.99). The Penrose view, as Franklin says, is that "not only can computers not experience the things we experience consciously, they can't *do* the things we do consciously" (ibid). Tesfatsion opts for a fudge: "lacking a definitive answer to this question, **ACE** researchers argue more pragmatically that agent-based tools facilitate the modelling of cognitive agents with more realistic social and learning capabilities (hence more autonomy) than one finds in traditional Homo economicus" (Tesfatsion 2006 p.844).

The point that seems to be missing from Tesfatsion's argument is that Penrose could be right and agent modelling still valid. The existence of the simplest supply and demand dynamic shows this: regardless of whether some irreducible nugget of consciousness places human buying decisions beyond the reach of any model's initial conditions, those decisions can still produce aggregate order. If they do, there are definitely valid reasons for disaggregating into 'unrealistic' agents where aggregate regularities exist.

The whole enterprise of microsimulation (e.g. Ballas and Clarke 2001) is based on using those regularities to cycle between macro to synthetically generated micro models and back as a way of leveraging the information in different datasets. The key point is this: there is no 'fallacy' of composition or division taking place, despite the fact that the model actors are being treated as microscopic statistics-processing machines. The unrealism involved - the lack of coherence to their actual internal sophistication or embeddedness - does not *a priori* invalidate the approach.

In some senses, this makes Penrose's argument *less* problematic than metis, to the extent that metis means all actor decisions have massive interdependence. Actors with metis may sit in the middle of the 'complexity curve' between equation-based and statistical predictability (Flake 1998 p.135), and thus summary statistics may be of no use. If predictive power were the goal, attempting what Friedman calls 'photographic representations' in this situation would not be sensible.

Tesfatsion's insistence that, regardless, agent models can manage more cognitively plausible agents suggests the persistent implicit preference for descriptive mapping: human cognitive sophistication does not necessarily require model agents to attempt to mimic it, and doing so does not make them better models. Ross summarises why: "you... have enough in common with economic agents, especially in modern institutional settings, that non-trivial predictions about your individual behaviour can be had by modelling you as if, within temporal and institutional constraints, you were such agents." Ross 2008 p.130)

At the other end of the scale are 'zero intelligence' agents: these can "bid randomly subject only to budget constraints [and] may achieve near perfect market efficiency" (Conlisk 1996 p.675). Their success suggests that, for important classes of problem, metis is of no relevance. Rather, it is the structure that actors must work within that constrains outcomes - so much so that random actions, properly structured, can reach optimal macroscopic outcomes.

As section 4.5.2 discusses shortly, the modelling of firm decision-making has all the same issues as for human beings. Yet, again, collosal simplifications are common. Taking the example of the 'engine' of **GE**, the **monopolistic competition** model, the equilibrium number of firms is completely constrained by the level of demand, as well as its **love of variety** nature. If a disaggregated version were modelled, would there be any difference between allowing firms entry and exit randomly, and giving them random decisions, or building a more sophisticated 'cognitive structure'? Firms' internal and external processes are obviously much more accessible than the human mind: shouldn't this make such simplicifications a clear violation of reality?

The function of the **monopolistic competition** model is to allow a simple analysis of what happens in a market of that sort which, as an added bonus, allows a spatial model to be created. It is not meant to be realistic. So would developing a disaggregated model supersede it? What if that model failed to recreate the **monopolistic competition** model? Would that falsify it? The Hayekian/Epstein-style argument would be, if the dynamic in the **monopolistic competition** model could not be produced through the interaction of agents, it is nothing more than an unacceptable distortion of reality, an assumption too far that claims to know the end process of monopolistic competition with no justification.

Friedman's 'as if' argument is often used in relation to this question. Firm entry and exit is standing in for a complex process of adaptation to economic reality. As Friedman says, "the process of 'natural selection' thus helps to validate the hypothesis - or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgement that it summarises appropriately the condition for survival" (Friedman 1953a p.22). Firms are acting 'as if' evolutionary pressure is selecting them: "they will evolve to the extent that selection processes quickly eliminate poorly administered behaviour" (Heiner 1983 p.586) In a book written specifically as a guide to action in business, Baumol makes the same point in more concrete way:

> "It is at least possible that shear business acumen and experience permit management and other economic units to arrive at decisions which come close to being optimal. Moreover, in business, competition may soon eliminate firms whose decision-making is consistently poor. To the extent that these assertions are valid, optimality analysis should serve as a relatively good predictor of economic behaviour; that is, it should provide a reasonably good explanation of actual economic decisions and activities." (Baumol 1976 p.5)

The fundamental point from the **monopolistic competition** model is that, for a given demand level, only a set amount of firms can survive. The *process* leading to that number is obviously completely glossed over - but for its purposes, that is taken to be a strength, not a weakness.

This is a useful example because it highlights one of the major divides between the physical and social sciences. Baumol wants to say that economists' relationship to operations research is "somewhat analogous to the physicist's relation to the engineer" (Baumol 1976 p.5) But - as the quote at the start of the chapter suggested - no physicist need attempt to educate atoms on how best to follow physical laws. McLuhan's aphorism applies: "we shape our tools and thereafter our tools shape us" (Mcluhan 1964). In mundane modelling terms, the problem is intractable: a model is unlikely to be able to accurately map a reality that includes itself.

To put it in rather more down-to-earth terms, Baumol is conflating applied and theoretic models here. It may be entirely feasible to treat firms in the way the **core model** does for gaining understanding, but arguing, as Baumol does, that such knowledge could actually be used by firms to 'engineer' outcomes seems a much more questionable proposition.

The parallel to ideal gas theory is also instructive in what it says about the *difference* between physical and social modelling. The connection between atomic motion and pressure needs to be robust. Such a solid link between layers of social analysis might seem an appealing goal, but the failure to achieve it would not automatically throw any particular level of explanation into doubt. Indeed, Summers suggests that "attempts to make empirical work take on too many of the trappings of science render it uninformative" (Summers 1991 p.130).

4.5 Production and utility in an agent context

4.5.1 Using utility in an agent model

The idea of utility is a simple way of thinking through how people react to cost changes, and should be considered as just that. As discussed in section 4.4.3, it is often seen to be umbilically linked to what Howitt calls an 'irrational passion for dispassionate rationality', which makes it "easy to dismiss as *ad hoc* or poorly grounded any theory that starts with behavioural rules not explicitly derived from rational foundations" (Howitt 2006 pp.1610) But pinning utility onto rationality is not necessary - any more than the theory of gravity requires planets to love each other.

Section 4.4.3 discussed the question of realism of assumptions. Friedman cites an early critic, Thorstein Veblen, pouring derision on the idea of people as "a lightning calculator of pleasures and pains, who oscillates like a homogenous globule of desire of happiness" (Veblen 1898 p.389 quoted in Friedman 1953a p.30). It was argued that pointing out people are simply not like that is not, by itself, a strong enough reason to reject the whole approach.

The **ABM** literature contains opposing views, though the overall bias towards complexity means utility is ignored more often than attacked. Vriend is a rare theorist, both an agent modeller and a careful critic of utility. As he points out, "economic behaviour simply means that an individual agent chooses (one of) the most advantageous options, given their preferences, in their perceived opportunity set" (Vriend 1994 p.33). Arguments for rational underpinnings are, for Vriend, just "another name for economic behaviour; a question of rhetorics" (Vriend 1996 p.265-6). Feynman's 'levels' argument helps here: this is just working in a framework where something is allowed to be true. As Vriend puts it:

> "Abstracting from an explanation of the individual agent's preferences, and from the mental processes by which he arrives at choices, economics is just a very specific abstraction from reality. Whether these fundamental abstractions are good approximations of reality depends upon the usefulness of the explanatory discourses one can build on it." (Vriend 1996 p.265-6).

This is just reiterating Friedman's point: the theory is essentially tautologous. As Blaug puts it, it does not need a 'hedonistic premise' (Blaug 1997 p.338): there is no need to gain access to people's internal states for utility to be a useful tool for understanding the effect of cost changes. But treating utility theory as a self-consistent tautology is hardly unproblematic. Becker's take on rationality illustrates the issue well:

> "When an apparently profitable opportunity is not exploited, the economic approach does not take refuge in assertions about irrationality... Rather it postulates the existence of costs, monetary or psychic, of taking advantage of these opportunities that eliminate their profitability - costs that may not be easily 'seen' by outside observers" (Becker 1976, p.7).

In this way, any action taken by economic actors becomes their 'revealed preference': if one accepts *a priori* that the action is rational, the action itself must logically be the outcome of a rational choice. This way of thinking helped Becker produce ideas like rational addiction theory, where the focus becomes precisely about 'rationality as explanation'. This leads, unsurprisingly, to incredulous critiques that argue it "raises the question of how they can be taken seriously" (Rogeberg 2004 p.264). (For a survey, see Melberg and Rogeberg 2010.) Becker's argument also seems to make this approach to utility completely immune to empirical testing. As Conlisk puts it: "whatever the truth about the particular case, economic research often seems to work backwards from empirical findings to whatever utility maximisation will work. Where the empirical arrow falls, there we paint the utility bullseye" (Conlisk 1996 p.685).

This is certainly how Becker's approach sounds: a classic case of allowing *ad hoc* modifications (see e.g. Chalmers 1999 pp.75) to keep the theory of rationality through revealed preference coherent. A simple thought experiment should illustrate. If a model of allotment production versus growing food in one's backyard were built, how would utility be used? The factors can be reduced to time and distance. Does an actor stand to gain more utility from travelling to the allotment and putting time in there, or by staying at home and growing, thus eliminating travel costs? (This is a similar approach to optimising time used in the models in this thesis.) For a set of actors who can choose between allotment and home-garden plots of similar size, rationally, it seems they should choose to eliminate travel costs and produce in their own backyard. But what if reality does not conform to this - there is a set of people who choose allotments over their own, perfectly suitable garden. Now one might propose an alteration to the theory, making sure to keep the assumption of rationality, but assuming a previously hidden element: utility is gained from leisure use of the home garden that would be lost by turning it into a vegetable patch. That would need to be outweighed by the disutility of distance before an actor would dig up their begonias.

From a philosophy of science point of view, this makes utility theory - especially of the type grounded in rational choice - beyond the falsificationist pale. It is immune to almost any challenge, since all questions posed by reality can simply be tidied up as a hidden cost or benefit.

However, should utility theory be treated as a scientific theory in this way?

Thinking about how people react to *space* cost changes actually helps give the discussion some concreteness. Transport economics' use of revealed preference is an excellent illustration of the usefulness of the idea in practice. Approaches to utility in the field have used the idea of revealed preference successfully to say useful things about people's choices. Button points out that "the general conclusion about the idea that some overall budget mechanism governs individual travel decisions, however, must be that, to date, the evidence available still leaves many questions unanswered and the theory is still largely unproved" (Button 2010 p.92). Yet, despite this, many directly practicable ideas have emerged through studying people's reaction to space costs. One that stands out is finding revealed preferences for travel time: "if a person chooses to pay $x to save y minutes then he/she is revealing an implicit value of time equal to at least $(x/y) per minute." This has led to an understanding that -

> "savings in walking and waiting times are valued at between two and three times savings in on-vehicle time - parameters that have proved to be remarkably robust over the years." (ibid. p.104)

This is striking: how people value travel time is not simply a function of the time taken. Those willing to commute by car for an hour each way would be much less likely to walk for that length of time. As Button points out, these findings go back to at least 1967 (Quarmby 1967 p.297). One might counter: utility describes but does not explain this. But again, it is clear that the difference between the theory being a description or an explanation comes down to one's particular purpose, and the level of analysis. Certainly, the underlying explanation would require a deeper understanding of the factors affecting mode choice, but it would allow one to develop a predictive theory.

One last issue is worth mentioning. A particularly troublesome problem is the issue of comparability. Much of the early arguments related to utility revolved around this issue. This included early attempts in experimental economics: even searching for 'utils' and marginal utility by taking milk and bread away from people (Fisher 1927 in Blaug 1997 p.314). As mentioned in section 5.6.3, the development of Pareto optimality avoided the whole problem by ruling out inter-actor utility comparisons. In the thesis models, the situation is somewhat peculiar, in that from a 'modellers-eye' point of view, actors all use the same utility functions and share the same tastes, yet no direct comparisons are ever required of the model actors. The only ability actors are *required* to have is the basic microeconomic ones: complete, reflexive and transitive preferences. These are as follows. Complete: any two bundles of goods can be compared; reflexive: any bundle is at least as good as itself. Transitive: of three bundles of goods: x,y and z. if $x > y$ and $y > z$, then $x > z$ (Varian 2006 p.35). Section 5.3 explains how model actors go about assembling and comparing bundles to achieve this.

4.5.2 Production

Agent models of production face all the same issues as individual action and utility, not least because often precisely the same functions are used to describe both, and both are required to maximise those functions given a limited quantity of inputs. The problem of abstracting from the reality of firms' decisions is as thorny as for people. As Blaug notes, utility "no more 'explains' an individual's choices than a production-transformation curve 'explains' the state of technology" (Blaug 1997 p.337). Analysis of production, however, is clearly a more accessible problem than building testable models of people's internal mental processes. From Adam Smith's analysis of pin factories (Smith 1776) to modern work on the collective cognitive processes of productive activities (e.g. Hutchins 1996, a detailed study of a navy navigation crew), the mechanics of production have been a key focus for economics. Innovation is an ideal subject for **CAS** theory, and is often analysed as an evolutionary process. (For a recent overview see Safarzynska and Bergh 2010 pp.347; or earlier, Dawid 2006.)

So the simple approach to production taken in the thesis model needs even more careful caveats than for utility. As Storper notes of the **core model**, attempting to explain agglomeration and growth through "an indeterminate, simultaneous dance of firms, consumer-workers and product varieties and scales... is not very convincing" (Storper 2010 p.317). In particular, linking the quantity of labour input directly to an instantaneous increase in output efficiency clearly does not capture the process of developing those efficiencies. It is important, therefore, to keep claims about modelling that development separate.

Ellerman's essay on Jacobs captures the difference between growth and development nicely: development is not just 'growth' but "differentiation, diversification, and transformation in the products and in the underlying processes of production - all of which might be hidden in the black box of "total factor productivity' ". (Ellerman 2002 p.4). Referring to the common use of capital and labour inputs into production functions (K and L), Ellerman sees Jacobs' take on development as "more like the process of epigenetic transformation, not blowing up a small balloon - with more K and L - to make a big balloon" (ibid).

From this point of view, reducing development to a difference between Jacobs and Marshall externalities (see section 3.4.1) is dubious: Jacobs' argument goes beyond measuring the diversity of sectors in a region, as Glaeser does. Again, however, it comes down to the level of analysis: it is perfectly possible that Glaeser's measuring of diversity is a sensible proxy for just the kind of development Jacobs talks about, without denying the fact that the underlying mechanics are much deeper.

Firm dynamics go beyond internal structure. Section 3.5.3 described a facet of this dynamic: the trans-national corporation as a method of managing risk over space and time. As Conlisk says, production structures are "critically shaped by a need to economise on various transaction costs" (Conlisk 1996 p.675): a force for consolidation. Jacobs' take on cities as "symbiotic nests of suppliers" (Jacobs 1986 p.76) pulls the other way, making productivity a function of the 'tangled bank' of firm inter-relations (Ellerman 2002 p.6). This points more towards a focus on the combination of evolutionary dynamics and the traditional ‘forward and backward linkages’ between diverse firms (see e.g. Brakman and Heijdra 2011 p.5).

The approach to **increasing returns** in both the **core model** and the production model presented here is thus a huge simplification. But the goal is to examine the connection between production and welfare in as simple a form as possible; any conclusions must be made in the light of the complications described here.

4.5.3 The link between production and utility

The problem of modelling production starts with demand. If constrained optimisation is used, a utility function is constrained by a budget. As outlined in section 2.1.5, this produces a set of equations giving the ‘objective’ optimal quantity of goods, given that budget. The approach in **GE** (and many other economic models) is to assume the objective demand implies the correct level of production to meet it, and thus the correct number and scale of firms. In terms of agents, as Leijonhufvud puts it -

> “in this theory utility or profit maximisation is a statement about actual performance, not just motivation... The theory does not leave room for failures to realise the relevant optima.” (Leijonhufvud 2006 p.1628)

It is this kind of assumption in particular that exercised Hayek. Production is assumed as the mirror image of consumption: “consumers in evaluating (‘demanding’) consumers’ goods *ipso facto* also evaluate the means of production which enter into the production of these goods.’ ” (Schumpeter 1942 p.175 in Hayek, ibid.) Hayek’s reasoning for rejecting the ‘implicit production’ assumption differs markedly from how Friedman says assumptions should be judged. Hayek refuses to allow these assumptions, labelling them a figment of the modeller’s imagination imposed after the fact.

Jane Jacobs took a very similar line towards any descriptive production theory not grounded in the underlying mechanics. She attacked arguments that comparative advantage originates from the division of labour as teleological; “one might as well say rain is beneficial to plants and that is why it rains.” (Jacobs 1986 p.70) For Jacobs, as well as Hayek, the logic of causation is wrong. As she says, of the definition of efficiency intrinsic to **increasing returns**, claiming a country is more efficient because of specialisation “is to stand reality on its head” (ibid p.71). Jacobs wants to build a causal argument about the role of ‘nests of symbiotic suppliers’; any black box approach makes this impossible.

A similar argument occurred over the earliest use of the Cobb Douglas function to describe output for the whole economy, reducing capital and labour to single terms. As Fisher put it: “the suggestion is clear, however, that labour’s share is not roughly constant because the diverse technical relationships of modern economies are truly representable by an aggregate Cobb-Douglas but rather that such relationships appear to be representable by an aggregate Cobb-Douglas because labour’s share happens to be roughly constant” (Fisher 1971 in Robinson

1971 in Wong 1973 p.324). Wong replies: "of course, the way is open for the rebuttal that the Cobb-Douglas is not really to explain but to describe the empirical relationships found in economies" (ibid p.324).

There is a common thread that has found its way into **ABM** modelling: if the macro result has not been produced by interaction alone, it must be a *post hoc* arbitrary assumption hiding the true causal structure of the system.

It seems a little too easy to simply point to Feynman's quote and say 'it depends on what level a model wants to analyse', but - following Friedman - there is very little else to do if one accepts a model's assumptions alone cannot condemn it. At any rate, the approach to production used in the thesis model is the same as Fowler's: to stick with objective demand. As he puts it, "workers select their optimal bundle of goods without reference to the actual supply of the good" (Fowler 2011 p.10). This avoids the need to solve rationing of limited goods. It also sends "clear signals to firms about the actual level of demand for their goods at the current price" (ibid). As section 6.7 explains, it is entirely possible to do this without violating the zero stock limit. Section 6.3 outlines that, while it is possible to create a utility structure that *can* deal with subjective demand (where stock may not be available), it is not especially useful to do so for the current model goals.

4.6 Summary

How model mapping is theorised has a profound effect on the kind of approach to building models deemed 'valid'. This chapter has examined Friedman's argument in some depth and compared it to the current way in which agent modelling is done.

The argument about 'levels' suggests that, if many economists were left to investigate gas theory, they would conclude that all atoms must be temperature and pressure maximisers. On the the other hand, many agent modellers would insist temperature and pressure were of no use as concepts, and the only theoretically important feature is how atoms interact. It has been argued that there are valid reasons for treating components of a system as small 'divisions' of the larger system (as in microsimulation). **GE** demonstrates that under some limited circumstances, treating a collection of agents as a single **representative agent** is a useful modelling trick, but persuasive arguments have been presented that suggest doing so may make potentially simple behavioural models more complex than they need to be.

The result is that giving simple utility and production to disaggregated agents is a potentially useful thing to try. The next chapter presents the model framework developed for the thesis for doing this, and is followed by the results of that framework.