Information, Work and Value

Allin Cottrell & Paul Cockshott & Greg Michaelson & Ian Wright

November 21, 2007

CONTENTS

10 Money and the form of value

11 Credit and Capital

12 Understanding profit

INTRODUCTION

We live today in what is called an information economy. The growth in all things computer-related is obvious to all. Shares in companies making information processing products—the “new economy”—have been the stars of Wall Street, London and Frankfurt. At the same time an ever increasing proportion of the workforce in advanced economies has moved into jobs concerned with the handling of information: financial services, tele-centres, advertising and the media.

In the face of this obvious change we want to address some basic questions. What is information? Why is it valuable? What is the relationship between money and information?

In answering these questions we draw upon three areas of study that were until recently quite distinct: classical political economy, thermodynamics and information theory. Classical political economy links the creation of value to work. Thermodynamics, arising from pragmatic studies of the limits to our ability to perform work, became, with the concept of entropy, a cornerstone of our understanding of the physical world. Information theory, originally another pragmatic branch of engineering, has revealed unexpected links between information and entropy.

In the process we will show how concepts derived from thermodynamics have proven themselves to be amazingly fruitful in confirming the hypotheses of the classical economists.

Whilst the four authors of the book share a substantially similar perspective on the topics covered, our agreement is not total, for which reason we have chosen to identify authorship of individual chapters.

CHAPTER 1

PROBLEMATIZING LABOUR

Cockshott

1.1 WATT ON WORK

Prior to the eighteenth century, muscles—whether of humans, horses or oxen—remained the fundamental energy source for production. Not coincidentally, the concepts of work, power, energy and labour did not exist in anything like their modern form. People were, of course, familiar with machinery prior to the modern age. The Archimedean machines and their derivatives—levers, inclined planes, screws, wheels, pulleys—had been around for millennia to amplify or concentrate muscular effort. Water-power had been in use since at least the first century A.D.,[1] initially as a means of grinding grain; during the middle ages it was applied to a wide variety of industrial processes. But water-power, and its sister wind-power, were still special-purpose technologies, not universal energy sources. Limited by location and specialized use they did not problematize effort as such.

A note on terminology is in order here. The (admittedly not very elegant) verb 'to problematize' derives from the work of the Althusser and Balibar (1970) who coined the term *problématique* (problematic) to refer to the field of problems or questions that define an area of scientific enquiry. The term is fairly closely related to Thomas Kuhn's idea of a scientific 'paradigm'[2]. So, to problematize a domain is to transform it into a scientific problem-area, to construct new concepts which permit the posing of precise scientific questions. In the pre-modern era engineers and sea captains would

---

[1] See Strand (1979), Ste-Croix (1981)(p. 38).

[2] Kuhn (1970)

know from experience how many men or horses must be employed, using pulleys and windlasses, to raise a mast or obelisk. Millers knew that the grinding capacity of water mills varied with the available flow in the mill lade. But there was no systematic equation or measure to relate muscular work to water's work, no scientific problematic of effort. That had to wait for James Watt, after whom we name our modern measure of the ability to work.

Watt, the best-known pioneer of steam, did not actually invent the steam engine, but he improved its efficiency. As Mathematical Instrument Maker to the University of Glasgow he was called in to repair a model steam engine used by the department of Natural Philosophy (we would now call it Physics). The machine was a small scale version of the Newcomen engine that was already in widespread use for pumping in mines.

The Newcomen engine was an 'atmospheric engine'. It had a single cylinder, the top half of which was open to the atmosphere (Figure 1.1). The lower half of the cylinder was connected via two valves to a boiler and a water reservoir. The piston was connected to a rocking beam the other end of which supported the heavy plunger of a mine pump. The resting condition of the engine was with the piston pulled up by the counter-weight of the pump plunger.

To operate the machine, the boiler valve was opened first, filling the cylinder with steam. This valve was then closed and the water-reservoir valve opened, spraying cold water into the piston. This condensed the steam, resulting in a partial vacuum. Atmospheric pressure on the upper surface of the piston then drove it down, providing the power-stroke. The two phase cycle could then be repeated to obtain regular pumping.

Watt observed that the model engine could only carry out a few strokes before the boiler ran out of steam and it had to rest to 'catch its breath'. He ascertained that this was caused by the incoming steam immediately condensing on the walls of the cylinder, still cool from the previous water spray. His solution was to provide a separate condenser, permanently water cooled, and intermittently connected to the cylinder by a valve mechanism. The cylinder, meanwhile, was provided with a steam-filled outer jacket to keep its inner lining above condensation temperature (Figure 1.1). His 1769 patent was for "A New Method of Lessening the Consumption of Steam and Fuel in Fire Engines".

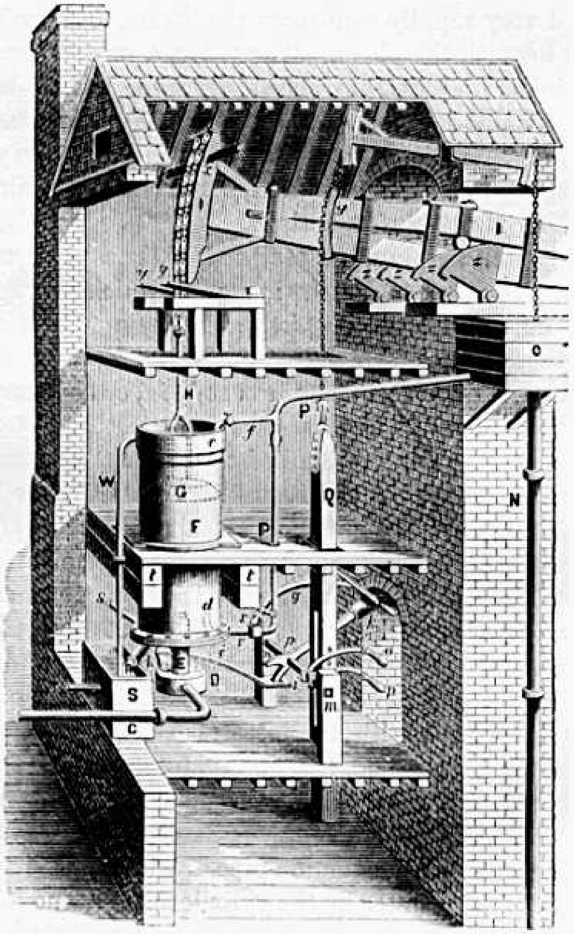


Figure 1.1: The Newcomen engine built by Smeaton (reproduced from Thurston)

Watt's later business success was based directly on this gain in thermal efficiency. His engines were not sold outright to users, but were leased. The rental paid was equal to one-third the cost of coal saved through using a Watt engine rather than a Newcomen engine[3]. This pricing system worked so long as the Newcomen engine provided a basis for comparison, but as Watt's engines became the predominant type, and as they came to be used to power an ever-widening range of machines, some system of rating the working capacity of the engines was needed. Watt needed a standardized scale by which he could rate the power, and thus the rental cost, of different engines. His standardized measure was of course the horsepower: users were charged £5 per horsepower year.

Watt's horse was not a real horse of course, but the abstraction of a horse, a standardized horse. The abstraction is multiple: at once an abstraction from particular horses, an abstraction from the difference between flesh and blood horses and iron ones, and an abstraction from the particular work done. The work done had to be defined in the most abstract terms, as the overcoming of resistance in its canonical form, namely raising weights. One horsepower is 550 ft lb/sec, the ability to raise a load of 1 ton by 15 feet in a minute.

While few real horses could sustain this kind of work, its connection to the task performed by Watt's original engines is clear. The steam engine was a direct replacement for horse-operated pumps in the raising of water from mines. But with the development of mechanisms like Watt's sun and planet gear, which converted linear to rotary motion, steam engines became a general purpose power source. They could replace water wheels in mills, drive factory machines by systems of axles and pulleys, pull loads on tracks. Engine capacity measured in horsepower abstracted from the concrete work that was being performed, transforming it all to **work** in general. Horsepower was the capacity to perform a given amount of work each second. By defining power as work done per second, work in general was itself implicitly defined. All work was equated to lifting. Work in general was defined as the product of resistance overcome, measured in pounds of force, by the distance through which it was overcome.

Mechanical power seemed to hold the prospect of abolishing human drudgery and labour. As Matthew Boulton proudly announced to George II: "Your Majesty, I have at my disposal what the whole world demands;

---

[3] Tann (1981)

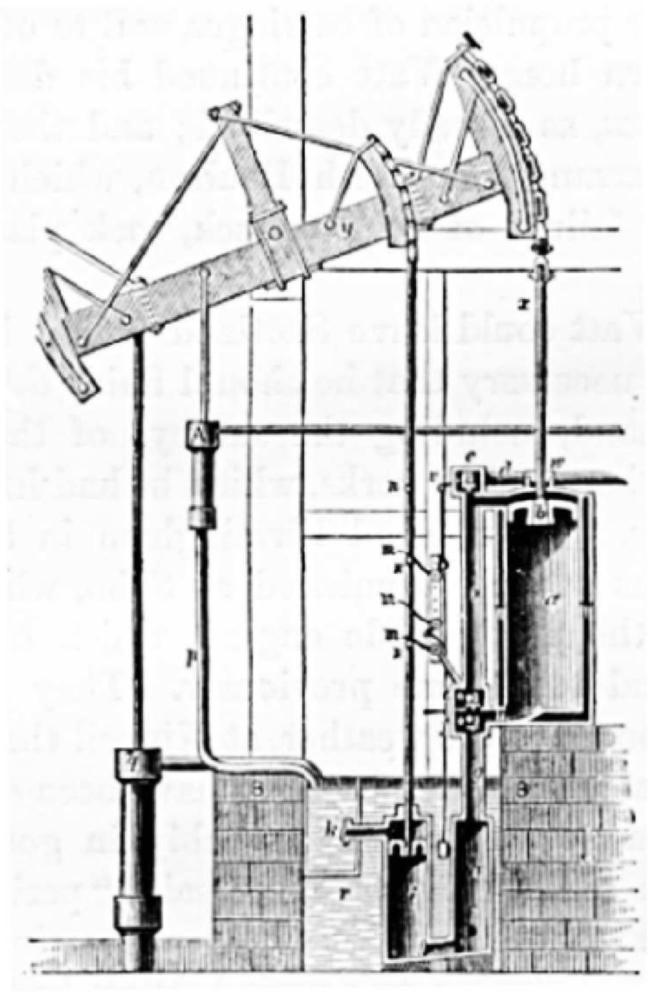


Figure 1.2: Watt’s steam engine with separate condenser (reproduced from Thurston)

something which will uplift civilization more than ever by relieving man of undignified drudgery. I have steam power."[4] To a world in which human muscle was a prime mover, this equation of work in the engineering sense with human labour was exact. Work on ships, in mines, at the harvest, was work in the most basic physical sense. Men toiled at windlasses to raise anchors, teams pulled on ropes to set sails and hauled loads on their backs to unload cargo. Children dragged coal in carts from drift mines, women carried it up shafts in baskets on their backs. The 'navigators' who built canals did it with no mechanical aid more sophisticated than the wheelbarrow (a combination of lever and wheel, two Archimedean devices).

As horsepower per head of population multiplied, so too did industrial productivity. The power of steam was harnessed, first to raise weights, then to rotate machinery, then to power water-craft, next to trains—and eventually, through the mediation of the electricity grid, to tasks in every shop and home—while human work shrank as a proportion of the total work performed. More and more work was done by artificial energy, yet the need for people to work remained. A steam locomotive might draw a hundred-ton train, but it needed a driver to control it. Human work became increasingly a matter of supervision, control and feeding of machines. Thus the identification of work with the overcoming of physical resistance in the abstract, and of human labour-power with power in Watt's sense, contained both truth and falsehood. Its truth is shown by the manifest gains flowing from the augmentation of human energy. Its falsity is exposed by the residuum of human activity that expresses itself in the control, minding and direction of machinery.

Indeed, the introduction of powered machinery had the effect of lengthening the working day while making work more intense and remorseless. The cost of powered machinery was such that only men with substantial wealth could afford it. Cheap hand-powered spindles and looms could

---

[4] Compare Antipater of Thessalonika's eulogy on the introduction of the water mill:

> Stop grinding, ye women who toil at the mill
> Sleep on, though the crowing cocks announce the break of day
> Demeter has commanded the water nymphs
> to do the work of your hands
> Jumping one wheel they turn the axle
> Which drives the gears and the heavy millstones

not compete with steam-powered ones. Domestic spinners and hand-loom weavers had to give up their independence and work for the owners of the new steam powered ‘mules’ and looms. Steam power brought no increase in leisure for weavers or spinners. The drive to recoup the capital cost of the new machinery brought instead longer working hours and shift-work, to a rhythm dictated by the tireless engine. The fact that the machinery was not owned by those who worked it, meant that it enslaved rather than liberated.

A particular pattern of ownership was the social cause of machine-enforced wage slavery, but that is only half the story. We may ask why the new machine economy needed human labour at all. Why did ‘self acting’—or as we would put it now, ‘automatic’—machines not displace human labour altogether? A century ago, millions of horses toiled in harness to draw our loads. Where are they now? A remnant of their former race survives as toys of the rich; the rest went early to the knackers. Why has a similar fate not befallen human workers? Why has the race of workers not been killed off, to leave a leisured rich attended by their machines?

Watt’s horsepower killed the horse, but the worker survived. There must be some real difference between work as defined by Watt, and work in the sense of human labour.

1.2 MARX: THE ARCHITECT AND THE BEE

Karl Marx proposed an argument which seems at first sight to get to the essence of what distinguishes human labour from the work of an animal or a machine, namely purpose.

> An immeasurable interval of time separates the state of things in which a man brings his labour-power to the market for sale as a commodity, from that state at which human labour was still in its first instinctive stage. We pre-suppose labour in a form which stamps it as exclusively human. A spider conducts operations that resemble those of a weaver, and a bee puts to shame many an architect in the construction of her cells. But what distinguishes the worst of architects from the best of bees is this, that the architect raises his structure in the imagination before he erects it in reality. At the end of every labour process we get a result that already existed in the imagination of the labourer at its commencement. He not only effects a change of form in the material on which he works, but he also realises a purpose

> of his own that gives the law to his modus operandi, and to which he must subordinate his will. (Marx (1954) pp. 177–8)

This suggests that animals, lacking purpose, can be replaced by machines, but that humans are always required, in the end, to give purpose to the machine. We cite Marx's statement because it articulates what is probably a rather widely held view, yet it has several interesting problems. This is an issue where it is difficult to go straight for the 'right answer'. It may be profitable to beat the bushes first, to scare up (and shoot down) various prejudices that can block the road to a scientific understanding.

First, are animals really lacking in purpose? The spider may be so small, and her brain so tiny, that it seems plausible that blind instinct, rather than the conscious prospect of flies, drives her to spin. But it is doubtful that the same applies to mammals. The horse at the plough may not envisage in advance the corn he helps to produce, but then he is a slave, bent to the purpose of the ploughman. Reduced to a source of mechanical power, overcoming the dumb resistance of the soil, he is readily replaced by a John Deere. The same cannot be said of animals in the wild. Does the wolf stalking its prey not intend to eat it? It plans its approach with cunning. Who are we to say that the result—fresh caribou meat—did not "already exist in the imagination" of the wolf at its commencement? We have no basis other than anthropocentric prejudice on which to deny her imagination and foresight.

Turn to Marx's human example, an architect, and his argument looks even shakier. For do architects ever build things themselves? They may occasionally build their own homes, but in general what gives them the status of architects is that they don't get their hands dirty with anything worse than India Ink. Architects draw up plans. Builders build. (In eliding this distinction Marx showed an uncharacteristic blindness to class reality).

An office block, stadium or station has, it is true, some sort of prior existence, but as a plan on paper rather than in the mind of the builders. If by collective labour civilized humans can put up structures more complex than bees, it is because they can read, write and draw. A plan—whether on paper or, as in earlier epochs, scribed on stone—coordinates the individual efforts of many humans into a collective effort.

For building work then, Marx is partially right, the structure is raised *on paper* before it is raised in stone. But he is wrong in saying that it is built in the imagination first, and in implying that the structure is put up by the architect. What is really unique to humans here is, first, the social division of labour between the labour of conception by the architects and the work of execution by the builders, and second, the existence of *materialized plans*: configurations of matter that can control and direct the labour of groups of humans.

While insect societies may have a division of labour between ‘castes’, for example between worker and soldier termites, they do not have a comparable division between conception and execution, between issuers and followers of orders. Nor do insects have technologies of record and writing. They can communicate with each other. Dancing bees describe to others the whereabouts of flowers. Walking ants leave scent trails for their companions. These messages, like human speech, coordinate labour. Like our tales, they vanish in the telling. But, not restricted to telling tales, we can can make records that persist, communicated over space and time.

Our tales are richer too. The set of messages that can be expressed in our languages is exponentially greater than in the language of bees. Each works by the sequential combination of symbols—words for us, wiggles for bees—but we have many more symbols and can understand much longer sequences. The number of distinct messages that can be communicated by a language is proportional to v^m where v is the number of distinct symbols that can be recognized in the language and m is the maximum message length. If bees have a repertoire of six types of wiggles and can understand ‘sentences’ of three wiggles in succession then they can send $6^3 = 216$ different messages. A human language with a vocabulary of 3000 words and a maximum sentence length of 20 words could convey about $3.486784401 \times 10^{69} = 348,678,440,100,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000$ distinct sentences. Of course, not all of these would be grammatically correct, and a rather small proportion of those would make any sense, but the number of messages is still astronomically greater than what insects can manage. And we can keep piling on the sentences until the listener loses track.

All this leaves open another interpretation of what Marx had to say. True enough, architects may not build theatres themselves, any more than Hadrian built his wall[5] or Diocletian his baths. But Hadrian caused the wall to be built and Diocletian’s architect caused the baths to be built to a specific design. (This use of the word ‘built’ is of course common in class societies,

---

[5] It was of course the rank and file legionnaires who built the wall; see Davies (1989).

where real builders get no credit for their creations. Their labour contributes instead to the fame of a ruler or architect.) If the architect creates only a paper version of a theatre, can we say, at any rate, that he creates this drawing in his mind before setting it down on paper? This interpretation of Marx's story of the architect and the bee seems to make sense, but it's not clear that it's a true description of what an architect actually does.

1.2.1 Emergent buildings

Some individuals, autistic infant prodigies or 'idiot savants', do seem to have the ability to hold in their minds almost photographically detailed images of buildings they have seen. Working from memory they are able to draw buildings in astonishing and accurate detail.[6] But it is questionable whether professional architects work this way. Some may, but for others the process of developing a design is intimately tied up with actually drawing it. They start with the broad outlines of a design in their minds. As this is transferred to paper, they get the contexts within which the mind can work to elaborate and fill in details. The details were not in the mind prior to starting work, they emerge through the interaction of mind, pen and paper. Pencils and paper don't just record ideas that exist fully formed, they are part of a production process that generates ideas in the first place.

At any one time our consciousness can focus on only a limited number of items. On the basis of what it is currently conscious of, its context, it can produce responses related to this context. In reverie the context is internal to the brain and the responses are new ideas related to this context. In an activity like drawing a plan or engineering diagram, the context has two parts

(1) an internal state of mind; and

(2) that part of the diagram upon which visual attention is fixated,

and the response is both internal—a new state of mind—and external—a movement of the pencil on the paper.[7] Where in reverie the response, the new idea, slipped all too easily from grasp, paper remembers.[8] Architecture

---

[6] It may be worth seeing if we could reproduce some images by such autistic artists

[7] The reader may notice that this argument is a thinly disguised version of Alan Turing's famous argument, see: Turing (1937).

[8] Cite the passage in Tacitus, I think it is in the Annals, where he says that civilization depends upon Papyrus.

exchanges for the fallibility and limited compass of memory the durability of an effectively infinite supply of A0. One might say that complex architecture rests on paper foundations.

If the idea of the architect as creating buildings spontaneously out of the imagination is dismissed as an almost religious myth, redolent of the Masonic characterization of the deity as the *Great Architect*, what then remains of the antithesis between architect and bee? Well, how do the bees shape their hive? We can be sure there are no drawings of hexagons, made by the 'queen',[9] and executed by her worker daughters. We are talking here of *apis mellifera* not the solitary bumble bee. The labour of the honey bees is collective, like that of workers on a building site, yet although they have no written plans to work from they create a geometrically precise, optimal and elegant structure.

1.2.2 Apian efficiency

Consider the problem to which the honeycomb is the answer: to come up with a structure that is interchangeably capable of storing honey or sheltering bee larvae, is waterproof, is structurally stiff, provides a platform to walk on and which uses the minimum material. Given this design brief it is unlikely that a human engineer could come up with a better structure.

The structure has to be organized as a series of planes to provide access. Within the planes, the combs, the space has to be divided into approximately bee-sized cubicles. These could be triangular, square, or hexagonal (the only three regular tessellations of the plane). Our architects have a predilection for the rectangular, but the hexagonal form is superior.

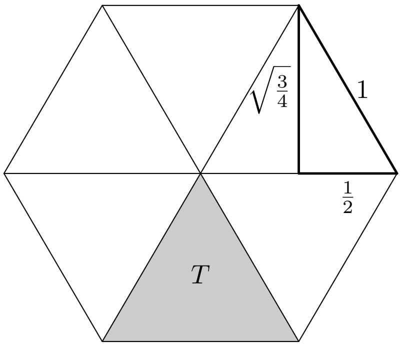
A tessellation of unit squares has a wall length of 2 per unit area, since a single unit square has four sides of unit length, each shared 50 percent with its neighbours. A tessellation of hexagons of unit area has a wall length of $\frac{2}{\sqrt{3}}$ per unit area, a reduction by a factor of $\sqrt{3}$ (see Digression 1.1). The honeycomb structure used by bees is thus more efficient in its use of wax than a rectilinear arrangement would be.

The fact that hexagonal lattices minimize boundary lengths per unit area means that they can arise spontaneously, for example in columnar basalts.

---

[9] The breeding female is no more an architect or Caesar than the Pope is the genetic father of his followers. Monarchy and patriarchy project dominance relations onto genetic relations and vice versa. Apian Mother becomes queen, the Vatican monarch, Holy Father.

Digression 1.1 Apian efficiency



(1) A hexagon of unit side is made up of 6 identical equilateral triangles, thus its area is $6T$ where T is the area of an equilateral triangle of unit side.

(2) The area of an equilateral triangle of unit side is $\frac{1}{2}bh$ where b the base $= 1$ and h the height $= \sqrt{\frac{3}{4}}$. So $T = \frac{1}{2}\sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{4}$.

(3) The area of one hexagon is then

$$6\frac{\sqrt{3}}{4} = \frac{3\sqrt{3}}{2}$$

(4) The hexagon's six sides are each shared 50% with a neighbour.

(5) Wall per unit area for a hexagonal tessellation is then $3/\frac{3\sqrt{3}}{2} = 2/\sqrt{3}$ which is better than the wall to area ratio for squares.

The Honeycomb Conjecture has been debated since at least 36BC when it was mentioned by Varro in his book on agriculture. It has been remarkably difficult to prove. Here we have considered only a comparison between hexagonal tesselations and square ones. There remains the possiblity that some layout using curved walls might be still more efficient. A full proof of the conjecture was not produced until 2001 Hales (2001).

Figure 1.3: Tessellation of the plane using hexagons

Here the tension induced in rocks as they cool encourages cracking, preferentially giving rise to six sided columns. We might suspect that the beehive too, gained its structure from a process of spontaneous pattern formation analogous to columnar basalts or packed arrays of soap bubbles. But this doesn't tally with the way the cells are built up, or with the uniformity of their dimensions. In a partially constructed honeycomb the cells are of a constant diameter; those in the middle of the comb are all of uniform height while towards the edge the depth of the cells falls. The bees build the cells up from the base, laying wax down on the upper margins of the cell walls, just as bricks are added to the upper margin of a wall by a bricklayer. The construction process takes advantage of the inherent stability of a hexagonal lattice, allowing the growing cells to form their own scaffolding. But the process also demands that the bees can deposit wax accurately on the growing cell walls, and that they stop building when the cells have reached the right height. That is, it depends on purposeful activity on the part of the bees.

A similar process takes place in the human construction of geodesic domes, hexagonal lattices curved through a third dimension. These have an inherent stability that becomes more and more evident as you add struts to them. You build them up in a ring starting at ground level. The structure initially has a fair bit of play in it, but the closer the structure comes to a sphere the more rigid it is. Human dome builders, like bees, exploit the inherent structural properties of hexagonal lattices, but they still need to cut struts to the right length and put them in the correct place. The bees likewise must select the right height for their cell walls and place wax appropriately.

Spontaneous self-assembly of hexagonal structures similar to geodesic domes does occur in nature. The Fullerenes are a family of carbon molecules

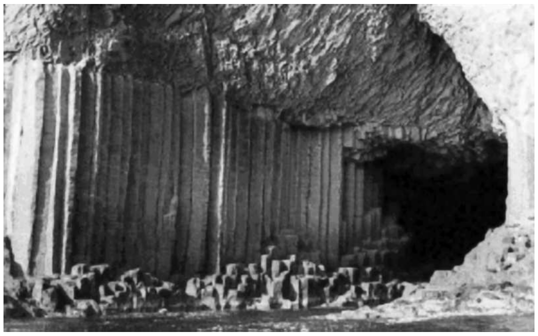


Figure 1.4: Nature is the architect of the hexagonal columns of Fingal’s cave (Photo by Andrew Kerr)

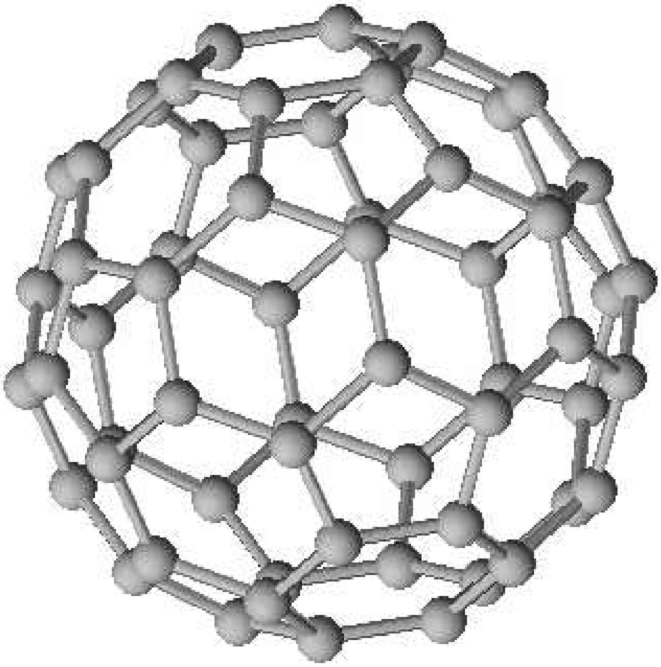


Figure 1.5: C_{60} a spontaneously formed dome structure

named after Buckminster Fuller, the inventor of the geodesic dome. The first of these to be discovered, C_{60}, has the form of a perfect icosahedron (see Figure 1.5). Condensed out of the hellish heat of a carbon arc, it depends on thermal vibrations to curve the familiar planar hexagonal lattice of graphite onto itself to form a three dimensional structure. No architect or bee is required. Atomic properties of carbon select the strut length. Thermal motion searches the space of possible configurations; a small fraction of the molecules settle into the local energy minima represented by C_{60} and its sisters.

If the bees can't rely upon spontaneous self-assembly to build their hives, must they have a plan in mind before they start? Since they can't draw, the mind would have to be where they held any plans. While we can't rule this out, it seems unlikely. The requirement is that they can execute a program of work. A bee arriving on the construction site must, in the darkness, find an appropriate place to put wax, for which they need a set of rules:

```
If the cell is high enough to crawl into, put no more wax on it,
otherwise if the cell has well formed walls add to their height,
otherwise if it is a cell base smaller than your own body diameter,
    expand it,
otherwise start building the wall up from the base...
```

No internal representation of a completed comb need be present in the bee's mind. The same rules, simultaneously present in each of a hive full of identical cloned sisters, along with the structural properties of beeswax, produce the comb as an emergent complex structure. The key here is the interaction between behavioral rules and an immediate environment that is changed as the result of the behaviour. The environment, the moulded wax, records the results of past behaviour and conditions future behaviour. But for rules to be converted into behaviours by the bees, the bees must have internal 'states of mind', and be able to change their state of mind in response to what their senses are telling them. A bee that is busy laying down wax is in a different state of mind from one foraging for pollen and their behavioral repertoire differs as a result.

As we have argued above, what an architect does is not so different. Architects produce drawings, not buildings or hives, but producing a drawing is an interactive process in which the architect's internal state of mind, his knowledge of the rules and stylistic conventions of the epoch, produces behaviour that modifies the immediate environment—the paper. The change to the paper creates a new environment, modifying his state of mind and calling into action other learned rules and skills. The drawing is an emergent property of the process, not something that pre-existed as a complete internal representation before the architect put pencil to paper.

1.3 THE DEMONIC CHALLENGE

Purposeful labour depends upon the ability to form and follow goals. A goal is a representation of a state of affairs that does not exist plus a motivation to achieve it. Although bees do not have the goal processing capabilities of the human mind, they nonetheless follow simple goals. Goal processing, from simple, reactive programs hard-wired in the neural circuitry of insects, to the much more adaptive and sophisticated rational planning capabilities of humans, is the mechanism that distinguishes the constructive activity of humans and bees from the blind efforts of Watt's engines. An engine transforms energy in one form to another, but it does not act to achieve states of affairs, unlike bees that build or humans that labour.

There is a hidden connection between purposeful labour and work in the engineering sense. Any purposeful activity overcomes physical resistance and involves *work*, measured in watts, for which we must be fueled by calories in our food; the hidden connection comes from the realization that, at least in principle, purposeful labour could itself be a source of fuel.

Recall that Watt's key invention was the separate condenser for steam engines, which saved fuel by preventing wasteful condensation of steam within the cylinder of the engine. In the years after Watt's invention, it came to be realized that the thermal efficiency of steam engines could be improved by maximizing the pressure drop between the boiler and the condenser. A series of inventions followed to take advantage of this principle: Trevithick's high pressure engine, the double and then the triple expansion engine. These had the effect of increasing the amount of effective work that could be extracted from a given amount of heat. But successive gains in efficiency proved harder to come by. The amount of work obtained per calorie of heat could be increased, but not without limit.

It was understood that work could be converted into heat, for instance through friction, and heat could be converted back into work, for instance by a steam engine. But if you convert work into heat, and heat back into work, you always end up with less work than you put in. In converting work into heat, the number of calories of heat obtained per kilowatt hour of work is constant—conversion of work into heat can be done with 100 percent efficiency. The reverse is not true. Heat can never be fully converted into useful work.[10] The practical imperative of improving steam engines gave rise to the theoretical study of the laws governing heat, the laws of thermodynamics.

One of the first formulations of the second law of thermodynamics was that heat will never spontaneously flow from somewhere cold to somewhere hot.[11] This implied that, for instance, there was no chance of transferring the heat wasted in the condenser of a steam engine back to the boiler where it would boil more water. Thermodynamics ruled out perpetual motion machines.

But James Clerk Maxwell, one of the early researchers in thermodynamics, came up with an interesting paradox.

> One of the best established facts of thermodynamics is that it is impossible in a system enclosed in an envelope which permits neither change of volume nor passage of heat, and in which temperature and pressure are everywhere the same, to produce any inequality of temperature or of pressure without the expenditure of work. This is the second law of thermodynamics, and it is undoubtedly true as long as we can deal with bodies only in mass, and have no power of perceiving or handling the separate molecules of which they are made up. But if we can conceive of a being whose faculties are so sharpened that he can follow every molecule in its course, such a being would be able to do that which is presently impossible to us. For we have seen that the molecules in a vessel full of air at a uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to

---

[10] Carnot was able to show that the efficiency of heat engines depended on the temperature difference between heat source, for example the boiler, and the heat sink, for example a steam engine's condenser.

[11] This formulation was due to Clausius in 1850; see (Porter (1946), pp. 8–9).

> B, and only the slower ones to pass from B to A. He will thus, without the expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics. (Maxwell (1875), pp. 328–329)

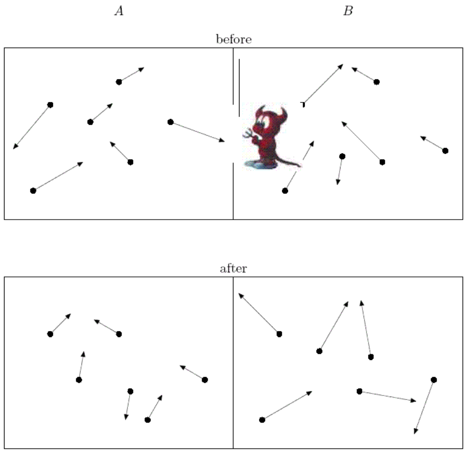


Figure 1.6: Gas initially in equilibrium. Demon opens door only for fast molecules to go from A to B, or slow ones from B to A. Result Slow molecules in A, fast in B. Thus B hotter than A, and can be used to power a machine.

The configuration of the thought experiment is shown in Figure 1.6. As the experiment runs the gas on one side heats up while that on the other side cools down. The end result is a preponderance of slow molecules in cavity A, fast ones in cavity B. Since heat is nothing more than molecular motion, this means that A has cooled down while B has warmed up. No net heat has been added, it has just re-distributed itself into a form that becomes useful to us. Since B is hotter than A, the temperature differential can be used to power a machine. We can connect B to a boiler and A to a condenser and obtain mechanical effort. An exercise of purposeful labour by the demon outwits the laws of thermodynamics. (Norbert Wiener coined the term 'Maxwell demon' for the tiny 'being' envisaged in the thought experiment.) It seems that the second law of thermodynamics expresses the coarseness of our senses rather than the intractability of nature.

1.4 ENTROPY

One perspective on the devilment worked by Maxwell's demon is that it has *reduced the entropy* of a closed system. The idea of entropy was introduced by Clausius in 1865 (see Harrison, 1975) with the equation

$$\Delta S = \Delta Q / T \tag{1.1}$$

where ΔS is the change in entropy of a system consequent upon the addition of a quantity of heat ΔQ at absolute temperature T.[12] According to Clausius's equation adding heat to a system always increases its entropy (and subtracting heat always lowers entropy) but the magnitude of the change in entropy is inversely related to the initial temperature of the system. Thus if a certain amount of heat is transferred from a hotter to a cooler region the increase in entropy in the cooler region will be greater than the reduction in entropy in the hotter, and overall entropy rises. Conversely, if heat is transferred from a colder to a hotter region entropy falls. Clausius's concept of entropy as an abstract quantity allowed him to give the second law of thermodynamics its canonical form: the entropy of any closed system tends to increase over time.

Using (1.1) we can readily see that Maxwell's demon violates the second law of thermodynamics. Suppose the demon has been hard at work for some time, so that B is hotter than A, specifically B is at 300° Kelvin and A is at 280° Kelvin. He then transfers ΔQ =1 joule of heat from A to B. In doing so he reduces the entropy of A by $\frac{1}{280}$ joules per degree and increases the entropy of B by $\frac{1}{300}$ joules per degree giving rise to $\Delta S = \frac{1}{300} - \frac{1}{280} = -\frac{1}{4200}$, a net reduction in entropy, contrary to the second law.

Clausius's formulation of entropy did not depend in any way upon the atomic theory of matter. Maxwell's proposed counter-example to the second

---

[12] At this stage the concept of entropy remains firmly linked to the sort of practical considerations, namely steam engine design, that gave rise to thermodynamics. Later, as we shall see, it becomes generalized.

law was explicitly based on atomism. With Boltzmann, entropy is placed on an explicitly atomistic foundation, in terms of an integral over molecular *phase space*.

$$S = -k \int f(v) \log f(v) dv \tag{1.2}$$

where v denotes volume in six-dimensional phase space, $f(v)$ is the function that counts the number of molecules present in that volume, and k is Boltzmann's constant.

The concept of phase space is a generalization of our normal concept of three-dimensional space to incorporate the notion of motion as well as position. In a three-dimensional coordinate system the position of each molecule can be described by three numbers, measurements along three axes at right angles to one another. We usually label these numbers x, y, z to denote measurements in the horizontal, vertical and depth directions. However each molecule is simultaneously in motion. Its motion can likewise be broken into components of horizontal, vertical and depth-wise motion which we can write as m_x, m_y, m_z, representing motion to the left, up and back respectively. This means that a set of six coordinates can fully describe both the position and motion of a particle.

In Boltzmann's formula, the letter v denotes a range of possible values of these co-ordinates. For example, a volume 1 mm cubed on the spatial axes and 1 mm per second on the motion axes. The function $f(v)$ would then specify how many molecules there were in that cubic millimeter with a range of velocities within 1 mm per second in each direction. Boltzmann's formula relates the entropy of a gas, for instance steam in a piston, to the evenness of its distribution in this six dimensional space: the less even the distribution the lower the entropy. This point is illustrated in simplified manner in Table 1.1. Suppose we have just two cells in phase space, and eight atoms that can be in one cell or the other. The table shows how the entropy depends on the location of the atoms, lowest when all 8 are in one cell, and highest when they are evenly divided between the cells. (Note that the minus sign in Boltzmann's formula is needed to make entropy increase with the evenness of the distribution, consistent with Clausius's earlier formulation.)

Boltzmann also showed that it is possible to reformulate the idea of entropy using the concept of the 'thermodynamic weight' of a state:

$$S = k \log W \tag{1.3}$$

Contents of cells 1, 2	$f(1)\log f(1)+f(2)\log f(2)$	Entropy, S
8, 0	8(2.079) + 0 = 16.636	$-16.636k$
7, 1	7(1.946) + 1(0) = 13.621	$-13.621k$
6, 2	6(1.792) + 2(0.693) = 12.137	$-12.137k$
5, 3	5(1.609) + 3(1.099) = 11.343	$-11.343k$
4, 4	4(1.386) + 4(1.386) = 11.090	$-11.090k$

Table 1.1: Boltzmann's entropy: Illustration

The thermodynamic weight W is the number of physically distinct microscopic states of the system consistent with a given 'macro' state, described by temperature, pressure and volume. This concept is the key to understanding the second law. Recall that the entropy of closed systems tends to increase, that is they move into macro-states of progressively higher thermodynamic weight until they reach equilibrium. States with higher weight are *more probable.* So the second law of thermodynamics basically says that systems evolve into their most probable state.

A simple analogy may be helpful here. Suppose a 'fair' coin is flipped ten times. What is the most likely ratio of heads to tails in the sequence of flips? The obvious answer, 5/5, is correct. Now, what is the most likely specific sequence of heads and tails? Trick question! There are $2^{10} = 1024$ such sequences and they are all equally likely. The sequence featuring 10 heads has probability $\frac{1}{1024}$; so does the sequence with 5 heads followed by 5 tails; so does the sequence of strictly alternating heads and tails, and so on. The reason why a 5/5 ratio of heads to tails is most likely is that there are more specific sequences corresponding to this ratio that there are sequences corresponding to 10/0, or 7/3, or any other ratio. It's easy to see there is only one sequence corresponding to all heads, and one corresponding to all tails. To count the sequences that give a 5/5 ratio, imagine placing the 5 heads into 10 slots. Head number 1 can go into any of the ten slots; head number 2 can go into any of the remaining 9 slots, and so on, giving $10 \times 9 \times 8 \times 7 \times 6$ possibilities. But this is an over-statement, because we have treated each head as if it were distinct and identifiable. To get the right answer we have to divide by the number of ways 5 items can be assigned to 5 slots, namely $5 \times 4 \times 3 \times 2 \times 1$. This gives 252 possibilities. Thus the

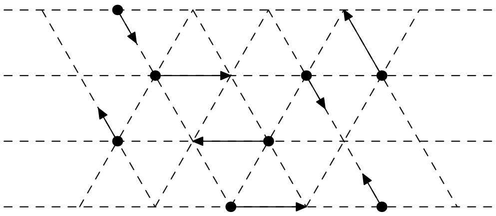


Figure 1.7: The molecules in a lattice gas move along the lines of a triangular grid with fixed velocities

'macro' result, equal numbers of heads and tails, corresponds to 252 out of the 1024 equally likely specific sequences, and has probability $\frac{252}{1024}$. By the same reasoning we can figure that a 6/4 ratio corresponds to 210 possible sequences, a lower 'weight' than the 5/5 ratio.

The number of possible states of a real gas in six-dimensional phase space is hard to visualize, so to explicate the matter further we'll examine a simpler system, namely a two-dimensional *lattice gas* Frisch et al. (1986). The 'molecules' in such a stylized gas move with constant speed, one step along the lattice per unit time (see Figure 1.7). Where the lines of the lattice meet, molecules can collide according to the rules of Newtonian dynamics, so that matter, energy and momentum are conserved in each collision. The different ways in which collisions occur can be summarized by two simple rules:

(1) If a molecule arrives at an intersection and no molecule is arriving on the diagonally opposite path, then the molecule continues unimpeded.

(2) If two molecules collide head on they bounce off in opposite directions, as shown in Figure 1.8.

Lattice gases are a drastic simplification of real gases, but they are useful tools in analysing real situations. The simple rules governing the behaviour of lattice gases make them ideal models for simulation in computer software or special purpose hardware (Shaw et al, 1996).

Since the velocity of the molecules in a lattice gas is fixed, the temperature of the gas can't change (this would involve a rise or fall in the molecules' speed). So Maxwell's original example of a being with precise senses, able to sort molecules by speed, is inappropriate. But we can invent another demon to guard the trapdoor. Instead of letting only fast molecules

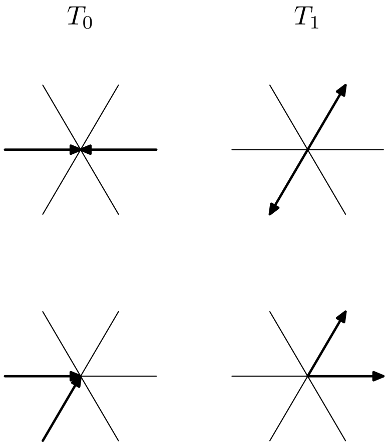


Figure 1.8: Collisions in a lattice gas: 'Molecules' colliding head on bounce off at 60° angles (above). In other cases the collision is indistinguishable from a miss (below). In all cases Newtonian momentum and energy are conserved.

through from A to B, this being will keep the door open unless a molecule approaches it from side B. Thus molecules approaching from side A are able to pass into B, but those in B are trapped. The net effect is to raise the pressure on side B relative to A while leaving temperature unchanged.

A lattice gas has only a finite number of lattice links on which molecules can be found, and since the molecules move with a constant velocity, Boltzmann's formula (1.3) simplifies to:

$$S = -kn \sum_i p_i \log p_i \tag{1.4}$$

where p_i is the probability of the node being in state i and n is the number of nodes. The weighted summation over the possible states has the effect of giving us the mean value of $\log p$. Suppose we have a very small pair of chambers, A and B, each of which initially has n nodes, and each containing $3n$ randomly distributed molecules. Then each of the six incoming paths to a node will have a 50 percent chance of having a molecule on it. We have $6n$ incoming paths to our nodes, and each of these has two equally likely states: a particle is or is not arriving at each instant. Each incoming path contributes $k \log 2 = 0.693k$. The total entropy of the chamber is then six time this or:

$$\text{Entropy of A in equilibrium} = 4.158kn.$$

Now suppose that our demon has been operating for some time, letting n particles pass from A to B, so that A now contains $2n$ particles and B contains $4n$ particles. In A, the probability of a molecule coming down any one of the paths is now only $\frac{1}{3}$. We can calculate the current entropy contribution of each incoming path as follows:

Number of particles	probability, p_i	$\log p_i$	entropy, $-kp_i \log p_i$
0	$\frac{2}{3}$	-0.405	$0.27k$
1	$\frac{1}{3}$	-1.098	$0.366k$
total			$0.636k$

The entropy of A after n particles have been transferred by the demon is $3.816kn$ which is less than before he got to work. By symmetry of complementary probabilities the entropy of chamber B will be the same,[13] thus the whole closed system has undergone a reduction in entropy.

This establishes that when an initially dispersed population of particles—the gas molecules in our case—is concentrated, entropy falls.[14] This is because there are a greater number of possible microstates compatible with dispersion than with concentration, and entropy is just the log of the number of microstates.

Consider in this light the work of the bees building their hive. There are two aspects to the work:

(1) The bees first have to gather wax and nectar from flowers dispersed over a wide area and bring it to the hive.

(2) They must then form the wax into cells and place the concentrated nectar in these as honey.

Both processes are entropy-reducing with respect to the wax and the sugar. The number of possible configurations that can be taken on by wax within

---

[13] This will not generally be the case; we have chosen the particle densities so as to ensure this.

[14] This is true on the assumption that the potential, gravitational or electrostatic, of the particles is unchanged by the process of concentration as in our example.

the few litres volume of a hive is enormously less than the number of possible configurations of the same wax, dispersed among plants growing over tens of thousands of square meters of ground. Similarly the chance that the wax, if randomly thrown together within the hive, should assume the beautifully regular structure of a comb, is vanishingly small. That the wax should be in the hive in the first place, is, in the absence of bees, highly improbable; that it should be in the form of regular hexagons even more so.

The second law of thermodynamics specifies that the total entropy in a closed system tends to increase, but the bees and their wax are not a closed system. The bees consume chemical energy in food to move the wax. If we include the entropy increase due to food consumed, the second law is preserved.

1.4.1 Men and horses

Let us return to the question we asked in section 1.1: Why did the introduction of the steam engine, which made redundant the equine workers of the pre-industrial age, not also replace the human workers? We can make a rough analogy between the work done by horses in past human economies and the work done by the bees in transporting wax and nectar from flower to hive. This is in the main sheer effort, work in Watt's sense. Horses bringing bricks to a building site or bees transporting wax are doing similar tasks. What remains, the construction of the hive after the work of transportation is done or the building of the house once the bricks are delivered, is something no horse can do. Construction involves a complex program of actions deploying grasping organs, hands, mandibles, beaks etc., in which the sequence of operations is conditioned by the development of the product being made. Human construction differs from that of a bee or a bird in:

(1) the way in which the program of action comes into being;
(2) the way in which it is transmitted between individuals of the species; and
(3) the form in which it is materialized.

In the social insects the programs of action largely come into being through the evolutionary process of natural selection. They are transmitted between parents and their offspring genetically encoded in DNA, and they are materialized in the form of relatively fixed interactions between components of the nervous system and general physiology. In humans the programs of action are themselves products that can have a representation external to the organism, in speech or some form of notation. Speech and notation act both as a means of transmission between individuals, and as a possible form of materialization of work programs while the work is being carried out—as for example, when one cooks from a recipe or follows a knitting pattern. The ability to make and distribute new work programs distinguishes human labour from that of bees and is the key to cultural evolution.

But even the work of transport requires a program of action, requires guidance if it is to reduce entropy. Transport is not diffusion. It moves concentrated masses of material between particular locations, it does not spread them about willy nilly. Without guidance there is no entropy reduction. A horse, blessed with eyes and a brain as well as big muscles, will partially steer itself, or at least will do better than a bicycle or car in this respect. But teams still needed teamsters, if only to read signposts.

The steam railway locomotive revolutionized land transport in the nineteenth century, quickly replacing horse traction for long overland journeys. Guidance by steel track made steam power the great concentrator, bringing grain across prairies to the metropolis. Railway networks are action programs frozen in steel, their degrees of freedom discrete and finite, encoded in points. Point settings, signaled by telegraph, coordinate the orderly movement of millions of tons according to precise published timetables. Human work did not all lend itself so readily to mechanization.

CHAPTER 2

PROBLEMATIZING INFORMATION

Cockshott

We have suggested that doing purposeful productive labour typically reduces entropy. Such entropy-reducing work requires information in two forms, an action plan or capacity for behaviour, and information coming in from the senses to monitor the implementation of the action plan. Productive labour also involves work in Watt's sense of overcoming physical resistance. As such it consumes energy and produces an entropy increase in the environment that more than compensates for the entropy reduction effected in the object of labour. We have also seen how Maxwell postulated that it should be possible to reduce the entropy of a gas if there existed a being small enough to sort molecules. In this case the being would be using information from its senses, and in its action plan, to produce an entropy reduction in the gas with no corresponding increase elsewhere. Up to now we have not rigorously defined what we mean by information. Once this is done, we shall see the deeply hidden flaw in Maxwell's argument.

2.1 THE SHANNON–WEAVER CONCEPT OF INFORMATION

The philosopher Bachelard (1970) argues that the formation of a science is characterized by what he calls an 'epistemological break', which demarcates the language and ideas of the science from the pre-scientific discourses that appeared to deal with the same subject matter. Appeared to deal with the same subject, but did not really do so. For one of the characteristics of an epistemological break is a change in the *problematic*, which means roughly, the set of questions to which the science provides answers. With the establishment of a science the conceptual terrain shifts both in terms of the answers given and, more importantly, in terms of the questions that researchers regard as relevant.

The epistemological break that established information theory as a science occurred in the middle of the last century and is closely associated with the name of Claude Shannon. We saw how Watt, seeking to improve the efficiency of steam pumps, contributed not only to an industrial revolution, but to a scientific revolution when he asked questions about the relationship between work and heat. From this problematic were born both a convenient source of power, and our understanding of the laws of thermodynamics. Shannon's revolution also came from asking new questions, and asking them in a very practical engineering context. Shannon was a telephone engineer working for Bell Laboratories and he was concerned with determining the capacity of a telephone or telegraph line to transmit information. Watt formalized the concepts of power and work in an attempt to measure the efficiency of engines. Shannon (1948) formalized the concept of information through trying to measure the efficiency of communications equipment. Practice and its problems lead to some of the most interesting truths.

To measure the transmission of information over a telephone line, some definite unit of measurement is needed, otherwise the capacity of lines of different quality cannot be meaningfully compared. According to Shannon the information content of a message is a function of how surprised we are by it. The less probable a message the more information it contains. Suppose that each morning the radio news told us "We are glad to announce that the Prime Minister is fit and well." We would soon get fed up. Who would call this news? It conveys almost no information. "Reports are just reaching us of the assassination of the Prime Minister." That is news. That is information. That is surprising.

A daily bulletin telling us whether or not the Prime Minister was alive would usually tell us nothing, then on one day only would give us some useful information. Leaving aside the circumstances of his death, if an announcement were to be made each morning, there would two possible messages

0 'The P.M. lives'

1 'The P.M. is dead'

Binary Code	Length	Meaning	Probability
0	1	False, False	$\frac{4}{9}$
10	2	False, True	$\frac{2}{9}$
110	3	True, False	$\frac{2}{9}$
111	3	True, True	$\frac{1}{9}$

Table 2.1: A possible code for transmitting messages that are true $\frac{1}{3}$ of the time

If such messages were being sent over the sort of telegraph system that Shannon was concerned with, one could encode them as the presence or absence of a short electrical pulse, as a binary digit or 'bit' in the widely understood sense of the word. Shannon defines a bit more formally as the amount of information required for the receiver of the message to decide between two equally probable outcomes. For example, a sequence of tosses of a fair coin can be encoded in 1 bit per toss, such that heads are 1 and tails 0.

What Shannon says is that if we are sending a stream of 0 or 1 messages affirming or denying some proposition, then unless the truth and falsity of the proposition are equally likely these 0s and 1s contain less than one bit of information each. In that case there will be a more economical way of sending the messages. The trick is not to send a message of equal length regardless of its content, but to devise a system where the more probable message-content gets a shorter code.

For example, suppose the messages are the answer to a question which we know a priori will be true one time in every three messages. Since the two possibilities are not equally likely Shannon says there will be a more efficient way of encoding the stream of messages than simply sending a 0 if the answer is false and a 1 if the answer is true. Consider the code shown in Table 2.1. Instead of sending each message individually we package the messages into pairs, and use between one and three binary digits to encode the 4 possible pairs of messages. Note that the shortest code goes to the most probable message, namely the sequence of two 'False' answers with probability $\frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$. The codes are set up in such a way that they can be uniquely decoded at the receiving end. For instance, suppose the sequence

‘110100’ is received: checking the Table, we can see that this can only be parsed as 110, 10, 0, or True, False, False, True, False, False.

To find the mean number of digits required to encode two messages we multiply the length of the codes for the message-pairs by their respective probabilities:

$$\frac{4}{9}+2\times\frac{2}{9}+3\times\frac{2}{9}+3\times\frac{1}{9}=1\frac{8}{9}\approx 1.889 \tag{2.1}$$

which is less than two digits.

Shannon came up with a formula which gives the shortest possible encoding for a stream of distinct messages, given the probabilities of their individual occurrences.

$$H=-\sum_{i=1}^{n} p_i \log_2 p_i \tag{2.2}$$

The mean information content of an ensemble of messages is obtained by weighting the log of the probability of each message by the probability of that message. He showed that no encoding of messages in 1s and 0s could be shorter than this. The formula gave him an irreducible minimum of the number of bits needed to transmit a message stream: this minimum was, he said, the real information content of the stream. Using Shannon’s formula we can calculate the information content of the data stream encoded in the example above.

$$-\frac{4}{9}\times\log_2\frac{4}{9}-\frac{2}{9}\times\log_2\frac{2}{9}-\frac{2}{9}\times\log_2\frac{2}{9}-\frac{1}{9}\times\log_2\frac{1}{9}\approx 1.837 \tag{2.3}$$

Since our code used $1\frac{8}{9}\approx 1.889$ bits for each pair of messages, we see that in principle a better code may exist.

In his 1948 article Shannon notes:

> Quantities of the form $H=-\sum_{i=1}^{n} p_i \log p_i$ play a central role in information theory as measures of information, choice and uncertainty. The form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of a system being in cell i of its phase space. H is then, for example the H in Boltzmann’s famous H theorem. We shall call $H=-\sum p_i \log p_i$ the entropy of the set of probabilities $p_1,\dots,p_n$.

input from		output to		Comment
A	B	A	B	
No	No	No	No	No molecules involved
No	Yes	No	Yes	Door shut, molecule bounces back to B
Yes	No	No	Yes	Molecule goes from A to B
Yes	Yes	Yes	Yes	Molecules bounce off one another

Table 2.2: The action plan of the demon

Shannon thus discovers that his measure of information is the same as Boltzmann's measure of entropy and decides that entropy and information are the same thing. Armed with this realization we can go back to the problem left to us by Maxwell. Could a sufficiently tiny entity violate the laws of thermodynamics by systematically sorting molecules?

Physicists have concluded that it is not possible. Szilard (1964), for example, pointed out that to decide which molecules to let through, the demon must measure their speed. He showed that these measurements (which would entail bouncing photons off the molecules) would use up more energy than was gained. Maxwell's demon, to vary the theological metaphor, was a *deus ex machina* (like Newton's God), able to know by immaterial means; Szilard's advance was to emphasize that knowledge or information is physical and can only come about by physical means. Brillouin (1951) extended Szilard's analysis by pointing out that at a uniform temperature, black body radiation in the cavity would be uniform in all directions, preventing the demon from seeing molecules unless he had an additional source of light (and hence energy input).

It is possible, however, to build an automaton that acts as a Maxwell demon for a lattice gas. As we said before such gases can be simulated in software, or in hardware (see Figure 2.1), with each gas cell represented by a rectangular area of silicon and the paths taken by the molecules represented by wires. In such a system the demon himself is an automaton, a logic circuit, as in Figure 2.2. A circuit like this really does work: it transfers virtual gas molecules from chamber A to chamber B. Why does this work in apparent conflict with the laws of thermodynamics?

The behaviour of the demon is summarized in Table 2.2. Notice that while there are 4 possible combinations of input conditions, there are only

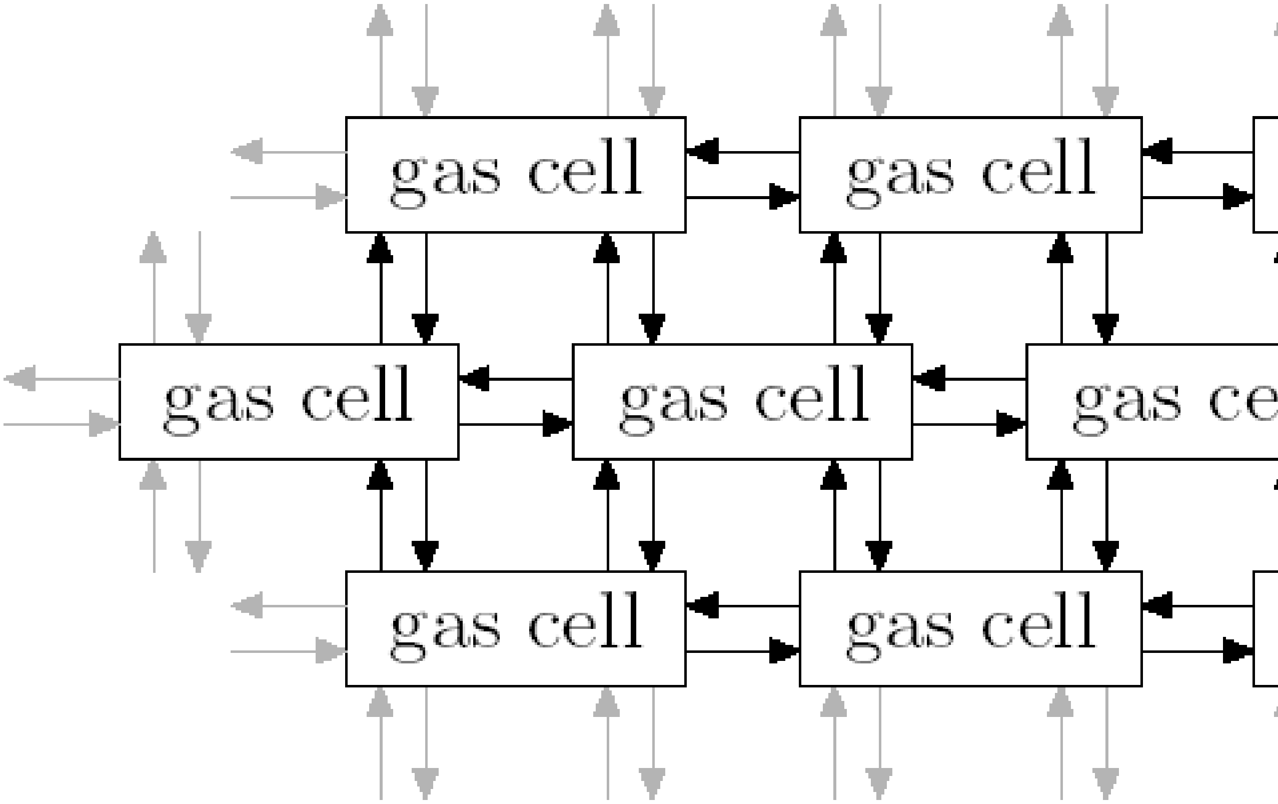


Figure 2.1: A lattice gas can be built in electronic hardware: each gas cell is represented by a rectangular area of silicon and the paths taken by the molecules are represented by wires.

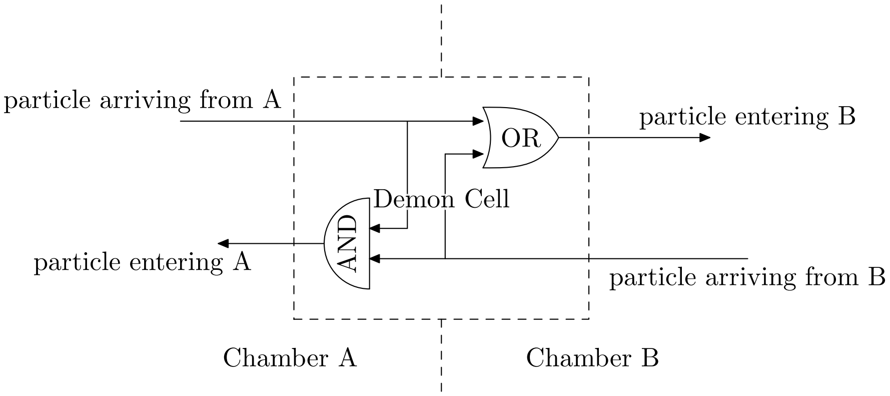


Figure 2.2: In a lattice gas, Maxwell's demon can be implemented with this logic circuit.

3 combinations of output conditions. This implies that we are moving from a system with a higher thermodynamic weight to one with a lower weight, which is what we would expect for an entropy-reducing machine. Just how much it reduces entropy depends on the probabilities of occurrence of incoming particles from each side.

Suppose that the system is in equilibrium and that the probability of occurrence of a particle on the incoming paths on each side is 50 percent in each time interval. In that case each of the 4 possible input configurations in Table 2.2 is equiprobable and has an entropy of 2 bits= $\log_2 4$. Applying Shannon's formula (2.2) to the output configurations we get

$$\frac{1}{4}\log_2 4+\frac{1}{2}\log_2 2\frac{1}{4}\log_2 4=\frac{1}{4}\times 2+\frac{1}{2}\times 1+\frac{1}{4}\times 2=1\frac{1}{2} \tag{2.4}$$

an entropy reduction of half a bit per time step. The key to how this can happen lies in the nature of the components used, logic gates for the functions AND and OR.

Landauer (1961) pointed out that any irreversible logic gate must destroy encoded information and in the process must dissipate heat. An irreversible logic gate is one whose inputs can't be determined from an examination of their outputs. Consider gates with two inputs and one output, such as the AND and OR gates whose truth functions are tabulated in Table 2.3. Roughly speaking they take two bits in and generate one bit out, thus de-

x	y	x AND y	x OR y
false	false	false	false
false	true	false	true
true	false	false	true
true	true	true	true

Table 2.3: Tabulation of the functions x AND y, x OR y

stroying information within the system defined by the lines connecting the gates. Landauer argues that the lost information, i.e., the entropy reduction within the logic circuit, results in an increase in the entropy of the environment. Each time a logic circuit of this type operates, the lost internal entropy shows up as waste heat. By applying Shannon's formula (2.2) to the output of the AND gate we get the following:

Output	p_i	$-p_i \log_2 p_i$
false	$\frac{3}{4}$	≈ 0.311
true	$\frac{1}{4}$	0.5
	1	0.811

The output has an entropy of *less* than one bit. Given that 2 bits of information went into the gate, a total of 1.189 bits are lost in processing the inputs. Since the probability structure of OR gates is the same, a similar information loss occurs going through these.

2.1.1 Information engines as heat engines

Boltzmann's constant (see equation 1.2) has the dimension joules per log-state degree Kelvin. Landauer saw that one can use this constant to convert entropy in Shannon's form, measured in log-states, to energy. The equation he established is

$$e = \ln(2)ktb \tag{2.5}$$

e represents the energy-equivalent, t is temperature in degrees Kelvin, b is the number of bits, and k is Boltzmann's constant, which has a value of about 1.38×10^{-23} joules per degree Kelvin. The remaining term in the conversion is the natural log (ln) of 2, to get us from the natural logarithms used by Boltzmann to the base-2 logarithms used in Shannon's information theory.

Using Landauer's equation we can calculate the heat energy, e_{AND}, generated by a single operation of an AND gate, in which 1.189 bits are lost:

$$e_{\mathrm{AND}} = 1.189 \ln(2) kt$$

At room temperature, or roughly 300° Kelvin, this is 3.4×10^{-21} joules each time the gate switches. This is a very, very small quantity of energy which is at present mainly of theoretical interest. What it represents is the theoretical minimal energy cost of operating a two-input irreversible logic gate.

Now look again at the demon cell in Figure 2.2, which has a pair of input logic gates. The process of deciding whether to open or close the trapdoor must consume certain minimum Landauer-energy. The energy consumed by the logical decision to open or close the barrier makes the demon ineffective as a power source.

Watt started out investigating how to convert heat into work efficiently; he was concerned with minimizing the heat wasted from his engines. Since Landauer we have known that information processing, too, must dissipate heat, and that information processing engines are ultimately constrained by the same laws of thermodynamics as steam engines. We can calculate the thermodynamic efficiency of an information processing machine just as we calculate the efficiency of a steam engine. If a processor chip of the year 2000 had roughly 6 million gates and was clocked at 600Mhz, its dissipation of Landauer energy would then be $(600 \times 10^6) \times (6 \times 10^6) \times (3.4 \times 10^{-21}) = 16.3\mu$w, or 16 millionths of a watt. This is insignificant relative to the electrical power consumption of the chip, which would be of the order of 20 watts. It implies a thermodynamic efficiency of only around 0.0001%. As a point of comparison, steam engines prior to Watt had an efficiency of about 0.5%. The steam turbines in modern power stations convert around 40% of the heat used into useful work. Two centuries of development raised the efficiency of steam power by a factor of about 100.

In thermodynamic terms a Pentium processor looks pretty poor compared to an 18th century steam engine: the steam engine was 500 times more efficient! But if compare a Pentium with the Manchester Mk1, the first electronic stored program computer[1], we get a different perspective.

---

[1] See:Lavington (1978),and Lavington (1980)

The Pentium has at least a thousand times as many logic gates, has a switching speed a thousand times greater and uses about one hundredth as much electrical power as the venerable valve-based Mk1. In terms of thermal efficiency, this represents an improvement factor of 100,000,000 in fifty years. If improvements in heat engine design from Watt to Parsons powered the first two industrial revolutions, the third has benefited from an exponential growth in efficiency that was sixteen times as rapid.[2]

We know from Carnot's theory that there is little further room for improvement in heat engines. Most of the feasible gains in their efficiency came easily to pioneers like Watt and Trevithick. We're now left with marginal improvements, such as the ceramic rotor blades that allow turbine operating temperatures to creep up. In the case of computers too, efficiency gains will eventually become harder to attain. There is still, to quote Feynman, "plenty of room at the bottom". That is, there is mileage yet in miniaturization. We have room for about a million-fold improvement before computers get to where turbines now are. However, as we take into account the growing speed and complexity of computers, the thermodynamic constraint on data processing will come to be of significance. On the one hand, if the efficiency of switching devices continues to grow at its current rate, they will be at close to 100% in about 30 years. On the other hand, as computers get smaller and faster the job of getting rid of the Landauer-energy, thrown out as waste heat, will get harder. In the 27 years following the invention of the microprocessor the number of gates per chip rose by a factor of some 3000. Processor speeds increased about 600-fold over the same period. Table 2.4 projects this rate of growth into the next century.

From being insignificant now, Landauer heat dissipation becomes prohibitive in about 30 years. A microprocessor putting out several kilowatts, as much as several electric heaters, is not a practical proposition. There is a time limit on the current exponential growth in computing power.

That is not to say that computer technology will stagnate in 40 years. Landauer's equation (2.5) has a free variable in *temperature*. If the computer is super-cooled, its heat dissipation falls. But once we're in that game the rate of improvement in computer performance comes to be limited by improvements in refrigeration technology, and these are unlikely to be so dramatic.

---

[2] Heat engine efficiency improved about ten-fold per century. Information engines have been improving at a factor of about 10^{16} per century.

year	*gates*	*clockspeed*	*landauer watts*
2000	8×10^6	600Mhz	16.3μw
2005	3.4×10^7	1.9Ghz	230μw
2010	1.5×10^8	6.4Ghz	3.24mw
2015	6.4×10^8	21Ghz	45.7mw
2020	2.8×10^9	68Ghz	643mw
2025	1.2×10^{10}	224Ghz	9.06w
2030	5.1×10^{10}	733Ghz	128w
2035	2.2×10^{11}	2.4Thz	1.80Kw
2040	9.5×10^{11}	7.8Thz	25.4Kw

Table 2.4: Projected Landauer heat dissipation in 21st century computers operating at 300° Kelvin.

2.2 ENTROPY REDUCTIONS IN ACTION PROGRAMS

Maxwell's demon cannot exist for real gases, but it can for lattice gases. If the demon really existed, he would reduce the laws of thermodynamics to the status of an anthropocentric projection onto reality. Lattice-gas devils, on the other hand, are not a threat to physics. They reduce the entropy of the gas, but only because they use logic gates with an external source of power. Nonetheless, their structure suggests something important. The demon reduces the entropy of the gas thanks to an action program which has four possible input states and only three possible output states.

We would suggest that this is not accidental: it would seem that *all production processes that produce local reductions in entropy are guided by an entropy-reducing action program*. Consider the bee once again, this time in its capacity as forager. In Maxwell's original proposal, the demon used its refined perception to extract energy from chaos. In reality a bee uses its eyes to enable it to extract energy from flowers. Were bees unable to see or smell flowers, their energy would be expended in aimless wandering followed by starvation. The bee uses information from its senses to achieve what, from its local viewpoint, is a reduction in entropy—the maintenance of homeostasis—albeit at a cost to the rest of the universe. To achieve this it requires a nervous system that performs entropy reduction on the input data coming into its visual receptors. At any given instant the bee's compound eyes are receiving stimuli from the environment. The number of possible different combinations of such stimuli is vastly greater than the number of instantaneous behavioural responses that it has while in flight—the modulation of the beat strength of a small number of thoracic muscles. In selecting one appropriate behavioural response out of a small repertoire, in response to a relatively large quantity of information arriving at its eyes, the bee's nervous system functions in the same sort of way as the AND gate in the demon-automaton of Figure 2.2. Having fewer possible outputs than inputs, it discards information and reduces entropy.

2.3 ALTERNATIVE VIEWS OF INFORMATION

We have come across two approaches to the idea of entropy so far, deriving from classical thermodynamics and Shannon's communication theory respectively. From the 1960s onwards a third version has developed: that of computational complexity. Where classical concepts of entropy derived from mechanical engineering, and Shannon's concept from telecommunications engineering, the latest comes from computer science. The key concepts appear to have been independently developed by Chaitin (1999) in the US and Kolmogorov in Russia. Their presentation, while not contradicting what Shannon taught, gives new insights that are particularly helpful when we come to consider the role that information flows play in mass production industries.

2.3.1 The Chaitin–Kolmogorov concept of information

Chaitin's algorithmic information theory defines the information content of a number to be the length of the shortest computer program capable of generating it. This introduction of numbers is a slight shift of terrain. Shannon talked about the information content of *messages*. Whereas numbers as such are not messages, all coded messages are numbers. Consider an electronically transmitted message. It will typically be sent as a series of bits, ones and zeros, which can be considered as a binary number. An information theory defined in terms of numbers no longer needs the support of a priori probabilities. Whereas Shannon's theory depended upon the a priori probability of messages, Chaitin dispenses with this support.

As an example of the algorithmic approach consider the Mandelbrot set picture in Figure 2.3. This image is created by a very simple computer

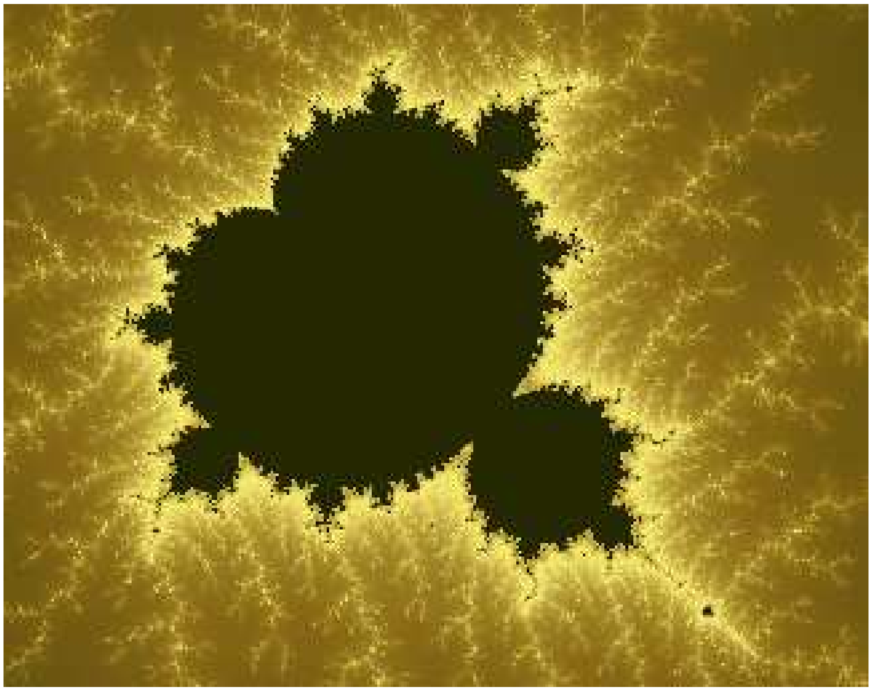


Figure 2.3: The Mandelbrot set, a complex image generated from a tiny amount of information.

program.[3] Although the image file for the picture is large, about 6 million bits, a program to generate it can be written in a few thousand bits. If one wanted to send the picture to someone who had a computer, it would take fewer bits to send the program than to send the picture itself. This only works if both sender and receiver have computers capable of understanding the same program. Chaitin's definition of information has the disadvantage of seeming to make it dependent upon particular brand of computer used. One could not assume that the length of a program to generate the picture would be the same on an Apple as on an IBM.

In principle one could chose any particular computer and fix on it as the standard of measure. Alternatively one could use an abstract computer, much as Watt used an abstract horse. Chaitin follows Watt, using a *gedankenapparat,* the Universal Turing Machine, as his canonical computer. Thus

---

[3] In fact it uses the formula $z = z^2 + c$ where z is a complex number.

he defines the information content of a sequence S as the shortest Turing machine tape that would cause the machine to halt with the sequence S on its output tape.[4]

Randomness and pi

An unsettling result from information theory is that random sequences of digits contain more information than anything else. According to common sense, information is the very opposite of randomness. We feel that information should be associated with order, but Shannon's identification of information and entropy amounts to equating information with *dis*order. To illustrate this let's compare a long random number with π. We know from Shannon that 1 million tosses of a fair coin generates 1 million bits of information. On the other hand, from Chaitin we know that π to a precision of a million bits contains much less than 1 million bits, since the program to compute π can be encoded using much fewer bits. Thus π must contain less information than a random sequence of the same length.

But what do we mean by random? And how can we tell if a number is random? The answer now generally accepted was provided by Andrei Kolmogorov, who defined a random number as *a number for which there exists no formula shorter than itself*. By Chaitin's definition of information a random number is thus incompressible: a random number of n bits must contain n bits of real information.

A fully compressed data sequence is indistinguishable from a random sequence of 0s and 1s. This not only follows directly from Kolmogorov and Chaitin's results but also from Shannon, from whom we have the result that for each bit of the stream to have maximal information it must mimic the tossing of a fair coin: be unpredictable, random.

We have a paradox: one million digits of π are more valuable and more useful than one million random bits. But they contain less information. They are more valuable because they are harder to come by. They are more useful because a host of other formulae use π. They contain less information because each and every digit of π was determined, before we started

---

[4] There is, in principle, no algorithm for determining the shortest Turing Machine tape for an arbitrary sequence. $3 \div 7$ is a rule of arithmetic, an algorithm that generates the sequence 0.428571428571. So this sequence is presumably less random than 0.328571428771 (we changed two digits). But in practice we can never be sure.

calculating it, by π's formula. Thus in a sense the entire expansion of π is redundant if we have its formula. Valuable objects are generally redundant. We thus have three concepts that we must distinguish with respect to sequences: their information content, their value, and their utility.

Concept	*Meaning*
Information	Length of program to compute the sequence.
Value	Cycles it takes to compute the sequence.
Utility	The uses to which the sequence can be put.

The *value* of a sequence is measured by how hard we must work to get it. π is valuable because it is so costly to calculate. We can measure the cost by the number of machine cycles a computer would have to go through to generate it.[5] As with information content, this definition is dependent upon what we take as our standard computer. A more advanced computer can perform a given calculation in fewer clock cycles than a more primitive one. For theoretical purposes any Universal computer will do. Information theorists typically use machine cycles of the Universal Turing Machine (UTM) for their standard of work. We will follow them in defining the information content of a sequence in terms of the length of the UTM program that generates it, and the value of a sequence in terms of the UTM cycles to compute it.

Now the UTM is an imaginary machine, a thought experiment, living in the platonist ideal world of the mathematician. Its toils are imaginary, consuming neither seconds nor ergs; its effort is measured in abstract cycles. But any physical computer existing in our material world runs in real time, and needs a power supply. Valuable numbers—tomorrow's temperature for example—whose computation requires large number of cycles on the Met Office super computers, take real time and energy to produce. The time depends on clock speed, and the energy depends on the computer's thermodynamic efficiency.[6] If we abstract from changes in computer technology, information value in UTM cycles is an indication of the thermodynamic cost

---

[5] We are identifying the value of a sequence with what **?**althusser70 calls its logical depth. The homology with Adam Smith's definition of value should be evident.

[6] The UTM plays, for computational complexity theory, the role of Marx's "labour of average skill and intensity" in the economic theory of value. Improvements in computer technology are analogous to changes in the skill of the worker.

of producing information. It measures how much the entropy of the rest of the universe must rise to produce the information.[7]

Having traced the conceptual thread of entropy from Boltzmann through Shannon to Chaitin, it is worth taking stock and asking ourselves if Chaitin's definition of entropy still makes sense in terms of Boltzmann's definition. To do this we need to move from numbers to their physical representation. A material system can represent a range of numbers if it has sufficient well-defined states to encode the range. Will a physical system in a state whose number has, according to Chaitin, a low entropy, have a low entropy according to classical statistical mechanics?[8]

What we will give is not a proof, but at least a plausible argument that this will be true. As a *gedanken* experiment we will consider a picture of the Mandelbrot set rendered on digital paper. Digital paper is a proposed display medium made of thin films of white plastic. In the upper layer of the plastic there is a mass of small bubbles of oil, in the middle of each of which floats a tiny ball. One side of the ball is white and the other black. Embedded within the ball is a magnetized ferrite crystal with its North pole pointing towards the black end.[9] If the paper is embedded in an appropriate magnetic field all of the balls can be forced to rotate to have their white half uppermost, making the paper appear white. Applying a South magnetic pole to a spot on the paper will leave a black mark where the balls have rotated to expose their dark half. When it is passed through an appropriate magnetic printer, patterns can be drawn. A sheet of digital paper with a Mandelbrot set image on it nicely straddles the boundary between an industrial product and a number or information structure.

According to algorithmic information theory, the Mandelbrot set image represents a relatively low entropy state, since the length of the program to compute it contains fewer bits than the image. Does it also represent a low entropy state in statistical mechanics?

The second law of thermodynamics states that the entropy of a closed system is non-decreasing. So we would expect that a picture of the Mandelbrot state drawn on digital paper would tend to change into some other

---

[7] This is what Norretranders (1998) calls *exformation*.

[8] We need this step if we are to apply Chaitin's theory to labour processes that produce real physical commodities. We need an epicurean not a platonist theory.

[9] We are giving a somewhat stylized account of digital paper for the purposes of this argument.

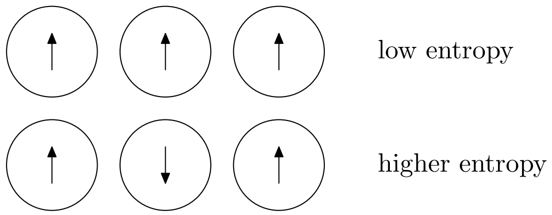


Figure 2.4: Configurations of parallel poles are unstable and tend to evolve towards the anti-parallel configuration.

picture whose state would represent a higher entropy level. In fact there are good physical reasons why this will take place. If a local area is all white or all black, the magnetic poles are aligned as shown in the top of Figure 2.4. In this configuration the like poles tend to repel one another, and over time some of the poles will tend to flip to the configuration shown in the bottom half of the diagram.

The rate at which this occurs depends upon the temperature, the viscosity of the fluid in which the balls are suspended, and so on, but in the long run entropy will take hold. The image will gradually degrade to a higher entropy state, both in thermodynamic terms and in algorithmic terms. The program necessary to produce the degraded picture is bound to be longer than the program that produced the pristine one. Hence thermodynamic and algorithmic entropy measure the same scale.

The example we have given is stylized but the thermodynamic degradation of digital information is not hypothetical. Magnetic tape libraries have a finite life because of just this sort of flipping of the magnetized domains on which the information is stored.

2.4 RANDOMNESS AND COMPRESSIBILITY

You may find at this point that reason in you rebels at the idea that information content and randomness are equivalent. But this is what information theory teaches us, so it is worth considering and trying to resolve several apparent paradoxes that arise from information theory.

Kolmogorov identifies the randomness of a number with its incompressibility (via his “no shorter formula” proposition). There seems to be a contradiction—or at least a strong tension—between this conception of randomness as a property *of a number* and the "ordinary" conception of randomness as a property of a *mechanism for generating numbers*. (As in the statisticians' talk of a "random variable" as a variable whose values are determined by the outcome of a "random experiment".)

2.4.1 Random numbers contain non-random ones

To expose the tension, consider a random number generator (RNG). Suppose it's a true quantum RNG, set to produce a series of uniformly distributed ten-digit numbers. The standard definition of randomness would be that every ten-digit number is produced with equal probability (and the drawings are independent, previous drawings do not affect subsequent ones). Thus if we leave our RNG running for a while, it's bound to produce numbers such as 1111111111 and 0123456789. But these are not "random numbers" on the Kolmogorov definition. The paradox is then that the output of a random number generator (i.e. a device that generates numbers at random) is bound to include nonrandom numbers.

In these examples we have non-random sub-sequences of the output of the RNG. This is not a valid objection, as we have to take the entire output of the RNG up to some large number of digits, in order to obtain these sub sequences that appear non-random. So these short subsequences are not produced by the random number generator, but, strictly speaking, by a Turing machine program that is a prefix to the random number generator, and which searches for patterns like 1111111111111 in the output of the RNG. The Algorithmic Information Theory approach to this would be to add the information content of the program which generated the sequence to the program which selected for the "non-random" sub sequences.

2.4.2 Randomness of a number as opposed to of a generator.

In standard statistical parlance it doesn't really make sense to talk of a random *number* as such, as opposed to a random *variable* or a random number *generator* (where the adjective "random" attaches to the generator, i.e. it's a random generator of numbers rather than a generator of random numbers). Kolmogorov defines "random number", in a way that seems to conflict with the standard view.

But this is just a divergence between what we commonly understand as a number in statistics and how a number is defined in computational complexity theory. By number the Algorithmic Information Theory just means a sequence of digits. Since any sub-sequence of digits is also a number, formalisations in terms of numbers also provide for formalisation in terms of finite sequences of numbers. Thus a sufficiently large number can be treated as a generator of smaller numbers.

2.5 INFORMATION AND RANDOMNESS

To get at the second paradox we will report a little experiment. We have an ASCII file of the first eleven chapters of Ricardo's *Principles*: it's 262899 bytes. We ran the `bzip2` compressor on it and the resulting file was 61193 bytes, a bit less than quarter of the size. Suppose for the sake of argument that `bzip2`[10] is a perfect byte-stream compressor: in that case the 61193 bytes represent the incompressible content of the Ricardo chapters. They measure the true information content of the larger file, which contains a good deal of redundancy. That idea seems fair enough.

The second part of the experiment was to generate another file of 262899 bytes of printable ASCII characters (the same length as Ricardo), this time using a random number generator[11], and running `bzip2` on the resulting file produced a compression to slightly over 80 percent of the original size.

The first question is why we get any compression at all on the "random" ASCII files?

Our bytes are printable characters. These are drawn from a subset of the possible byte values[12], and as such all, the possible byte values are not equiprobable. Thus the stream is compressible.

The next question concerns the information content of the various files. Suppose we have already accepted the idea that the 61193 bytes of bzipped Ricardo represent the irreducible information content of the original Ricardo file. Then by the same token it seems the 218200 (or so) bytes of bzipped rubbish from the random number generator represent the true information content of the (pseudo-)random byte stream. The rubbish contains almost four times as much information as the Ricardo. This is very hard to swallow.

The point here is that standard data compression programs use certain fixed algorithms to compress files. In this case an algorithm known as

---

[10] A publicly available data compression program.

[11] the `rand()` function in the GNU C library

[12] There are $256 = 2^8$ possible values for 8 bit bytes.

Lempel-Ziv[13] is used. Lempel-Ziv does not know how to obtain the maximum compression of the stream—which would be an encoding of the random number generating program. One can not make a general purpose compressor that will obtain the maximum possible compression of any stream. One can only produce programs that do a good job on a large variety of cases.

We make the distinction between information as such and utility, and in those terms it's clear that the Ricardo is of much greater utility than the rubbish. Even so, intuition rebels at the idea that the rubbish carries *any* information. We have a conception of "useless information" alright, but it seems doubtful that a random byte stream satisfies the ordinary definition of useless information. In ordinary language information has to be *about* something; and it's useless if it's about something that is of no interest. For me, the weekly guide to Cable TV programming may contain useless information. It's of no more interest to me than a random byte stream. Nonetheless, I recognize that it does contain (quite a lot of) information; it is certainly about something.

In classical political economy use-value is neither the measure nor the determinant of value, but nonetheless it's a *necessary condition* of value. If a product has no use-value for anyone then it has no value either, regardless of how much labour time was required for its production. Can we say that the utility of a message is not the measure of its information content, but if a "message" is of no potential use to anyone (is not about anything) then it carries no information, regardless of its incompressible length?

No. Information exists even if it is not useful. Take the case of hieroglyphs prior to the discovery of the Rosetta stone.[14] They were meaningless until that was discovered, useless in other words. Once it was discovered they became useful historical documents. Their information content was not created *ex-nihilo* by Champollion, but must have been there all along. Similarly, the works of Ricardo in Chinese contain no information to me, are of no use to me, but they still contain information.

---

[13] Ziv and Lempel (1978).

[14] The inscription on the Rosetta Stone, is a decree for King Ptolemy V Epiphanes dating from March 196 BC. It is repeated in hieroglyphs, demotic and Greek. By using the Greek section as a 'key' scholars realised that hieroglyphs were not ideograms, but that they represented a language. Jean-François Champollion (AD 1790-1832), realised in 1822 that they represented a language which was the ancestor of Coptic.

In the end, whether information is useful to us concerns our selfish thermodynamic concerns. Does it enable us to change the world in a way that saves us work or produces us energy?

This is an anthropospective projection. It is not a property of the information but a property of the user of the information, which is cast back onto the information itself. Information theory in its epistemological break, had to divest itself of anthropospective views, just as astronomy and biology had to.

The "digital paper" example suggests one further paradox on the issue here. Let's go back to the ASCII Ricardo. Its incompressible length was (according to bzip2) 61193 bytes. Now suppose the hard drive is exposed to radiation that results in random bit-flipping, which changes some of the bytes in the Ricardo file. At some later point we try compressing the file again. We find that it won't compress as well as before. Its information content has increased due to the random mutation of bytes! Meanwhile, of course, its value as representation of what Ricardo said is eroding. Is it possible to make any sense of this?

Yes. The degraded work contains more information since to reconstruct it one would need to know the trajectories of the cosmic rays which degraded the stored copy, plus the original copy. We may not be interested in the paths of these cosmic rays,[15] but it is additional information, provided to us courtesy of the Second Law.

---

[15] In other circumstances, archeological dating for example, such radiation damage gives us useful information.

CHAPTER 3

LABOUR PRODUCTIVITY
Cockshott

> Those who possess rank in a manufacturing country, can scarcely be excused if they are entirely ignorant of principles, whose development has produced its greatness. The possessors of wealth can scarcely be indifferent to processes which, nearly or remotely have been the fertile source of their possessions. Those who enjoy leisure can scarcely find a more interesting and instructive pursuit than the examination of the workshops of their own country, which contain within them a rich mine of knowledge, too generally neglected by the wealthier classes. (Babbage (1832), Preface.)

3.1 RAISING PRODUCTION IN GENERAL

In this chapter we examine the means by which labour productivity increases over time. The level of our analysis here is essentially technical. We are looking at productivity in physical terms rather than in value terms. We are not at this point interested in how many Euros' or dollars' worth of output each worker produces per hour. Instead we are looking at physical production—tons of steel, meters of rope, numbers of cars, and so on.

This concentration on physical productivity means that our focus is not only limited to technical considerations, it is also narrow, looking at one industry at a time. We cannot yet look at the economy in general since, by abstracting from prices or other means of valuation, we have deprived ourselves of any scale by which we could measure the national product. The total product of the economy comprises a heterogenous mixture of goods.[1] For the moment we will consider one product at a time, and the natural units of that product will provide us with our scale.

We are primarily interested in the flow of product per unit time—17 million tons of steel per year, 15 meters of cloth per hour. We are also interested in product flow per unit time per worker since this is the dimension along which the wealth of society in general increases.[2]

There are three fundamental ways by which the flow through any production process can be increased.

(1) Accelerating the production cycle.
(2) Parallelizing production.
(3) Eliminating wasted effort.

These basic methods apply whether the production process is human or animal, mechanical or biological, carried out by men, bees or robots. Examination of them will provide the main substance of the chapter.

3.1.1 Entropy analysis

Before going into the above-mentioned methods of increasing productivity, we shall first extend our analysis of information and entropy to look at the changes in entropy that take place in during industrial production.

We have already considered the thought experiment of digital paper. We showed that if you wrote text on it, although this text represented information, it contained much less information than the paper potentially could. If we transfer what we have learned from this example to ordinary paper and the process of producing a book we see that the production process encompasses two opposite phases.

First, we have the production of the paper. This is an entropy-*reducing* process. The blank sheets of paper obviously have low information content with respect to human language, but they also constitute a low entropy state with respect to the raw material. In a sheet of paper the cellulose fibres

---

[1] Technically speaking, it is a *vector*, a list of numbers: $[x$ tons of steel, y cars, z barrels of oil, ... $]$. Vectors are a means of describing positions in multi-dimensional space. To get an unambiguous measure of changes in production you need a scale to measure the changes, a *scalar* quantity like \$$w$.

[2] Abstracting for now from the division of this wealth between the different classes in society.

are constrained in both orientation and position. With regard to orientation, the fibres must lie in a plane rather than being free to take up any angle. This implies a reduction in the volume of state space that the fibres occupy, and thus, from Boltzmann, a corresponding reduction in entropy. The fibres are also constrained to exist within a small volume a few hundredths of a millimeter thick. This restriction in physical space obviously entails a smaller entropy, as shown in our discussions of Maxwell's daemon.

Second, there is the writing of the text—whether by hand, as in the distant past, or using a printing press. This is an entropy-*increasing* process. Imagine that the text to be printed exists as binary data in a file on disk, encoded using ASCII or UNICODE.[3] Clearly the book contains this information, since by sending the book in the post to someone we enable them to recreate the relevant binary file. Thus, by the equivalence of information and entropy, we have increased the entropy of the book relative to the blank sheets of paper. For another perspective, consider the fact that while all blank sheets are alike, printed sheets can be different. The number of possible different pages that can be printed is so huge as to dwarf the concept of astronomically large.[4] Since entropy is logarithmically related to the number of possible states, the increase in the number of possible states implies a rise in entropy.[5]

In the first phase a low-entropy material is created; in the second phase the entropy of this material is increased in a controlled way. Initially *natural* information is removed; subsequently *anthropic* or human-created information is added. The natural information removed in the first stage is of no interest to us, while that added in the second stage is dictated by our concerns.

---

[3] ASCII is the American Standard Code for Information Interchange, a code which uses 7 bits to represent each letter or symbol. It is restricted to the characters appearing on US typewriters. UNICODE is a newer 16-bit code that can represent every letter or glyph used in any of the world's languages, including ideographic scripts like those of China and Japan.

[4] If we allow 40 lines of 60 characters, with these characters drawn from a lexicon of all of the world's languages, we have of the order of some 10^{10000} possible printed pages. For comparison, the volume of the universe in terms of the Planck dimension—the quantum of space, 10^{-35}m—is of the order of 10^{210}

[5] It may be objected that while there are a vast number of possible pages that could be printed, we are only interested in printing a particular page. This is true, but it is the particularity of the page that constitutes the added information and thus the added entropy.

The first process—pulping wood, bleaching it, forming it into sheets, drying it—has to use energy to produce the reduction in entropy. Thermodynamics gives any local reduction in entropy its energy price. The second process, increasing entropy, could in principle be done at no energy cost.[6]In practice our technologies are not that efficient. Still, the power consumption of a printworks is a lot lower than that of a paper mill.

Considerable research is currently underway to develop nano-technologies that use self-assembly of microstructures. In this case the increase in entropy that occurs as the structures acquire form and information occurs directly by thermodynamic means, albeit starting off from precisely controlled compositions and temperatures (FIXME citation Witesides 95).

3.1.2 Replicated parts

Consider two books by two different authors, each 200 pages long, printed with the same size of letters. Each has roughly the same amount of information added to the paper in the printing process, but in each case the information is different. On the other hand two copies of a book have the same information added. The added information is what on the one hand differentiates books, and on the other makes replication possible.

It is easy to see the relevance of information theory to the printing industry. Its product, after all, contains information in the everyday as well as the technical sense of that word. Does this approach provide insights into how other production processes function?

For a rather different example, consider the process of producing cloth. The starting material is wool or cotton fibres in a random tangled state. This is first carded to bring the fibres into rough alignment, and then simultaneously twisted and drawn to spin the fibres into yarn. In the yarn both the volume and orientation are sharply reduced. Energy is used to reduce the entropy of the cotton. The weaving of the cotton then increases the entropy by allowing two possible orientations of the fibres at right angles to one another (or more if we take into account the differences in possible weave).

In the case of man-made fibres the extrusion and drawing processes that precede spinning are designed to align the polymer molecules with the axis of the fibres, again this is clearly an entropy reducing process.

---

[6] The not gate proposed for quantum computing is in principle a mechanism by which a process analogous to the printing of information onto blank paper can take place in a reversible and thus non energy consuming way DiVincenzo (1995).

Other industries that use thin, initially flat materials clearly have a lot in common with printing. The manufacture of car body parts from sheet steel, or the garment industry, share the pattern of producing a low entropy raw material and adding information to it. In pressed steel construction, added information is encoded in the shape of the dies used to form the car doors, roof panels etc. We can quantify it using Chaitin's algorithmic information theory, as proportional to the length of the numerically controlled machine tool tape that is used to direct the carving of the die. In the making up of garments from bolts of cloth, the added information comes in the form of the patterns used to cut the cloth.

All of these involve the replication of standard products, dependent on the existence of materialized information in the form of patterns and dies. If steam powered the industrial revolution, the technologies of replication were the key to mass production. The classic example of the importance of accurate replication was in the production of the Colt revolver in the mid 19th century. Prior to Colt establishing his factory the gun trade was dominated by handicraft manufacturing techniques. The different parts of a gun's mechanism were individually made by a gunsmith so that they fitted accurately together. While the components of an individual fowling piece might fit together beautifully, if the hammer were removed from one gun, it would be unlikely to fit accurately into another. Mass production required the use of replicated interchangeable parts. For parts to be interchangeable they must be made to very precise tolerances. This improvement in accuracy of production involves the parts having a lower entropy, occupying a smaller volume of phase space, than the old hand made parts. Again by the equivalence of information and entropy this means that the standardized parts embody less information than the hand made ones. This makes sense; for example it may have been possible to identify the maker of a hand made gun, whereas this would be impossible with a standardized Colt.

In the 19th century, prior to the introduction of numerically controlled machine tools, replicated parts had to be composed of circular and planar elements which could be produced on lathes or milling machines. The limited information content of these can be seen when you consider that in turning a smooth bore gun barrel one only has to specify the inner and outer radii and its length. If an axle and a bearing are being produced separately to fit together, then one wants the uncertainty in the surface of the bearing, given the surface of the axle, to be reduced below a certain limit.

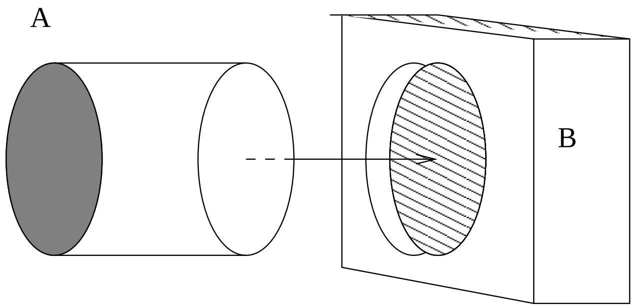


Figure 3.1: When inserting axle A into bearing B we want to minimize the conditional information $H(B|A)$, between B and A.

Information theory analyses this in terms of *conditional entropy*. The Chaitin formulation of this is as follows: the conditional entropy of a character sequence B dependent upon a sequence A, which we write as $H(B|A)$, is given by the length of the shortest prefix Turing machine program that when fed with the program for A will generate B.

How can we apply this concept to our previous mechanical example? Let A stand for an encoding of our axle and B an encoding of our bearing (Fig. 3.1). We divide space up into cells of a fixed size, let us say a $\frac{1}{10}$th of a millimeter on edge. If the space is occupied by metal we denote this with a 1 otherwise we denote it by a 0. We can then use arrays of characters like those in Fig. 3.2, to represent slices through the axle.

According to the Chaitin view, the information content of the cross section through the axle is given not by this array of 1s and 0s but by the shortest program to generate it. Here is an example of a short program that will print out the pattern in Figure 3.2:

```
program circ ;
const
    b: array [boolean ] of char =( ‘1’ , ‘0’ );
    c =20;
    r =18;
var
    a: array [-c ..c ,-c ..c ] of boolean ;
begin
```

```
00000000000000000000000000000000000000000
00000000000000000000000000000000000000000
00000000000000000000000000000000000000000
00000000000000000111111100000000000000000
00000000000000111111111111100000000000000
00000000000011111111111111111000000000000
00000000001111111111111111111110000000000
00000000011111111111111111111111000000000
00000000111111111111111111111111100000000
00000001111111111111111111111111110000000
00000011111111111111111111111111111000000
00000011111111111111111111111111111000000
00000111111111111111111111111111111100000
00000111111111111111111111111111111100000
00001111111111111111111111111111111110000
00001111111111111111111111111111111110000
00001111111111111111111111111111111110000
00011111111111111111111111111111111111000
00011111111111111111111111111111111111000
00011111111111111111111111111111111111000
00011111111111111111111111111111111111000
00011111111111111111111111111111111111000
00011111111111111111111111111111111111000
00011111111111111111111111111111111111000
00001111111111111111111111111111111110000
00001111111111111111111111111111111110000
00001111111111111111111111111111111110000
00000111111111111111111111111111111100000
00000111111111111111111111111111111100000
00000011111111111111111111111111111000000
00000011111111111111111111111111111000000
00000001111111111111111111111111110000000
00000000111111111111111111111111100000000
00000000011111111111111111111111000000000
00000000001111111111111111111110000000000
00000000000011111111111111111000000000000
00000000000000111111111111100000000000000
00000000000000000111111100000000000000000
00000000000000000000000000000000000000000
00000000000000000000000000000000000000000
00000000000000000000000000000000000000000
```

Figure 3.2: A slice through the axle.

```
    $a \leftarrow \sqrt{\iota_0^2+\iota_1^2} < r$;
    write($b_a$);
end .
```

We cannot guarantee to have found the shortest such program.[7] Indeed Chaitin shows that in the general case one can never prove that a given program is the shortest to produce a particular output. But the program is considerably shorter than the pattern that it produces, and with the alteration of the definition of two variables c and r it will generate arbitrary sized circular patterns of 1s in a field of 0s.

Clearly if the bearing exactly fitted the axle the expanded encoding for a slice through the bearing would be an array similar to Fig 3.2 but with 1s and 0s swapped round. This can be produced by a trivial change to the program circ, the addition of a single statement. Indeed all that is required is that the line:

```
    write($b_{\textbf{not}a}$);
```

replaces the line:

```
    write($b_a$);
```

in the program. This must come close to minimizing the conditional entropy of the two parts.

Suppose that the parts were less than perfectly made, so that there were rough spots on the surface of the axle. Figure 3.3 shows a cross section through a pin A' that should be circular, but has a step on it, generated perhaps by improper turning.

Suppose we have our perfectly formed circular hole B, then as before the conditional entropy $H(B|A')$ of the hole and the imperfect pin is much greater than before. Working in the domain of generator programs we would need to add the following lines to the generator of A' to make the bitmap for B:

```
    $a_{r,1} \leftarrow$ false;
    $a_{r,2} \leftarrow$ false;
    $a_{r+1,0} \leftarrow$ false;
    $a_{r,0} \leftarrow$ false;
    $a_{r-1,-1} \leftarrow$ true;
```

---

[7] The program is in Vector Pascal which is fairly concise, see Cockshott and Renfrew (2004).

```
0000001111111111111111111111111111000000
0000001111111111111111111111111111000000
0000000111111111111111111111111110000000
0000000011111111111111111111111100000000
0000000001111111111111111111111000000000
0000000000011111111111111111100000000000
0000000000001111111111111111000000000000
0000000000000001110111111000000000000000
0000000000000000000111000000000000000000
0000000000000000000100000000000000000000
0000000000000000000000000000000000000000
```

Figure 3.3: Part of a pin with a fault on its circumference.

write$(b_{\textbf{not}a})$;

This obviously contains extra information, required to correct the bitmap of A’ to generate that of B. In pre-industrial production, the extras steps in the generator program would translate into additional steps of filing and grinding to make parts fit. The aim of standardized production is to arrive at a situation where independently made parts, derived from a common technical specification, fit together because the conditional information of the mating parts is minimal.

3.2 ACCELERATED PRODUCTION

The most obvious way in which production can be increased is by accelerating the production process itself, by making people and machines work longer and faster.

3.2.1 Longer days

If the working day is increased from 8 hours to 12 while the same tempo of work is maintained, then output per worker will rise by a half. The effect, over a 24-hour day, is analogous to increasing the average intensity of labour. Similarly if a machine is used for 12 hours a day rather than for 8, we have the same effect as if the machine ran 50% faster.

From the standpoint of society as a whole, however, there are real differences. If machines are scarce, an economy can increase its output by using them on a 24-hour shift system. But if a system of three shifts each of 8 hours is used, then three times as many workers are required. Total production will rise threefold, but output per worker remains the same.[8] If on the other hand, the working day is extended to 12 hours, and two shifts are worked, both total output and output per worker go up. This fact encourages employers to lengthen the working day whenever the labour supply is limited. Further, since daily wages rarely rise in proportion to hours, longer hours mean more profit. But the scope for extending the working day is still relatively limited. The maximum feasible working day is perhaps 16–18 hours under the most exploitative conditions, less than a doubling of the pre-industrial working day.

These are small gains compared to those available from technology. No free workers would willingly work such hours. It is rather the fate of slaves, either bonded labourers or wage slaves without access to free trades unions. The working day is ever the inverse reflection of workers' liberty. As workers gain political rights and influence, the working day comes down and other ways have to be found to increase productivity.

3.2.2 Studied movements, intensified labours

Today we think of mass production in terms of the mechanized production line introduced by Henry Ford at the start of the last century. But mass production started much earlier. In the 18th century, before steam or water power were generally applied, mass production took place in manufactories.[9]. In a manufactory, the work was done with hand tools,[10] by groups of workers using a division of labour.

It is a common enough observation that a person's speed improves with practice. Through practice, sequences of muscle movements cease to be under conscious control and become reflexes. We no longer have to think about them. We do them automatically and we do them fast. Early manufacturing based itself upon this principle. Each worker had a simple repetitious

---

[8] The labour required to produce one unit of output may fall slightly, since the depreciation of the machines may not rise proportionately with their intensity of use.

[9] *manu*factory, from manus, the latin for hand.

[10] 'The difference between a tool and a machine is not capable of very precise distinction; nor is it necessary, in a popular explanation of those terms, to limit very strictly their acceptation. A tool is usually more simple than a machine; it is generally used with the hand, while a machine is frequently moved by animal or steam power. The simpler machines are often merely one or more tools placed in a frame, and acted on by a moving power.' *Charles Babbage* Economy of Machinery and Manufactures, 1832, Chap 1.

task, performed largely under reflex control. Production was accelerated both by the increased speed that came from practice, and by eliminating the 'lost time' which would otherwise be spent changing from one task to another. The combination of faster movements and the elimination of wasted time could lead to remarkable improvements in productivity;[11] but the drawbacks of this form of production are obvious. People are, for the duration of the working day, used as automatons, their minds and imaginations rendered redundant. We use the present tense advisedly: plenty of consumer goods in our shopping-malls today come from third world manufacturies where children work as machines.

---

[11] 'To take an example, therefore, from a very trifling manufacture; but one in which the division of labour has been very often taken notice of, the trade of the pin-maker; a workman not educated to this business (which the division of labour has rendered a distinct trade), nor acquainted with the use of the machinery employed in it (to the invention of which the same division of labour has probably given occasion), could scarce, perhaps, with his utmost industry, make one pin in a day, and certainly could not make twenty. But in the way in which this business is now carried on, not only the whole work is a peculiar trade, but it is divided into a number of branches, of which the greater part are likewise peculiar trades. One man draws out the wire, another straights it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving, the head; to make the head requires two or three distinct operations; to put it on is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them. I have seen a small manufactory of this kind where ten men only were employed, and where some of them consequently performed two or three distinct operations. But though they were very poor, and therefore but indifferently accommodated with the necessary machinery, they could, when they exerted themselves, make among them about twelve pounds of pins in a day. There are in a pound upwards of four thousand pins of a middling size. Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made twenty, perhaps not one pin in a day; that is, certainly, not the two hundred and fortieth, perhaps not the four thousand eight hundredth part of what they are at present capable of performing, in consequence of a proper division and combination of their different operations.' Smith (1974) Chap 1

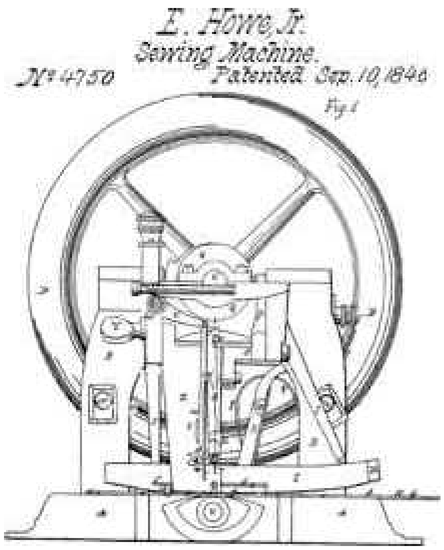


Figure 3.4: The lockstitch sewing machine of Elias Howe.

3.2.3 Mechanical sequencing and power

Nearly all human productive activity involves movements by the hands or limbs. The fingers must move in a precise sequence of motions to manipulate the tool and produce the desired effect on the product. The speed with which this can be done depends on both a flow of information and a flow of energy. The information is supplied by the brain in the form of nervous impulses, sent in the correct sequence to the hand. The energy is supplied by hand and arm muscles, which accelerate the hands plus tools while overcoming mechanical resistance.

There is a limit to how fast even the most practiced hand can move, a limit to how fast a seamstress or tailor could sew. This is imposed both by the brain's inability to provide the nervous impulses faster than a certain rate, and by the speed with which the fingers can be moved. A whole class of industrial appliances accelerated production by first providing a self-acting mechanism to supply the information input, and then providing an external source of power, allowing a previously manual production process to be accelerated.

The classic example of this was the sewing machine. The first functional sewing machine was invented by French tailor Barthelemy Thimonnier in 1830. He was almost killed by other enraged French tailors who burnt down his sewing machine factory because they feared unemployment. In 1834, Walter Hunt built America's first sewing machine. He later lost interest in patenting his sewing machine because he believed the invention would cause unemployment. Sewing machines did not go into mass production until the 1850s. The first commercially successful sewing machine was the one designed by Isaac Singer. The Singer machine used the Lock-stitch mechanism patented earlier by Howe (Figure 3.4). It therefore differed from a tailor in using two threads instead of one. The upper needle simply moved up and down while the cloth was dragged past it. Meanwhile a shuttle containing a second reel of thread was rotated through the loops created in the first thread. Singer's machine could be operated either by a treadle or by a crank. It was a huge success and Singer and Howe both became multimillionaires.

The key to its success was the fact that it greatly increased the productivity of sewing cloth together. The number of stitches a person could do per hour increased by an order of magnitude.

The speed of stitching could be made much higher for two reasons. First, the much stronger muscles of the leg replaced those of the hand in moving the needle. Second, the sequence of needle movements was no longer generated by the human nervous system translated in finger movements. Dexterity gave way to rotary action as cams, cranks and levers sequenced the thread movements to generate the lock stitch. The cams could operate far faster than the nimblest fingers, turning every tailor into a Rumpelstiltskin.

Training can accelerate manual skills immensely, as the control of our muscles is transfered from conscious to reflex action. But such acceleration meets its limits, set both by the reflex speed of our nervous system and ability of our hand muscles to accelerate and decelerate our fingers. A machine with an external power source is freed from these limits. The sequence of movements to be made is now encoded in the mechanical linkages. Rotate the drive shaft faster and the sequence speeds up. The ultimate limit now becomes either friction or the strength of steel exposed to sudden acceleration and deceleration. This can be a couple of orders of magnitude above the limits of human dexterity.

The automatic control mechanism of the treadle sewing machine allows muscular effort of the foot to produce an embodied information structure in the twists and loops of the stitches It is worth noting here, that once we deal with a repeated process like stitching, that the algorithmic and thermodynamic conceptions of entropy diverge. If an automaton is to produce a repeated pattern $P = c^n$, containing n repetitions of a basic cell c, then we would expect the algorithmic information to be bounded by $H(c) + H(n)$. It will be bounded by the information content of the basic cell plus the information content of the number n, But since an integral number can always be expressed in binary, the information content of n must be bounded by the number binary digits in n. Thus on algorithmic grounds we would expect $H(P) \leq H(c) + \log_2 n$. When analysing the thermodynamics of production this formula does not necessarily hold.

Thermodynamic analysis of production is more complex. Doing one hundred stitches clearly involves about one hundred times more physical work than doing one. Some of that work will be dissipated in frictional heat, a clear entropy increase. Another part goes into bending and twisting thread both in the stitches and in the cloth being worked on. This increase in thread entropy absorbs another portion of the work. Thus the thermodynamic entropy increase varies as $nh(s)$, where $h(s)$ is the increase in the entropy of the thread involved in doing a single stitch.[12]

3.3 PARALLELIZING PRODUCTION

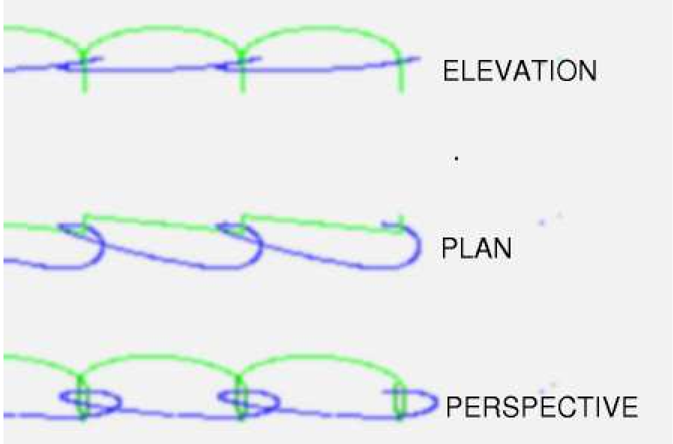
The sewing machine greatly increased the productivity of tailors, but it did not usher in a social revolution. Individual tailors could still afford to work on their own since the price of sewing machines was within their reach. The sewing machine in fact became a staple of domestic equipment, allowing women to clothe their families more cheaply. It was compatible with the continued self sufficiency of the farm household.

3.3.1 More people

Today most work done by sewing machines is done in factories. Millions of women are employed in Asia sewing garments for western chain stores. In these factories productivity will be somewhat higher than in domestic

---

[12] For macroscopic products the thermodynamic entropy changes are much larger than the algorithmic entropy changes. For sophisticated nanosystems which may be built in the future evolving along conservative lines, like Feynmann's proposed quantum simulator Feynman (1999) the thermodynamic and algorithmic entropies of repeated patterns may be equivalent.



```
procedure sew ;
begin
    $\theta \leftarrow \frac{2\times\pi\times\iota_0}{c}$;
    $s \leftarrow 2 \times r \times (\frac{\iota_0}{c})$;
    $x1 \leftarrow \begin{cases} r+s & \text{if } \iota_0 \textbf{ mod } c < h \\ s-r\times(\cos(\theta)) & \text{otherwise} \end{cases}$ ;
    $y1 \leftarrow 0.5 \times r \times \sin(\theta + \pi)$;
    $z1 \leftarrow 0.125 \times r \times \cos(\theta)$;
    $x2 \leftarrow x1 + r \times (-0.2 + 0.45 \times \sin(\theta))$;
    $y2 \leftarrow -2 + 0.1 \times r \times \sin\theta$ ;
    $z2 \leftarrow r \times (-0.35 + 0.35 \times \cos\theta)$;

end ;
```

Figure 3.5: The Lock-stitch presented in plan, elevation and perspective views, along with a generating action program. Note that all steps of the action program are generated by sine and cosine functions driven by θ which models the angular rotation of the sewing machine's drive wheel. The computer algorithm has to specify 6 degrees of freedom, 3 for each thread. This is to ensure that our modeled thread does not intersect itself. A practical sewing machine will work by controlling 4 degrees of freedom: a) the movement of the cloth, modeled in the algorithm above by s; b) The vertical movement of the needle, modeled by $y1$; and c) the circular movement of the lower thread, modeled by $x2$ and $z2$.

production, but not enormously so. Such productivity gains as there are stem from the mechanisms analysed by Smith over 200 years ago: the division of labour and the repeated execution of the same task. But these gains are not huge. What has happened to transform the sewing machine from a tool of family independence to an instrument of exploitation?

It is a combination of two factors. First the big difference in wealth between the already industrialized nations of Europe, North America, Australia and Japan means that there is a huge demand in these countries for cheaply made clothes. Since the goods are being exported across the world, the trade inevitably falls into the hands of capitalist middlemen. These, through their contacts and wealth are in a position to supply material to, and sell on the products made by, individual seamstresses. With the passage of time it becomes advantageous to them to bring the workers under one roof and make the seamstresses direct employees. In so doing they gain better control over the labour process, can impose stricter work discipline, and save the costs of distributing cloth to lots of home-workers.

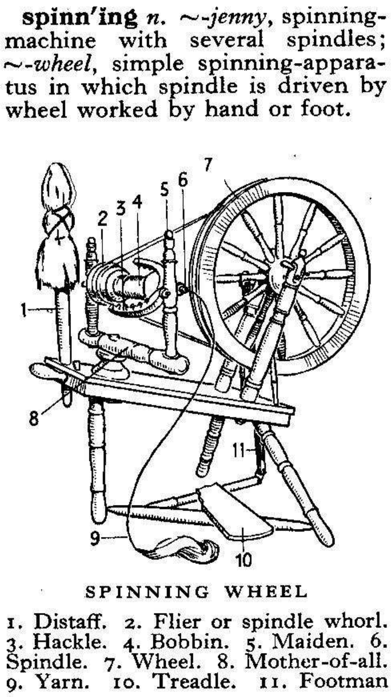
A second cause is the dominance of distribution in the developed capitalist world by big chain stores selling branded goods. These big companies can place contracts for large numbers of identical garments, with local manufacturers. They require cheap standardized garments produced either in sweatshops or by homeworkers subject to the control of subcontractors.

The employers can exploit the machinists because the employers are rich and well connected, whereas the machinists are poor. The employers don’t exert their control due to any particularly superior technology, but due to their social position. But they have this position because their role in an international capitalist trade network. And this network in turn depends upon the prior industrial development of richer nations going back two centuries.

3.3.2 More spindles

It was not sewing machines that drove the birth of the industrial revolution but spindles.

Immediately prior to industrialization yarn in Europe was produced by domestic treadle spinning wheels. The wooden spinning wheel looks a much more primitive machine than Singer’s sewing machine, but in many ways they were very similar devices. They were both driven by foot power. Both were, in a sense, single-threaded. The spinning wheel allows the twisting and drawing out of a single strand of thread. Both involve a modicum



spinn'ing *n.* ~-*jenny*, spinning-machine with several spindles; ~-*wheel*, simple spinning-apparatus in which spindle is driven by wheel worked by hand or foot.

SPINNING WHEEL

1. Distaff. 2. Flier or spindle whorl. 3. Hackle. 4. Bobbin. 5. Maiden. 6. Spindle. 7. Wheel. 8. Mother-of-all. 9. Yarn. 10. Treadle. 11. Footman

Figure 3.6: Treadle spinning wheel.

of hand control—guiding the cloth in one case, drawing out the yarn in the other. Like the early sewing machines the spinning wheel was essentially a domestic instrument of production. No factory system based on spinning wheels ever established itself. The mechanization of spinning took what was essentially a much more adventurous course than Singer did. Compton's Mule (see Figure 3.7) multiplied the number of spindles and also replaced the hand actions of the spinner with a sequence of mechanical movements.

The spindles were mounted on a moving carriage. The sequence of actions emulate those done by a hand spinner.

(1) The carriage moves out, drawing the as yet unspun yarn through rollers that impede its progress. As this happens the spindles impart a twist onto the yarn. This emulates the first action of the hand spinner as they move their hand away from the spindle stretching the yarn.

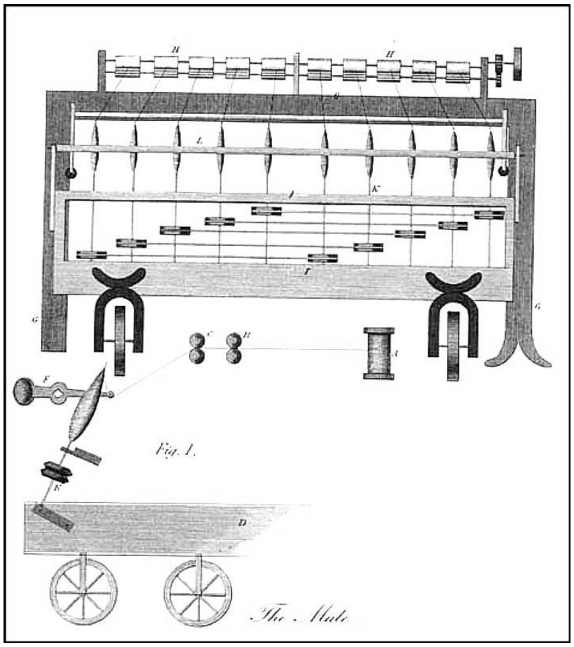


Figure 3.7: Compton's Mule. Note the multiplicity of spindles and the moving frame which substituted for the stretching movement of the hand spinner's arms.

(2) Next the carriage stops and the spindles start winding the thread onto the bobbins. Simultaneously the carriage moves back to the starting position as the thread is drawn in.

(3) The cycle repeats.

The mule was, in the terminology of the day, *self-acting*. We would now say it was *automated*. It carried out its basic sequence of operations so long as power was supplied. Human invervention was restricted to loading and unloading bobbins, and connecting broken threads.

While the fact that the mule was water or steam powered meant that it could spin each individual thread faster, this was not critical. The really important thing was the parallelism. Combined with self action this allowed the number of threads spun by each worker to grow enormously. The system illustrated in Figure 3.7 illustrates an 8-fold multiplication of productivity but later mules increased the level of parallelism to the order of 100 fold.

3.3.3 From Samian ware to UV lithography the development of printing like technologies

Pottery casting We will now look at a quite different method of raising productivity, one which has a long history and is transforming society even now. One of the early mass production industries was the Roman Samian ware industry which flourished from .. to .. It produced ochre coloured pottery kitchenware vessels with raised designs as illustrated in Figure **??**. These were unlike earlier pottery styles in that large numbers of identical pieces were produced. The key to this was the use of casting.

Pottery vessels went through two earlier stages of development. In the first phase pots were made by hand shaping the clay prior to firing. Next came the invention of the potters wheel. This, perhaps the earliest rotary production tool, accelerated the production of circular vessels. The rotation of the wheel meant that the potter had only to specify two parameters for each vertical position on the pot: its inner and outer radii. The 'specification' was done by where they placed their thumb and forefinger relative to the axis of the wheel. The wheel enabled pots to be made with a reduced algorithmic information content. The pots were more even and their production was easier. The potter's wheel was the progenitor of a whole class of rotary tools such as lathes and drills.

The next development dispensed with the wheel and introduced moulds. Clay was pressed into a pre-shaped mould and took on the entire shape in a single operation. With the wheel shaping was still a sequential process. A one dimensional path, a spiral was traced out in the frame of reference of the pot by the potter's grip. With casting the shaping became a parallel two dimensional process. The mould is a two dimensional surface with information encoded as raised and lowered details. Consider what this implies:

(1) The shape is impressed onto the whole surface simultaneously. Of course this is only approximately true with the mold for a curved vessel, but we can conceptualise this as a process in which an approximately flat die comes into contact with a roughly flat sheet of clay, imposing detail right across the surface. The archetypical model becomes more realistic with subsequent examples of this sort of production.

(2) Whereas the wheel accelerated production by reducing the algorithmic information in the product, moulding did not have this disadvantage. It allowed arbitrary and detailed artistic patterns to be embossed

on the piece. The product of the wheel must be a solid of revolution, and arguably, much of the beauty of hand turned pottery stems from this constraint. Moulding allowed decoration to run riot. Samian ware seems to have an almost Victorian love of fancy detail.

(3) No two pots turned on the wheel are the same, but the Samian ware industry was able to churn out masses of identical bowls. Moulding allowed standardized mass production. This was helped by the fact that moulding can be recursive. Pottery moulds were a negative image of the final pot, with raised areas on the pot being depressions on the mould. But if the mould was ceramic, it could be made by first pushing a positive pattern piece into an unformed mould which was then baked. Suppose that a mould could be used 100 times before it became too damaged. Suppose further that the master used to make the mould coul be used 100 times, then this two step process could turn out 10,000 copies of the original pattern piece.

It is worth returning to the paradox relating algorithmic to thermodynamic entropy in production that was mentioned in section 3.2.3. There we said that the algorithmic information in repeated production grows by a law of the form $H(P) \leq H(c) + \log n$ where P is the total product made up of n repetitions of c. If we look at the process of reproduction as a whole there are two terms the first given by the complexity of the original and a second logarithmic term given by the number of repetitions.

In the case of the Samian ware pottery there is the original work of producing the master or pattern piece which corresponds to $H(c)$, but then the number of copies that could be made grows exponentially with the number of successive steps of copying: if the master is used directly to produce the pots then L pots can be made, where L is the lifetime of the master. If the master produces moulds which in turn produce the pots then L^2 pots can be made, etc. Invert the relationship and we find that the number of successive steps of copying will be related to the number of pots produced n as $\log_L(n)$, a relationship suggested by the predictions of information theory.

Cast iron moulding The application of mass production to iron working required a similar path. The crucial step here was the ability to cast iron.

> The transition from the *stackofen* to the blast furnace was gradual. In the taller furnaces the iron ore remained exposed to the reducing action of charcoal for a longer period, and this, combined with higher

> temperatures from the water-driven blast, generally, but not always; caused some of the iron to melt and trickle from the bottom of the furnace, where it solidified. This iron, having absorbed enough carbon to transform it into cast iron, which is brittle and unworkable in the forge, was an annoyance to the smelter whose object was to produce low carbon wrought iron. As yet he had no use for cast iron and returned it to the furnace to be remelted. In the early part of the fourteenth century, a new term began to appear among iron smelters *flussofen*, that is, a flow oven, clearly indicating that it was capable of producing molten iron.
>
> It was also known in German as a hochofen and in French as a *haut fourneau*. The increasing appearance of molten iron running from the furnace presented the smelter with a problem. We are left to conjecture what may have passed through his mind. In the proportion that iron flowed from his furnace, the quantity of wrought iron which he obtained was lessened.
>
> At the same time, the return of the solidified iron to the furnace for remelting interfered with his operations as a producer of wrought iron. Bronze was then being cast in many forms. Among the chief, if not the chief, cast bronze products were church bells. The iron smelter was certainly familiar with the bronze foundry industry. What could have been more natural than for the producer of cast iron and the bronze foundryman to have been brought together? The circumstances under which this may have occurred are obscure, but it appears most likely that church bells were the first cast iron products extensively produced, followed by a much greater demand for cast iron cannon and cannon balls. (Fisher (1963), p. 27)

Prior to the development of the blast furnace iron objects required repeated hammering to forge a shape out of the bloom. The resulting wrought iron was tough but expensive to produce. Its use was limited to tools and weapons. Once iron could be heated enough to cast it, one could make shapes that would be hard to produce by hammering—for example cooking pots or cast iron stoves. These could be mass produced from a single pattern wooden pattern from which sand-moulds could be taken.[13] This allowed the mass production of iron utensiles for applications where high

---

[13] *Of casting iron and other metals.* Patterns of wood of metal made from drawings are the originals from which the moulds for casting are made: so that, in fact, the casting itself is a copy of the mould; and the mould is a copy of the pattern. In castings of iron and

tensile strength was not essential. Iron stoves, pipes and cookware became available for domestic use. Cast iron pillars could be used to support the large working areas of mills. Cast iron members operating in compressive mode could be used for bridges. As with Samian ware we see an exuberance in decorative detail made possible by the new technology.

The development of the Bessemer process then allowed the same sand mould technology to be applied to steel production so that even parts used in tension could be cast. The mass production of car engines for example, would have been impossible without castings.

Again we have a technology that utilizes the parallel formation of a product enabling a huge extension of production.

Plastic mouldings In the 20th century one saw a recapitulation of history as plastic moulding became available. As with cast iron, this enabled the mass production of domestic utensiles. The significant differences were that plastics were lighter, and could be made to higher dimensional accuracy than cast iron. If one considers products from vacuum cleaners to buckles, we see a progressive replacement of cast or pressed metal parts by cast plastic ones. Aside from the gain in weight, manufacturing costs are reduced by replacing a sequence of metal forming steps by the parallel forming of the product in a mould.

Printing press The casting of pottery vessels was not the first use of impressions. The use of seals as a certificate of authenticity in correspondance

---

metals for the coarser purposes, and, if they are afterwards to be worked. even for the finer machines, the exact resemblance amongst the things produced, which takes place in many of the arts to which we have alluded, is not effected in the first instance, nor is this necessary. As the metals shrink in cooling, the pattern is made larger than the intended copy; and in extricating it from the sand in which it is moulded, some little difference will occur in the size of the cavity which it leaves. In smaller works. where accuracy is more requisite, and where few or no after operations are to be performed, a mould of metal is employed which has been formed with considerable care. Thus, in casting bullets, which ought to be perfectly spherical and smooth, an iron instrument is used, in which a cavity has been cut and carefully ground; and, in order to obviate the contraction in cooling, a jet is left which may supply the deficiency of metal arising from that cause, and which is afterwards cut off. The leaden toys for children are cast in brass moulds which open, and in which have been graved or chiselled the figures intended to be produced. Babbage (1832) , Chapter 11, section 106.

certainly predated it. Sumerian cultures used cylinder seals that could be rolled onto wet clay tablets. Roman administrative authorities used circular stamps looking very like modern postmarks to mark government property. The stamping of coins is another similar example. The purpose in these cases was to have a mark that was unique and easy to apply. The information on the mark could be easily replicated but the master stamp or seal was difficult to replicate. A particular information structure then authenticates an object or claim on an object.

These are specialized activities though, not involving mass production. That changes with the development of the printing press and moveable type. Printing replaces the serial production of the scribe with parallel processing. An entire folio of several pages is formed with a single impression. Here we have the clearest, the archetypical, example of this class of production process. Information, encoded in the physical structure of the array of type is simultaneously transfered across an entire plane surface onto a receiving medium, the paper. It is clear that what we have transfered is information: we can read it. The transfer is done by a physical movement of the press at right angles to the paper.

But in printing, making marks on paper is the final step in the process of information copying. What made the printing press revolutionary in Europe was the moveable type.[14] One could in principle have carved an entire page of a book as a single block using etching or engraving, as was done with the earlier Chinese wood block printing. This would have speeded up the making of prints, but the work of engraving the master plate would still have been considerable. The use of pre-cast type reduces the labour required to make the master. The information in a page of type comes at two conceptual levels. The semantic level is given by the sequence of words, further decomposed, in Europe, to a sequence of letters from a small fixed alphabet. The shape of these letters comprises a second level of information. In hand written text each letter ‘B’, ‘W’ etc will be different. In printed text they are

---

[14] Printing from moveable types. This is the most important in its influence of all the arts of copying. It possesses a singular peculiarity, in the immense subdivision of the parts that form the pattern. After that pattern has furnished thousands of copies, the same individual elements may be arranged again and again in other forms, and thus supply multitudes of originals, from each of which thousands of their copied impressions may flow. It also possesses this advantage, that woodcuts may be used along with the letterpress, and impressions taken from both at the same operation. Charles Babbage *Economy of Machinery and Manufactures*, 1832, Chapter 11, section 93.

all identical ‘BBBB...B’ etc. The type used in each B is cast from the same mould. This means that the information in a page of printed text is much less than that in a page written by hand. The cheapness of printing stemmed from the following:

(1) Each folio off the press was identical to the preceding one. This means that in algorithmic terms we are exploiting the logarithmic term of the repeated production cost. In labour time it makes use of the fact that repeated copies cost only the labour required to load a sheet of paper and operate the press through one cycle.
(2) The fact that individual letters do not have to be carved reduces typesetting to the choice of appropriate letters.

Taken together these represented a huge change in the productivity of information copying. They were a material precondition for generalized literacy and the eventual development of industrial civilization.

The process of parallel transfer of information to the product initiated with ceramic casting creates an independent existence for the information source. With seals this independent existence was harnessed to certify the validity of documents. Only the holder of a particular seal could validate a document. But the invention of moveable type transformed this relationship. The particular configuration of type used in an edition of a book became incidental as the type themselves were re-usable. The printers plates are of little value in themselves. The information that is being impressed on the page is only secondarily the particular shapes of the letters used. A change in typeface alters all of these but leaves the book substantially unaltered. It becomes clear that what is being transfered is an information structure that has multiple possible representations. The book is an abstract identity surviving its impressions, defined purely as a sequence of characters.

We have a three-stage evolution of the relationship between labour and information in the product here:

(1) In handicraft work, the information is impressed on the product by the bodily movement of the artisan and has no idenpendent existence.
(2) In pattern or mould based production, the handicraft work is captured once in a pattern or mould from which multiple copies are made. The pattern piece is then an independently existing encoding of the information, whose possesion implies social power. This is either overt in the case of the holder of a seal of office, or implicit in the

iron-master's ownership of a store-room of pattern pieces for standard products. The pattern pieces embody much more labour than the individual products they inform and their monopolization gives market power to their owner.

(3) In printing the information structure becomes abstracted from the impressing apparatus and potentially mobile. Printer with a copy of a book can turn have it typeset and turn off an impression at will. All that is required is the labour of typesetting which is typically less than the labour of authorship. Printing breaks the link between material possesion and ability to reproduce.

If the labour of writing was to be recompensated in a society of independent commodity producers, the sequence of words itself had to be made an item of property. Hence the printing press in combination with bourgeois social relations gives rise to the law of copyright. Information becomes property independent of its material embodyment.

It is notable that whereas in the case of authorship, the direct producer of the information ususally ends up owning it, this has not been the case for pattern-making, the author owned his copyright, the iron-master owned the patterns, not the pattern maker. Whence the difference?

There appear to have been a number of contributory factors here. The ponderous nature of the patterns made them analogous to other products of direct labour which, in bourgois right, always belong to the employer. A pattern used in sand casting was apparently no different from any other piece of exact carpentry. The pattern-maker might be the more skilled worker and paid better than a moulder, but he was still an employee working at his masters' direction. Next we have to consider that in the casting of machine parts, the pattern would often be an embodiment of information already recorded as technical drawings by an engineer. But this can not have been decisive since the original designs would not necessarily be the property of the iron-master, but might belong to the customers to whom he was contracted to produce parts.

Prior to this one has to ask why the pattern-maker ends up as a wage labourer surrendering his right to the information he produces whereas the author typically remained an independent agent. The decisive factor has to be the extent to which the process of producing information structures can be carried out independently. An author can write 'on his own account', since there is little need for collective input to his production. The work of the pattern-maker forms part of an industrial division of labour. The function of any system of property law is to ensure the reproduction of the agencies of production, be these agencies individuals, firms or the state.

In a commodity producing society, non-state agents of production can only survive by the sale of their product. If that product is an information structure, the agency that bears the cost of making it, will tend to own it. They can then survive by selling either the information itself, or the use of the information. This is a sufficient cause for their survival, whether it is a necessary cause is another matter.

Photography versus painting Printing technology gave us the mass production of images.

Picture prints could be cheaply turned out provided that a human artist had made the master copy. This might be an etching or a lithograph but in either case the information on the page went via human eye, brain and hand. This meant making the master was an inherently serial process. The camera changes this.

Photography, literally translated means drawing with light, but this is an understatement.[15] It is printing with photons. Instead of a metal plate comming down on the paper at centimers per second, wavefronts of light traveling at 300,000 kilometers a second impose their image on the film. As in printing, they work on the whole frame simultaneously.

With photography all humanist mysticism relating information to concious agency is evaporated. With photography the creative subject vanishes. The image is a work of nature. The photographer, where he is even present, has his role reduced to selecting the vital instant at which nature can do its work. With photography Landauer's aphorism that 'information is physical' is literally made manifest. Photography was our first technology to encounter the limits that nature places on the handling and transmission of information. Consider some of its contraints.

Photon quantization Although the light waves that impinge on the film approach at the ultimate speed c, this does not produce the acceleration in process that one might anticipate. To form an image we need

---

[15] True photo*graphy* had to await the laser printer, whose hair thin beam, like the engraver's stylus, forms its image stroke at time

photons to interact with tiny crystals of silver iodide and seed their photodecomposition. Where struck by photons the crystals break down to leave black colloidal silver.

An individual tiny crystal makes a binary choice, it is either hit by a photon and decomposes or it does not. If struck it evolves to a black dot, if not, it will be dissolved away in the developing process. But we do not want our picture to be just black and white. We want shades of grey. Suppose we want to have 100 shades of grey available. Then we need 100 crystals in each small area that we can resolve. Suppose we have crystals that are $\frac{1}{10}$mm across. Then 100 crystals will fit into each square millimeter of film.

When we take a picture we exploit that probabalistic nature of the photodecomposition process. If an exposure caused all crystals to absorb photons then we would get a totally black surface. If it was so short that no photons hit any crystals the film would be left white. To get an acceptable grey to black range, we need to have an exposure such that, given the ambient light levels, we would expect that, on a randomly chosen part of the film, about 50% of the crystals will have decomposed. The longer we have the shutter open the more likely it is that we will have enough photons arrive at the surface. Because the arrival of photons is a random process the actual number of crystals triggered will vary. An area with 100 crystals 'should' have half its crystals black, but sometimes has 40 sometimes 60 etc. This gives the film a grainy, noisy look.

We can remove the graininess by using smaller crystals. As you increase the number of crystals in each small area the percentage of them that will turn black at a given light level becomes more predictable.

The noise introduced by photon quantization is refered to as "shot noise". The degree of uncertainty induced by the quatized nature of light is proportional to the square root of the number of photons arriving on a sensor.

Number of crystals	expected brightness error
10	16%
100	4.6%
1000	1.5%

As the number of crystals rises the error in our estimates of the light level falls. This is visually apparent as a smoother less grainy image. Having smaller crystals enables us to capture more information about the light falling on each small area of film. But this gain in information about light levels comes at a cost. It makes the film slower. Smaller crystals have a lower probability of absorbing a photon, so we have to open the shutter for longer to capture our picture. Gaining more information about intensity means that we are restricted to photographing static scenes.

We are up against the fact that information is not only physical, it is physically quantized. The information available about a scene is encoded in the trajectories of photons arriving from it. There are only a finite number of these available. The numbers of photons arriving sets limits on how much we can know about the scene. A fast film allows us to image rapidly moving objects, but the cost is a coarse and grainy image. An alternative is to supplement the supply of photons. In film studios where they want to capture motion and have high quality images, they have to use intense artificial lighting.

λ is the letter conventionally used to represent the wavelength of light or other electromagnetic radiation. Visible light has a $\lambda \approx 0.5\mu = \frac{1}{2{,}000{,}000}$th of a meter. λ determines the smallest details that we can in principle represent by photography. You can not use photography to form a pattern of light and dark whose smallest features are smaller than light waves. This is mainly of relevance in microscopy or the manufacture of microscopic components. But as micromanufacture has become the governing technology of our age, this constraint weighs more and more heavily upon us.

Let us concentrate for now on photography at conventional scales. We are still faced with constraints imposed by the wave-length of light. The problem arises from diffraction. The wave nature of light imposes a relationship between the resolving power of a lens and its apperture. If you take a picture of a star , something which is effectively a point source of light, with an ordinary camera what you actually see is not a point but little fuzzy circle. The angular size of this circle of confusion is roughly given by the ratio of the wavelength of light to the aperture of the camera. If one has a tiny camera with an aperture of the order of a millimeter, you cant expect to be able to resolve more than few hundred distinct points across your image. As the aperture of your lens goes up, so does the resolving power so that the number of pixels you can have in your image rise in proportion to the area of your lens. The constraint on the production of pictures is then set by the amount of information actually passing through space as light.

Sound recording Sound recording involves copying in two senses. In the first sense a musician plays a piece, this is recorded and subsequently people can listen to a 'copy' of that performance. These copies are separated from the original in time. The second sense involves making distinct copies of the recording itself. These copies can then be separated in space allowing people to simulaneously listen to the performance in many different places.

Copying in the first sense is inherently sequential. The performance has to be done from begining to end and recorded as it takes place, while the act of listening is also sequential. Copying in the second sense can be either parallel or sequential.

The production of records—whether the old analog ones or the modern CDs—is a special example of a casting processs. A master disk on which a negative image of the tracks has been cut is used to press out the disks from hot soft plastic. As such it is a parallel process. The entire recording is transfered to the disk in a single step. When music is recorded onto tapes on the other hand the copying process is inherently sequential. Such parallelism as there is, is due to having large numbers of tape recorders operating at the same time.

The transition from Edison's original cylindrical phonograph to disc recording was driven by the need to economise on copying. A cylinder could not be pressed out but had to be cut sequentially. The cheapness of disk pressing is what created a mass market for sound recording. With the record industry it at first appears that what is involved is merely the mass production of a material object, and in this context the efficiency of the productive process was vital. But the internet has revealed what should always have been clear: records were merely an intermediary to the copying of performances. People were being forced to buy the material object to get the information it contained. All products contain information, added during production, but for some products—initially books, then records and now software—their use value is their information.

Radio and TV Radio and television take the technologies of sound recording and photography and add to these the principle of broadcasting. Here the product, namely, radio waves, is a direct physical, though immaterial embodiment of information. Once released, broadcast information is available to anyone within reach of the transmitter. The number of copies of a broadcast musical performance that are heard is limited only by the number of recievers within range. The marginal labour embodied in each heard performance tends to zero as the number of listners goes up. To produce say a live musical broadcast there is a certain fixed cost: the time of the musician, the time of the technicians operating the broadcasting equipment, the depreciation on the equipment. These costs are essentially unrelated to the number of listeners. The only component of the cost of broadcasting that relates to the number of listeners is the power used by the transmitter. This tends to be a relatively small part of the total cost.

From its inception therefore, broadcasting was an implicitly "communist" medium, where performances are given away free to listeners. This free distribution meant that the labour required to run the broadcasting system had to be in a sense directly social labour. The BBC provides a model of this where what is essentially a special tax, the Radio License, was levied to provide the service. The private-sector equivalent, broadcasting funded by advertising, essentially taxes the sellers of mass produced goods to meet broadcasting costs. Once radio and TV advertising is introduced, manufacturers of consumer goods are forced to finance TV, or lose out to competitors who do.

The free nature of broadcasting prefigures the general transition of the mode of material production to one favourable to communism. As production becomes more and more dominated by the principle of copying information—a principle that has been in development ever since pottery casting by the Romans—the underlying cause of commodity production and market mechanisms comes to be increasingly undermined. Commodity forms of production can only be sustained by increasingly elaborate and 'unnatural' legal constructs that enforce property rights over information.

Printed circuits and Integrated Circuits The dominant technology of the first decade of the 21st century is digital electronics. This technology has seen sustained rates of growth of productivity that outstrip anything seen in past generations. At the heart of this growth has been the progressive refinement of copying technologies. The key component of contemporary digital technology is the NMOS transitor. This is the basic element, which repeated millions of times over, builds our computers, cell phones etc. A transistor is basically an electrically controlled switch. Figure 3.8 illustrates a cross section through a transistor. It comprises 3 electrical contacts, the source, the gate and the drain. When the switch is on, current flows from the

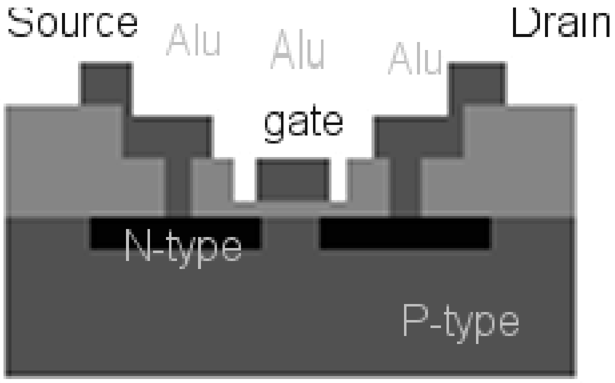


Figure 3.8: NMOS Transistor

source to the drain. When it is off, current can not flow, since current flowing from the source is impeded by having to pass from the N-type silicon aroung the source to the surrounding P-type semi-conductor. To turn the switch on a positive charge is applied to the gate. This creates repells positive charge carriers from below the gate, creating a temporary N-type channel under the gate linking the source to the drain. The components occupy a thin layer in the surface of a silicon chip. The key to their manufacture is to lay out and interconnect large numbers of these transistors on a chip.

Figure 3.9 shows the fabrication steps involved in making NMOS semiconductor chips. Reading left to right and top to bottom these are:

(1) Start with a polished wafer of P-doped silicon.

(2) Oxidize the wafer to form a SiO_2 layer about half a micron thick by heating the wafer to about 1000°C in an oxygen atmosphere.

(3) After oxidizing, a layer of photoresist is spread on the wafer. This is done by rapidly spinning the wafer so that drops of photoresist spread out to a uniform layer before drying.

(4) Next use photolithography to define the source and drain areas of the transistors (one transistor is shown). This involves shining UV light through a shadow mask ensuring that only some areas of the photo resist are exposed to the light. The exposed area undergoes chemical changes allowing it to be washed away by a developing fluid. After this step the information structure on the mask has been transfered to

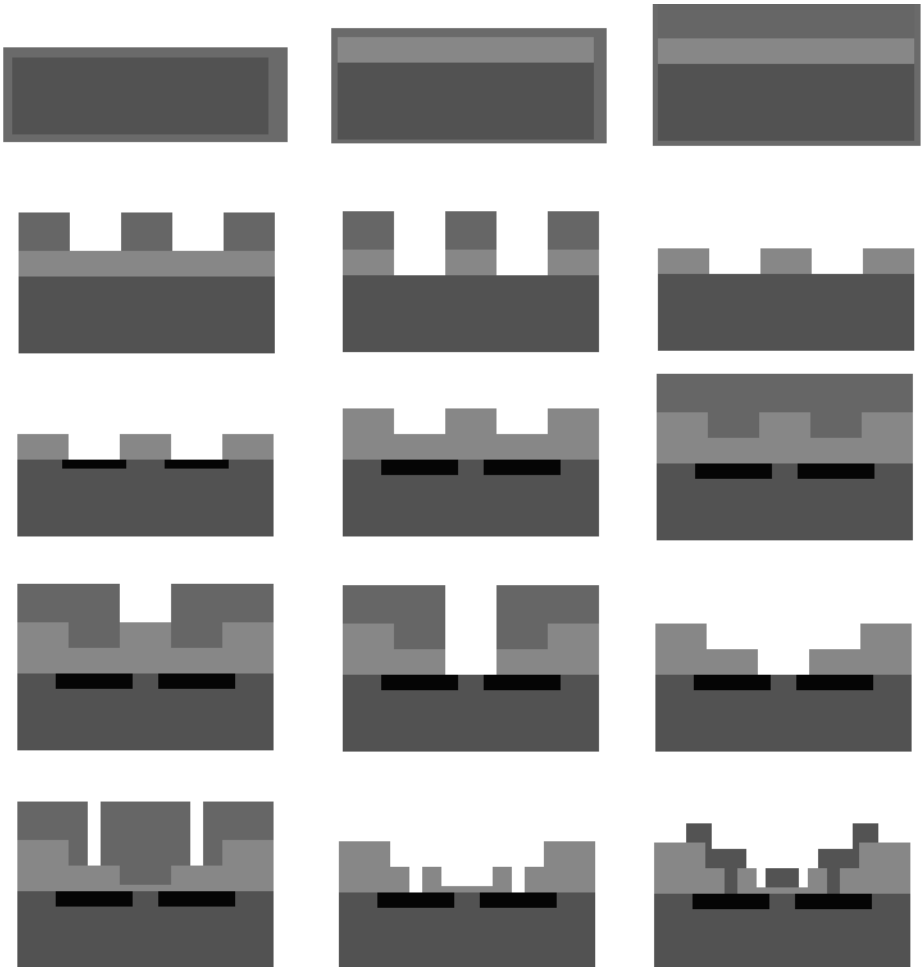


Figure 3.9: NMOS fabrication steps

a pattern of holes and lands in the photoresist. The photo resist is then baked so that it can resist acid which will be used to etch holes in the oxide layer.

(5) Expose the wafer to Hydrofluoric acid to disolve the the oxide wherever there are holes in the resist. The acid does not disolve pure silicon so the etching stops once it is through the oxide layer.

(6) The photo-resist has been dissolved by an organic solvent leaving a pattern of holes and lands in the oxide layer that matches the pattern

on the original mask. (We are now at the right end of the second row in Figure 3.9.)

(7) The silicon under the holes is now doped to N-type (black in the diagram) by diffusing phosphorous into it. This step forms the source and drain of the transistors.

(8) The oxidation of step 3 is repeated to grow a fresh layer of SiO_2.

(9) A new layer of photoresist is spun on. We are now at the end of the third row of the diagram.

(10) The photoresist is exposed under a new mask and a hole is etched through the oxide to expose the area of the silicon that will become the gate of the transistor. We are now at the end of the fourth row of the diagram.

(11) A short further oxidation step is used to place a very thin insulating layer (a few hundred angstrom) across the top of the gate area. This has to be thin to allow sufficiently strong electric fields through from the gate to switch the transitor. we dont show this step in the diagram, but the resultant oxide layer can be seen in subsequent images.

(12) Another sequence of photoresist coating, UV exposure and etching is used to cut contacts through the oxide down to the source and drain. We are now at the middle of row five.

(13) The wafer is coated with aluminium. This forms the wires on the surface of the chip. The wafer subjected to yet another round ot photo resist coating, exposure and etching to cut the uniform aluminium layer into a networkof wires joining the chips. This yields the final circuit.

It is evident that the crucial repeated step in this manufacturing process is photolithography. It is this that is used to transfer patterns from a mask to the chip. The ability to project a clear image of a very small feature onto the wafer is limited by the wavelength of the light that is used and the ability of the reduction lens system to capture enough diffraction orders off of the illuminated mask. Current state-of-the-art photolithography tools use Deep Ultraviolet (DUV) light with wavelengths of 248 and 193 nm, which allow minimum feature sizes on the order of 130-90 nm. Future tools are under development which will use 157 nm wavelength DUV in a manner similar to current exposure systems. Additionally, Extreme Ultraviolet (EUV) radiation systems are currently under development which will use 13nm wavelengths, approaching the regime of x-rays, and should allow feature sizes below 45 nm.

The number of transistors that can be produced per square centimeter of silicon obvious varies inversely as the square of the feature size. If you half the feature size you can produce, then the number of transistors you can make goes up four times. Productivity gains have also come through increasing the sizes of the wafers used, allowing more transistors to be printed with each processing cycle.

The production of ICs shows very clearly how manufacturing moves towards being a process of copying information. In making a new processor chip, the big costs are:

(1) The work of creating the original design for the chip. This is typically a Computer Aided Design file or set of files which is transfered to the master masks used in chip production.

    Each generation of chips uses smaller transistors. This means that the number of transistors used in this year's model is likely to be twice as many as was used in the previous model released two years earlier. In consequence the labour of design grows over time just as the cost of producing the individual components falls.

(2) The capital cost of setting up the IC fabrication line. This tends to rise from generation to generation since the equipment used must be increasingly precise, the standards of cleanliness in the production facilities become more stringent, and the imaging equipment becomes more and more esoteric.

The combined effects of these opposing movements means that while there has been a rapid exponential growth in the number of transistors produced, with a doubling time of the order of two years, the number of firms able to bear the development costs of new products falls. This has led to an increasingly monopolized system of manufacture. One company, Intel, has ended up dominating the world production of CPU chips, with only marginal competition from a few smaller firms.

PCR and genomics The 1950s saw both the birth of the electronic computer industry and the discovery of the structure of DNA. It became clear that living organisms could be seen as self-replicating information structures. The reproduction of cells had as a precondition the copying of genetic information. The biotechnology industry rests fundamentally upon these insights. But since the invention of the Polymerase Chain Reaction (PCR), a copying technology has become a key part of the industrial process for biotechnology.

PCR is a technique for copying DNA. A polymerase enzyme from a thermophilic bacterium is placed in a solution of DNA bases and an initial starter quantity of DNA. The temperature of the solution is then cycled up and down. Each time the solution is warmed up, the double strands of DNA disassociate. As it is cooled, the polymerase enzyme builds up a complementary strand of bases on each single strand. This regenerates a complete double stranded molecule of DNA. Thus each cycle doubles the number of molecules, each of which is a copy of the original starter molecule. Here in the PCR process we seen the full industrial application of the principle discussed in sections 3.2.3 and 3.3.3 whereby the algorithmic information in repeated production grows by a law of the form $H(P) \leq H(c) + \log n$ where P is the total product made up of n repetitions of c. If one wants to make n copies of a DNA molecule containing c bases by automated DNA synthesis followed by the PCR, then there will be two phases. In the first phase a small number, in principle as few as 1, copies are made of the DNA using an automated synthesis machine. The number of steps to be followed here will be of the order c. Next the PCR is used to repeatedly double the number of DNA molecules we have. This phase will have to be repeated of the order of $\log_2 n$ times.

With the PCR we see that the regulation of the productivity of an industrial process follows directly from the laws of algorithmic information theory.

CHAPTER 4

INFORMATION AND COMMUNICATION

Much of the previous discussion has hinged upon information as patterns of difference. What is most interesting about such patterns is that they can carry messages: as we shall see, patterns within information may be structured in different ways to reflect different meanings. In particular, two entities may communicate provided they are able to transmit patterns of information between each other, and to interpret each other's patterns consistently

"The medium is the message." — Marshall McLuhan

For communication to take place, patterned information must be embodied in a form that enables it to be transmitted. Embodying information is relatively straight forward, requiring a medium within which it is possible to discriminate differences.

For example, consider two people talking face to face. The speaker's voice sets up patterns of vibrations in the air between them and the listener. The vibrations pass through the air and enter the listener's ears, eventually enabling them to hear what was said. The communication can take place because vibrating air molecules can embody patterns in information.

If the people were talking by telephone, then speakers voice would, via the air, vibrate the microphone in their handset. In turn, this would set up variations in the electricity in the wire connecting their phone to the exchange. That wire is, courtesy of the telephone system, connected to the wire to the loudspeaker in the listener's handset. The loudspeaker vibrates the air, enabling the listener to hear what was said by the speaker. Here, while electricity is an additional medium for communication, the electric variations carry the same meanings as the vibrating air. This ability to embody information patterns in electricity lies at the bottom of the profound revolution in machine aided communication and information processing, that started in the mid 19th century with the telegraph.

If our telephonists were using a digital telephone system, then the vibrating electricity from the speaker's microphone would be converted by electronic circuits to digital patterns of electric 'on's and 'off's for transmission. Eventually they would be converted back to vibrating electricity to drive the listener's earpiece. These digital patterns contain the same meanings as the continuous patterns in the electricity or air, but are much better suited for manipulation by electronic means, especially computers. For example, its much easier to detect when something's gone wrong with digital transmissions and to correct them. That's why the record manufacturers feared CD's, and the video manufacturers feared DVD: digitally structured information can be copied perfectly, and manufacturers' recordings are no longer superior to home copies.

"Between thought and expression, there lies a lifetime." — Lou Reed

When two entities interact by sending patterned information through a medium, they take it in turns to transmit and receive information. For their communication to be effective, the receiver must be able to understand the transmitter's message. That is, there must be consistency between the meanings that the transmitter generates and that the receiver interprets. Curiously, such consistency is not a property of the information itself: the same pattern of information may be interpreted as having different meanings in different contexts.

Consider, for example, the pattern "111". This might be interpreted as "three" in Roman numerals $(1+1+1)$, or as "one hundred and eleven" in Arabic numerals $(1\times 100+1\times 10\times 1\times 1)$, or as "seven" in binary digits $(1\times 4+1\times 2+1\times 1)$. Here, the same pattern of information bears different meanings in different languages.

Consider for example, Dr Seuss' immortal refrain "Yes I like green eggs and ham". We can show that this has different meanings by placing brackets to indicate how the salient words are associated with each other. The more obvious meanings are "Yes I like green (eggs and ham)" i.e. "I like both my eggs and ham to be green", and "Yes I like (green eggs) and ham" i.e. "I like my eggs to be green and I like ham in general". A less obvious though equally valid alternative is “Yes I (like green eggs) and ham” i.e. “I like my eggs to be green and I’m a bad actor”. That is, we can discern different structures with different meanings within the same pattern of information, even in the same language.

Nonetheless, entities communicate most effectively when they use the same language. In general, a language defines a class of information structures that bear meanings. Its usual to distinguish between the alphabet, symbols, syntax and semantics.

The alphabet consists of the base units of embodiment in the medium, for examples phonemes in speech, letters in writing or ‘on’s and ‘off’s in digital communication. These base units don’t have any necessary meanings themselves. Rather, they provide a first level of structure to apparently raw information.

The symbols are the base units of meaning, for examples words in human languages. While symbols carry meanings, once again they don’t have necessary meanings. For example, in Scots “mind” means “remember”, as in “Should auld acquaintance be forgot and never brought to mind”. In other dialects of English, “to mind” means to object to. Hence the old joke:

> Scots person: Do you mind my face?
> English person: No, you look fine.

The syntax defines sequences of symbols which correspond to a further level of rules for structuring information. Thus, its usual to view many written human languages as composed of phrases, clauses, sentences and paragraphs.

Finally, the semantics describe what the well formed symbol sequences mean. As yet, we have little idea about how to define the semantics of human languages well enough for them to be understood by computers. Nonetheless, we all imbibe semantics as we learn our native tongue. We will return to this problem later on.

Human beings are raised in cultures that seethe with language. Children are immersed in language from the moment they are born: all around them people engage in social transactions using language. The sooner children acquire language they sooner they can take part in effective interaction. More to the point, languages carry ideologies. In learning language children become socialised, that is they imbibe memes that both reproduce and challenge societies.

It is striking how hard people with different languages find communication, even when their cultures are very similar. Although they share the same potential meanings they lack a common means of expressing them. Apparently trivial activities, like buying train tickets, involve astonishingly rich powers of expressiveness.

Contrariwise, people from different cultures but with a common language can communicate unfamiliar, culturally specific meanings. This was the dream behind Esperanto, an attempt to devise an international language. Of course, Esperanto bares the hallmarks of its time, having been invented in 1887. For example, male and female genders are distinguished in contexts in which they are irrelevant. Nonetheless, Esperanto was motivated by the lofty ideal to break down barriers between cultures. However, its failure to compete with, let alone displace, its contemporaries shows how firmly language is rooted in the societies in which it evolves.

While it is not clear whether or not there are linguistic universals that are shared across all human languages, it is likely that human brains are equipped to acquire arbitrary languages. Babies will learn the language they are brought up in and young children from one culture very quickly learn the language of another. Indeed, it is common for immigrant children to mediate between their parents and the host community, on first arrival in a new society. However, this linguistic plasticity is quickly lost. Generally, adults find it much harder to learn new languages. It seems that our brains come to depend on the first language that we acquire, as a high level way to structure accounts of reality.

No one knows how human languages originated. Animals certainly communicate using languages, but they are not able to generate and interpret rich ranges of meanings compared with humans. Most species have limited repertoires of behaviours and very restricted learning capabilities. One implication is that animal brains are not able to represent and process information in such a way as to enable them to change the world as well as interpret it.

This is somewhat unfair. Animals necessarily change their worlds simply by living in them. For example, the eat/shit cycle is central to the reproduction of the natural environment. But most change is localised to animals' habitats, if we conveniently overlook the impact the world's cow population has on global warming, and most animals cannot live outside of relatively constrained habitats. Animal brains and bodies seem best adapted to enabling them to survive in the presence of the sort of slow predictable change experienced in any one locale on our planet as it gently wobbles its way round the sun.

In contrast, human beings survive all over the globe, displacing and exterminating other species, as well as each other, as they spread. Human language is one of the fundamental tools that has helped us to become the dominant species on this peculiar planet of ours. It enables us to make models of our circumstances such that we can predict the effects of our behaviours and modify them to achieve diverse aims.

Of course animals also make models of their environments but much animal modelling is hardwired. For example, hares' coats change from brown to white as autumn turns to winter and from white back to brown again as spring approaches. But this is driven by the temperature and the length of daylight rather than some fiendish harey plan to escape detection amongst the snows.

Unlike hares, chimpanzees seem to have rich inner existences. They have highly structured societies, held together by well understood social rules that are passed from parents to children. Chimpanzees can learn to use rudimentary tools, for example using sticks to winkle honey out of hives, and such tool use is again passed on through the generations.

"If lions could speak, we wouldn't understand them." — Witgenstein

Alas, all attempts to teach chimpanzees human language so far have failed to progress beyond a simple vocabulary based, Paulo Freire fashion, on everyday experiences. Chimpanzees can learn symbols for food and places, and for colour and quantity, and can learn to use them to interact with humans. However, chimps are poor at learning grammar and rarely produce novel word sequences, instead repeating sequences that they know will elicit appropriate responses in their keepers. It is as if chimpanzee brains cannot form and manipulate information structures of adequate complexity to see how things might be different and to make them so.

Chimpanzees are far more bound to their hard wiring than we are. For example, chimps can be taught that if they nominate either of two quantities of food, it will be given to another chimp leaving them with the remaining plate. When pictures of food are used, chimps always point to the smaller quantity and receive the larger. But when real food in closed glass jars is used, chimps cannot stop themselves reaching for the larger quantity, even though they loose it as soom as they do this. In the wild, if you don't eat when you can then something else will instead. This response has become deeply ingrained as a powerful survival mechanism in most animals' psyches, and cannot be displaced by a little local learning.

We share 98 percent of our DNA with chimpanzees. The other 2 percent is why they use tools and we make them. Chimp brains are well adapted for living in unchanging forests, or in circuses and research laboratories. But their limited ability to manipulate information structures is why they don't hold human tea parties in their zoos.

Human information structures like natural language have helped make us masters of our own dung heaps. They have also enabled us to transmit descriptions of our knowledge of our circumstances to other people. Our information structures can be embodied in artefacts, like cave paintings and scrolls and books and floppy disks, that long outlast us as individuals. But that knowledge is only useful to others if it can be discerned from the information structures that carry it.

Consider the Rosetta stone, mentioned in the previous chapter, containing inscriptions in hieroglyphic, demotic and Greek alphabets. Because Greek was already understood, and it was assumed that all three inscriptions bore the same meanings, it was possible to decipher the hieroglyphic and demotic inscriptions. This assumption of consistency might be wrong. The demotic might actually be a cry for help from the unfortunate slave that carved the stone. However, there are sufficient structural similarities between the inscriptions, and with other inscriptions using the same alphabets, to give confidence that they do all represent the same message.

Consider, for example, the strange fate of the Pioneer 10 space craft, launched from the USA in March 1972. In 1997, Pioneer 10 was 10.10 billion kilometres from earth and is still heading out across the cosmos. On its side is a gold plaque showing, amongst other things: the behaviour of a hydrogen atom, the most common element in the universe; the location of our planet relative to the centre of the galaxy; a waving human man next to a smaller human woman. The intention was to persuade passing aliens that there was intelligent life on earth. Perhaps this symbolism is reasonable. We have no idea what information structures aliens will use, so our own are probably as good as any. But the aliens may utterly misconstrue our message of goodwill. Perhaps they will think its junk mail, advertising vacations in a hydrogen rich atmosphere. In truth, they may have no common point of reference whatsoever with our graphical information structures.

“If you want to know the taste of a pear, you must change it by eating it yourself.” — Mao Tse-Tung

But it isn’t just explicit messages that bear meanings. All information structures have meanings for the entities that interpret them. And manipulation of information structures is fundamental to extracting meanings from them. Our aliens will undoubtedly find out more about us by taking the space craft to bits and working out how it functions, that is by interpreting our artefacts themselves as information structures that reflect something about their origins. Similarly, in understanding language, brains and computers tear internal representations of messages to pieces, to identify and interpret their information structures.

Consider a URL in the HTML language used for communication between people and computers:

`http://www.klingon.planet/˜whorf/fun/fighting.jpg`

First of all, this URL is made up of a sequence of letters: ‘h’, ‘t’, ‘t’, ‘p’, ‘:’ and so on. These are the letters we type to enter the URL into a computer. We know which keys to use because they have similar, if not the same, letters on them. These letters are also close to the lowest level representations that computers use to communicate.

Within the letter sequence, different letters mark different regions of the URL. In HTML, as in most human languages, symbols are made up of sequences of a fixed set of alphabetic letters. Thus, we and the computer can distinguish the words “http”, “www”, “klingon” and so on, as being bearers of basic meanings.

The punctuation marks structure the symbol sequences and help us an or a computer elucidate the URLs meanings. The ‘://’ marks the separation between the type of message and the start of its main content. The first word, “http” tells us, and computers, that the rest of the words specify a WWW location. If the first word were “emailto” then we would know that the rest of the words describe an email address. This first word effectively tells us the dialect of HTML in which the rest of the URL is written.

The sequence separated by ‘.’s up to the first ‘/’ is the main address of the WWW location. Curiously its read from right to left rather than left to right. It tells us to look for a specific locale, often associated with an individual organisation, called “klingon”, amongst a number of systems associated with the common domain “planet”. The “www” tells us to start from a standard place in that system’s files, in order to move to the fine detail of the location.

The international standard for WWW locations and email addresses is to write from most specific to most general, as it has been for hundreds of years with top to bottom addresses on envelopes. However, from the point of view of finding an address it would be easiest to start with the most general locale and then home in. Both human and artificial languages suffer from the vagueries of the contexts in which they evolve. Indeed, early UUCP email addresses used to be from most general to most specific from left to rught.

Next, the ‘˜’ before “whorf” tells us to look in a standard directory (also called “www”) at the top level in the files belonging to the user whose login is “whorf”. The next ‘/’ marks the start of the name of a subdirectory in whorf’s “www” file. Here, that directory is called “fun”. The next ‘/’ marks the start of the name of a file in that subdirectory called “fighting.jpg”. The “.jpg” tells us that this file contains a picture, in yet another encoding.

Our understanding of English, and of American TV, helps us deduce that we may have been told to look at a picture of Mr Whorf, from the planet Klingon, fighting. On the other hand, the URL tells a computer to send a set of information structures from its memory to a display screen or printer. Computers don’t yet watch Star Trek and the computer is unable to be even indifferent to our crude representations of alternative ways of being.

Unlike Mao Tse-Tung’s pear, information structures need not necessarily be destroyed in the process of being understood. Even though an individual embodiment of an information structure may be taken to pieces in the process of understanding it, an identical copy may still embody the original information structure. Thus, if our aliens chance upon Pioneer 10’s companion, Pioneer 11, they will find exactly the same gold plaque on its side. And if they broke something in examining Pioneer 10, they can always start again with Pioneer 11. As with the Rosetta stone, they aliens may find reassurance in the duplication of information structures.

We might say that an information structure bears meaning for another information structure when the interactions between the structures cause the second one to change. Computers are information structures and not just in the basic sense of being a particular configuration of atoms. Computers have memories which at any one time can hold a very large number of different patterns of ”1”s and ”0”s. Those patterns will in turn determine how the computer will behave. The microprocessor in the computer inspects the patterns in the memory, one after another, treating them as instructions in a language which is understood by all other microprocessors of the same family. These instructions typically specify that other patterns should be manipulated. We can characterise the state of a computer in terms of all the patterns in its memory and the position of the pattern it is currently inspecting. If we can take a copy of this state then, in principle, we can load it into another computer which will then behave exactly like the first computer.

When two computers pass each other information they change each other’s internal states. Human beings are also information structures with internal states in brains, consisting of links between neurones, electric potentials between synapses where two neurones meet and chemical balances establishing the electric potentials within the neurones themeselves. When human beings interact they also change each other’s internal states. Suppose your neighbour has a pet rabbit. There will be some component of their brain state that determines how the presence of the rabbit affects their behaviours. For example they will routinely feed the rabbit and clean its hutch. When you tell your neighbour that their rabbit has been eaten by your cat, you have changed fundamentally their conception of their rabbit, and hence their subsequent behaviours. For example, they now need to learn not to feed the rabbit. These changes are ultimately mediated by changes in their brain states.

Brain states are fabulously complicated and characterising a person’s brain state as a sequence of symbol patterns is well nigh impossible. This effectively precludes the possibility of cloning truly identical individuals. It may be possible to clone genetically and physically identical individuals, but in the absence of a precise characterisation, they cannot be loaded with the same brain state. Computers are much simpler and the same program running on two computers with the same configurations will display consistent behaviour.

If we can give a precise characterisation of an information structure then we can measure its meaning in terms of the minimum amount of work necessary to generate or interpret it. There are profound mathematical results which say that it is impossible to make a precise measurement of the meaning of an arbitrary information structure. However we can make crude measurements which divide meanings up into broad classes.

CHAPTER 5

POLITICAL ECONOMY: VALUE AND LABOUR

Cottrell

5.1 SMITH AND WATT

Chapter 1 discussed the development of the physical concept of work by James Watt. It is probably no coincidence that Watt's colleague at the University of Glasgow, Adam Smith, was in the same period developing what would later be called the labour theory of value. We say 'probably' no coincidence because although we gather that Smith and Watt were friends and discussed intellectual matters together,[1] we don't know if there was any direct connection between Watt's development of the concept of work and Smith's conception of labour as the basis of value; this remains an intriguing speculation. Certainly Watt's work and Smith's labour are not the same thing—we have pointed this out above and we expand on the differences below—yet the abstraction is similar. As Smith remarks, "The greater part of people … understand better what is meant by a quantity of a particular commodity than by a quantity of labour. The one is a plain palpable object; the other an abstract notion, which, though it can be made sufficiently intelligible, is not altogether so natural and obvious." The "abstract notion" of labour as employed by Smith is not entirely new with him. His friend David Hume had written that "every thing in the world is purchased by labour" in his *Political Discourses* of 1752, and John Locke had hinted at a labour theory of value in the chapter on property in his *Of Civil Gov-*

---

[1] "Watt's workshop was a favourite resort of Smith's during his residence at Glasgow College, for Watt's conversation, young though he was, was fresh and original, and had great attractions for the stronger spirits about him" (Rae, 1965, p. 74).

ernment. But these earlier statements were undeveloped, and Smith was writing against the background of the 'natural law' tradition (imported to Glasgow via Gershom Carmichael's edition of Samuel Pufendorf's *De officio hominis et civis*): in that tradition value was analysed in terms of 'utility and scarcity' (Hutchison, 1988), and not, as in Smith, in terms of labour.

Smith began his career as a moral philosopher, particularly concerned with the analysis of human sympathy, but he later turned his attention to political economy and of course his magnum opus was *An Enquiry into the Nature and Causes of the Wealth of Nations* (1776). The opening sentence of this work announces a perspective in which labour plays a central role:

> The annual labour of every nation is the fund which originally supplies it with all the necessaries and conveniences of life which it annually consumes, and which consist always either in the immediate produce of that labour, or in what is purchased with that produce from other nations.

Smith is interested in the proportion this "produce" bears to "the number of those who are to consume it" (or real Gross Domestic Product per capita, as we might say today), and he remarks that

> this proportion must in every nation be regulated by two different circumstances; first, by the skill, dexterity, and judgment with which its labour is generally applied; and, secondly, by the proportion between the number of those who are employed in useful labour, and that of those who are not so employed.

We can think of this as the identity

$$\frac{\text{output}}{\text{population}} \equiv \frac{\text{output}}{\text{worker}} \times \frac{\text{workers}}{\text{population}}$$

where output per worker, or labour productivity, is governed by Smith's "skill, dexterity, and judgment".

The first three chapters of *The Wealth of Nations* are given over to a discussion of the division of labour, which Smith sees as the key to increasing labour productivity. (FIXME: say a bit more about this?) In a society where the division of labour has taken hold, individual producers do not produce their own subsistence; they produce a surplus, over their own requirements, of their own product, and rely upon others for articles they require but do not themselves produce. Smith takes for granted that the developed form of this interdependency is *commodity production* (the term is actually Marx's). That is, individual producers confront each other as independent property owners, and produce their respective goods as commodities, products destined for exchange via a market. In this respect Smith's argument is lacking in generality (as Marx would point out): commodity exchange via the market is one way—historically a very important way, to be sure—of organizing an economy based on a complex division of labour, but it is not the only way. The alternative is that the division of labour is planned, and that the goods produced by the specialized workers are *transferred* to their consumers rather than purchased by the consumers. This is the model followed in the division of labour within a peasant household or, on a larger scale, in the planned industrial economy that existed in the Soviet Union from the late 1920s till the late 1980s.

At any rate, talk of commodity exchange as a concomitant of the division of labour leads Smith to money in chapter IV of *The Wealth of Nations*, and thence to value. The term 'value', as applied to goods and services, has various meanings or shades of meaning. When we talk of a commodity being 'good value' or 'value for money' we mean that it has a favourable ratio of useful or desirable qualities to price. This corresponds to the first pole of the opposition Smith established, between 'value in use' (or use value) and 'value in exchange' (or exchange value).

> The word value, it is to be observed, has two different meanings, and sometimes expresses the utility of some particular object, and sometimes the power of purchasing other goods which the possession of that object conveys. The one may be called 'value in use'; the other, 'value in exchange'. The things which have the greatest value in use have frequently little or no value in exchange; and, on the contrary, those which have the greatest value in exchange have frequently little or no value in use. Nothing is more useful than water: but it will purchase scarce anything; scarce anything can be had in exchange for it. A diamond, on the contrary, has scarce any value in use; but a very great quantity of other goods may frequently be had in exchange for it.

In light of subsequent developments in modern economics, it is worth noting that for Smith (and for the classical political economists in general) 'value in use' seems to be understood as an objective category. Smith is perfectly confident in saying that water is highly useful and diamonds have little value in use; there is no suggestion that this could be a matter of 'individual tastes and preferences'. Even when objective, in the sense of being independent of individual tastes, value in use can depend on the situation. Which has the greater value in use, a hammer or a screwdriver? It's not a matter of opinion, but it depends on the task in hand. By contrast, modern economics has replaced the term 'value in use' by 'utility', and has cast utility not as a matter of the objective usefulness of goods but as a matter of the subjective 'psychic satisfaction' an individual derives from consumption of the good. This rather strange analysis would seem to apply best (if at all) to the highly refined luxury products of an advanced culture. Which has the greater utility, a novel by Charles Dickens or one by Jane Austen? A bottle of California Chardonnay or a Chablis? Here the satisfaction obtained by the individual is all we have to go on.

Although classical 'value in use' is not a subjective matter, it is clearly relative to the state of technology. We can infer from Smith's dismissal of diamonds as having "scarce any value in use", if we didn't know it already, that diamond-tipped drills were not in use for oil exploration in Smith's day.

Anyway, having made the distinction between use value and exchange value, Smith proceeds to concentrate on the latter. He sets himself three problems.

> In order to investigate the principles which regulate the exchangeable value of commodities, I shall endeavour to show:
>
> First, what is the real measure of this exchangeable value; or, wherein consists the real price of all commodities.
>
> Secondly, what are the different parts of which this real price is composed or made up.
>
> And, lastly, what are the different circumstances which sometimes raise some or all of these different parts of price above, and sometimes sink them below their natural or ordinary rate; or, what are the causes which sometimes hinder the market price, that is, the actual price of commodities, from coinciding exactly with what may be called their natural price.

In understanding these questions it is important to be clear on terminology. Smith's first question concerns the 'measure' of exchangeable value: he wants to know how exchange value is best measured or expressed. This is quite distinct from the question of the *determination* of exchange value.

Well, actually 'determination' can mean two things. It *can* mean measurement, as in 'How would you determine the height of that tree?' (By triangulation, perhaps.) Or it can mean causation: in this sense the height of the tree is determined by its genetic material (Is is a dogwood or a redwood?) and its environment (How much sunlight and water were available to it?). When we use the phrase 'determination of value' below we take it strictly in the second sense, to refer to the causal processes governing the exchange value of commodities.

Smith's second question (What are the different parts of which real price is made up?) relates to the determination of value, but note that he seems to prejudge the issue, taking for granted that exchange value is determined by an adding up of component parts. His third question introduces the important concept of 'natural price': this is the price that is just sufficient to call forth a supply of the product that meets the demand for it. In Smith's view natural price constitutes the "centre of gravitation" of actual, day-to-day market prices. To update Smith's Newtonian metaphor using the language of modern dynamics we might talk of natural price as an *attractor* for market price. We shall have more to say about this below.

5.2 LABOUR COMMANDED AS A MEASURE OF VALUE

The title of Smith's Chapter V—'Of the Real and Nominal Price of Commodities, or their Price in Labour, and their Price in Money'—tells us where he's headed on his first question. He is emphatic that the 'real' price of commodities must be measured by 'labour commanded'.

> Labour was the first price, the original purchase-money that was paid for all things. It was not by gold or by silver, but by labour, that all the wealth of the world was originally purchased; and its value, to those who possess it, and who want to exchange it for some new productions, is precisely equal to the quantity of labour which it can enable them to purchase or command.

In the day-to-day operations of a market economy it is 'natural' to express the exchange value of goods in terms of money: the money one would have to hand over to acquire the good, or that one could realize by selling it. But Smith argues this measure is superficial and potentially misleading. Superficial, because it does not take into account the point that the "real price of everything, what everything really costs to the man who wants to acquire it, is the toil and trouble of acquiring it." Potentially misleading, because money is not constant in its own value over time. A better measure of the exchangeable value of a commodity is the quantity of labour which it enables its possessor to "purchase or command". Smith's particular formulation of labour commanded is appropriate to an age when people (of a certain class, of course) were accustomed to hiring servants. Thus if I own a commodity with a market value of one guinea (twenty-one shillings), and if the labour of a servant can be had for one shilling per day, then with the money obtained by selling the commodity I can command the labour of a servant for three weeks. For the modern reader an alternative, equivalent version of Smith's calculation may seem more natural: the 'labour commanded' by a commodity represents the time you would have to work (say, at the average wage) in order to buy the commodity. In both cases—Smith's version and the modern one—the calculation of labour commanded is the price of the item divided by the average wage.

This is a good comparative measure of the cost of goods to a working consumer at widely separated points in time or across nations at a point in time, when the exchange rates of national currencies are a questionable guide to the respective purchasing power of the currencies in their home economies. Thus for instance a new Ford Model A car (4-door model) cost $570 in 1928, while a new Ford Escort 4-door cost about $11,000 in 2000. Is the Escort in 2000 really almost 20 times as costly as the 1928 model? Not in any meaningful sense. The average hourly wage for manufacturing workers was $0.56 in 1928, and $14.50 in 2000. If we take a working month to be 160 hours, this means that the labour commanded by the Model A in 1928 was 6.3 months, while the labour commanded by the Escort in 2000 was 4.7 months.[2]

Notice that the labour time required to produce a good and the labour it commands in exchange are not the same thing. Say a basic car in the USA today commands five months' labour at the average wage. What can we say about the labour time required to *produce* a car? Well, suppose that were also five months; in that case the average worker could work five months to obtain a commodity that embodies five months' labour. That is, his wage over the period would equal the value of the output he produces over the same period. But this means that the workers' wages would exhaust the

---

[2] The data in this paragraph were collected from The Bureau of Labor Statistics and *Collectibles Corner* for August 27, 1999, at www.krause.com.

value of the product—there would be nothing left over for profit. If the profit margin on car production is positive, it must be that wages per month are less the exchange-value produced per month, or in other words the labour-time required to make the car is less than the labour-time it commands. If the car commands five months' labour, it might take, say, three worker-months to produce. Further, the factors making for changes in the labour required to produce a commodity, and those making for change in the labour it commands, are not the same. A change in the wage rate will alter the amount of labour commanded by any commodity of given price, while it is changes in technology, not wages, that produce changes in the labour time required to produce things.

5.3 LABOUR TIME AND THE DETERMINATION OF VALUE

Having argued that labour commanded is the best measure of value, Smith turns in Chapter VI of *The Wealth of Nations* to "the Component Parts of the Price of Commodities".

> In that early and rude state of society which precedes both the accumulation of stock and the appropriation of land, the proportion between the quantities of labour necessary for acquiring different objects seems to be the only circumstance which can afford any rule for exchanging them for one another. If among a nation of hunters, for example, it usually costs twice the labour to kill a beaver which it does to kill a deer, one beaver should naturally exchange for or be worth two deer. It is natural that what is usually the produce of two days' or two hours' labour, should be worth double of what is usually the produce of one day's or one hour's labour.

Here we have the idea that the labour time required to produce a given product governs or determines the exchange value of the product. There is, Smith says, a 'natural' proportionality between required labour time and exchange value. He proceeds to qualify this idea, saying that labour which is harder, or requires more skill, will count for more than simple labour. But the more important qualification is the one he starts with in the quotation above: he confines the basic principle that exchange value reflects required labour time to an "early and rude state of society". Why does he do this?

Reading further, it seems that, for Smith, the distribution of the product of labour is the key factor. In the early and rude state, "the whole produce of labour belongs to the labourer; and the quantity of labour commonly employed in acquiring or producing any commodity is the only circumstance which can regulate" its exchange value. By contrast, in a developed market economy, where "stock [i.e. capital] has accumulated in the hands of particular persons", we have a state where

> the whole produce of labour does not always belong to the labourer. He must in most cases share it with the owner of the stock which employs him. Neither is the quantity of labour commonly employed in acquiring or producing any commodity, the only circumstance which can regulate the quantity which it ought commonly to purchase, command, or exchange for. An additional quantity, it is evident, must be due for the profits of the stock which advanced the wages and furnished the materials of that labour.

The profits of stock, says Smith, constitute a second "component part" of price, over and above the wages of labour. He then goes on to say that the rent due to the landlord constitutes a third component part of price. Exchange value can no longer be based on labour alone.

Smith has got into a muddle here. He seems to have persuaded himself that if the prices of commodities remained proportional to the labour time required to produce them then profit would be ruled out. But this doesn't follow at all. In a capitalist economy the exchange values of commodities cannot, in general, equal the *wages paid* in their production, else there would be no profit. But the propositions (a) that prices are proportional to the labour time required to produce things, and (b) that prices are equal to the wages paid in the production of things, are quite distinct: neither one implies the other.

Smith seems closer to getting it right when he writes, "The value which the workmen add to the materials … resolves itself … into two parts, of which the one pays their wages, the other the profits of their employer." That is, one can think of the value of a commodity as being determined by the labour time required to produce it, and then, as a distinct question, consider the 'resolution' or decomposition of this value into wages and profit. This was the position taken by David Ricardo, the first writer after Smith to make real progress in political economy.

CHAPTER 6

THE PROBABILISTIC APPROACH TO THE LAW OF VALUE

Wright

In this chapter we take a probabilistic approach to what was the foundational question of political economy: what is the relationship, if any, between the time it takes people to produce things and the prices they exchange for?

6.1 PROBABILISTIC MODELS

Once we know the possible outcomes of a situation it is natural to consider how *probable* each of those outcomes are. The probability of an event is a number in the interval $[0,1]$, where 0 represents an impossible event and 1 a certain event.

For example, if we perform a large number of coin tosses we soon discover that about half the outcomes are heads, and half are tails. So although we cannot predict the outcome of a particular flip, we can say that the outcome is equally likely to be heads or tails, or more precisely the probablity of heads or tails is one half, $\mathrm{P}(X = \text{heads}) = \mathrm{P}(X = \text{tails}) = \frac{1}{2}$, where X is the outcome of the toin coss.

Knowing that $\mathrm{P}(X = \text{heads}) = \frac{1}{2}$ means that about half the time a coin will land heads. In fact, this is a probabilistic prediction of the frequency of a particular outcome. It does not predict what will actually occur, but what will probably occur, given knowledge of the possible outcomes. Although weaker than a deterministic prediction a probabilistic prediction is still very useful for acting in the world. For example, knowing that an area has a high probability of earthquake activity tells us to build robust homes, even though we do not precisely know when an earthquake will strike.

The theory of probability is an appropriate tool for situations where we simply do not know the full range of causal mechanisms that determine the outcome of a situation, or cases in which we think we know what determines the outcome but in practice it is difficult to use our theories to make robust predictions. In such cases we give up on the idea of predicting actual outcomes but instead predict what will probably occur given the known possible outcomes. Instead of using a deterministic model to predict the *value* of a *variable* x, such as predicting whether $x = \text{heads}$ or $x = \text{tails}$, we use a probabilistic model to predict the *distribution* of a *random variable* X, such as predicting that $\mathrm{P}(X = \text{heads}) = \frac{1}{2}$ and $\mathrm{P}(X = \text{tails}) = \frac{1}{2}$, which is equivalent to stating that the distribution of X is uniform, that is all outcomes have an equal probability of occurring.

Consider the purchasing decisions of all individuals in the USA during one month. There are an enormous range of reasons why certain goods are sold in certain amounts for certain prices. Some goods are bought regularly in stable amounts, such as basic utilities like gas and water, other goods are ephemeral and their sales are contingent on transient fashions, such as the market for childrens' toys. The weather can affect sales. People are very different and have different goals and tastes. Some goods wear out periodically and may need to replaced. In sum, there are almost as many reasons for exchange events as there are events themselves.

The variability and contingency that necessarily occurs when complex and intelligent human beings competitively interact with each other implies that it is impractical to try to model market exchanges in detail. Although it is possible to model and predict human behaviour in controlled experimental settings that constrain the space of possible actions, or in situations where conventions or rules play an important part, it is not possible to model the everyday creativity of market participants aiming to satisfy their goals in open-ended and mutually constructed economic environments. It is evident that, ignoring special cases, predicting the actual price of a good on a particular day, or predicting the demand for a newly invented commodity type, is a lost cause. Prices and goods are always changing. A market economy is therefore an ideal candidate for probabilistic modelling.

6.1.1 *A simple exchange economy*

Recently physicists have turned their attention to economic phenomena, creating a new field called econophysics. Econophysics approaches to traditional economic problems are essentially probabilistic in nature. We can illustrate this approach by examining a very simple model of a market economy developed by the physicists Dragulescu and Yakovenko (2000).

Imagine a simple economy consisting of N people, which we shall call actors. Each actor has an amount of money m, which for the sake of concreteness let's assume is denominated in dollars. The total amount of money in the economy, which is simply the sum of all the individual money amounts held by each actor, is a fixed constant M.

In a market economy people exchange goods and services for amounts of money. But we'll completely abstract from the nature of those goods and services, the time they take to produce or complete, and who does what and when. We won't consider institutions either, so firms, banks and the economic operations of the state are out of the picture. Instead, we will focus on an essential characteristic of a dynamic monetary economy – the fact that money is continually exchanged between actors in different amounts but is almost always conserved.[1] We will not attempt to deterministically model all the local reasons why particular actors exchange particular amounts of money at particular instants of time, but instead assume that all this contingency can be modelled as random noise. Given these mighty abstractions a single rule can drive the dynamics of the simple model:

Exchange rule $\mathbf{E_1}$:

(1) Randomly pick an actor i ($1 \leq i \leq N$) according to a uniform distribution. Actor i is the buyer.

(2) Randomly pick an actor j according to a uniform distribution. Actor j is the seller.

(3) Randomly pick a price p from the interval $[0, m_i]$ according to a uniform distribution, where m_i is the amount of money held by the buyer i.

(4) Reduce the money held by i by p. Increase the money held by j by p.

---

[1] In 1994 scottish avant-garde artists Bill Drummond and Jimmy Cauty burnt one million pounds, earned from the sales of their pop records. But such events have low probability.

Economic change is simulated by repeatedly applying this rule to the economy of N actors. The rule transfers random amounts of money between randomly selected individuals. And that's all there is to it. Call this model the *simple economy* model. As mentioned, it is a very simple model. It is so simple it is perhaps difficult to believe that it can contribute much to our understanding. But in fact it is able to replicate one of the enduring and characteristic empirical regularities of market economies.

The number of actors with \$0, \$1, ..., M\$ in their pockets can be counted. Each dollar amount can be considered a 'bucket', and any particular actor at any particular time is 'in' one of these buckets, depending on how many dollars they hold. For example, if we initialise the model so that each actor has M/N dollars in their pocket, and then measure the size of each of the dollar buckets, we find that the money distribution is degenerate. Every bucket is empty, except for M/N, which is of size N. The distribution is called degenerate because there is only one possibility.

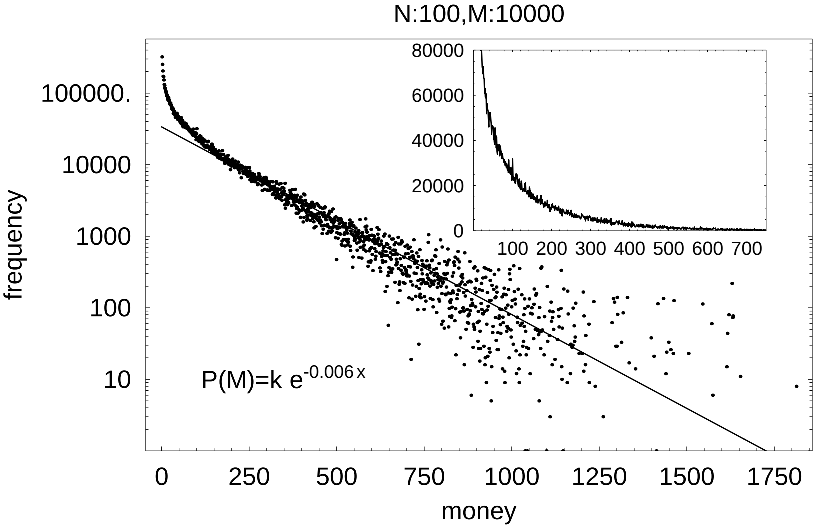


Figure 6.1: Stationary wealth distribution in a simple exchange economy plotted in linear-log scale. The straight line represents the Boltzmann distribution, $P(m) \propto \lambda e^{-\lambda m}$. The inset is a section of the wealth distribution in standard scale, which more clearly shows that most actors have very little wealth but exponentially few have a great deal.

But if rule $\mathbf{E_1}$ is repeatedly applied, let's say N^2 times, the distribution will begin to diverge from its degenerate state as money is exchanged in unequal amounts between the actors. Some actors will be lucky and enjoy a sequence of advantageous trades and obtain great wealth, while others may repeatedly spend money and get very little in return. If this process is continued the economy settles to a particular kind of distribution, illustrated in figure 6.1.1. This distribution is called an exponential distribution. The inset graph in figure 6.1.1 shows the money buckets along the x-axis and a count of the number of actors that were in that bucket over a period of time along the y-axis. The steep downward sloping graph can be fitted to an exponential function $\mathrm{P}(m) \propto \lambda e^{-\lambda m}$ as shown by the straight line fitted to the linear-log transformation of the data. The money distribution is highly unequal. The majority of actors have very little money, whereas exponentially few have a great deal. In fact, a very small number of individuals have relatively enormous amounts of money.

Remarkably the exponential distribution of wealth is found in real data from real economies. There is some divergence from an exponential distribution in the top 5%-10% of wealthy individuals, but an exponential distribution accurately describes the vast majority of the population (Nirei and Souma 2003b, Dragulescu 2003, Dragulescu and Yakovenko 2002), whichever advanced capitalist country is considered.[2] The distribution is also stable over long periods of time. Although mean wealth may change from one year to the next, the overall functional form of the wealth distribution remains exponential. In conclusion, the simple probabilistic model in spite of (or due to?) its high level of abstraction and simplicity has replicated an important feature of modern capitalist economies.

The probabilistic approach also provides some new economic insights, in particular the importance of statistical equilibria and entropy in economic phenomena.

6.1.2 The concept of a statistical equilibrium

The simple economy model illustrates the concept of a statistical equilibrium. Over time the distribution of wealth in the economy converges to an exponential distribution, and stays there. Even though the economic actors

---

[2] Some of the empirical studies use income over a time period as a proxy for instantaneous wealth, but the details do not matter here.

continue to exchange money, and ascend or descend the income scale, the overall distribution of wealth in the economy remains constant.

Contrast this kind of equilibrium with the better known concept of a mechanical equilibrium. For example, consider a set of weighing scales, 1kg on the left plate, and 1kg on the right: the scales balance, all forces equalise, and the arms are still. The system is is in mechanical equilibrium and will remain so until some external force is applied. A statistical equilibrium is a different kind of equilibrium. Unlike a mechanical equilibrium, in which the system configuration remains static, a system can be in statistical equilibrium even when its configuration is continually changing. It is the probability distribution over possible system configurations that remains constant over time. For example, if we sample the distribution of money in the simple economy over a period of time, and then repeat this experiment at a later time, the two distributions will be nearly identical with high probability, despite money continually changing hands. Therefore, unlike a mechanical equilibrium, there is always the possibility that a system in statistical equilibrium will spontaneously deviate from equilibrium. But the probability it will do so is small. For example, the probability that the simple economy will spontaneously return to its initial egalitarian wealth distribution is so vanishingly small it may as well be considered impossible.

The standard economic theories we have inherited from the twentieth century are deterministic models, following the path laid down by theorists in the nineteenth century who copied the tools and methods of the prevailing mechanical theories in the physical sciences (Mirowski 1989). The first definitive formulation of this approach is Debreu's short book *The Theory of Value* (Debreu 1959), in which a market economy is pictured as a huge deterministic calculator that computes a set of market exchanges between economic actors, agreeable to all, given initial endowments of goods. In this model the concept of a mechanical equilibrium is employed to understand the meaning of economic phenomena. But unlike mechanical configurations of matter, which do sometimes come to rest, a market economy never does: it is inherently a dynamic system, with economic actors whose agency continually upsets any possibility of the attainment of a mechanical equilibrium. A market economy is more like a bag of marbles vigorously shaken than a set of weighing scales at rest.

The simplest case of a statistical equilibrium analysed in the physical sciences is that of an ideal gas. An ideal gas consists of millions of identical particles enclosed in a container that is perfectly insulated. The volume and temperature of the gas are assumed to be constant. Every gas particle continually moves within the container, bouncing off the walls and other particles, changing direction, and gaining or losing speed depending on the local contingencies that determine collision outcomes. At the micro-level there is seeming chaos. Despite all the uncoordinated chaos, however, all the particles are connected to each other via the principle of the conservation of energy. Each collision conserves energy, therefore the total energy of the system is constant. Hence, if one particle is travelling unusually fast, and has a large kinetic energy, then this necessarily implies that some other particles must move at a slower speed. It is a physical impossibility that all particles have the highest kinetic energy at the same instant of time. In other words, there is a shared pool of available energy that is distributed amongst the gas particles. This total energy is a macro-level constraint on the micro-level disorder. All possible system configurations, that is possible distributions of kinetic energy amongst the gas particles, cannot violate this global constraint.

The fundamental law of equilibrium in statistical mechanics is the Boltzmann-Gibbs law, which states that the probability distribution of energy ϵ is $P(\epsilon) \propto \lambda e^{-\lambda\epsilon}$, where $1/\lambda$ is the temperature of the gas, or the average energy per particle. This is the exponential distribution once again. This is not too surprising when we consider that the simple economy model and the ideal gas are formally equivalent.

Simple economy	**Ideal gas**
Large number of identical actors	Large number of identical particles
Each actor has money m_i	Each particle has energy ϵ_i
Total money M is constant	Total energy E is constant
Exchange is money conserving	Collisions are energy conserving
Economy enters statistical equilibrium	Gas enters statistical equilibrium
Boltzmann-Gibbs money distribution	Boltzmann-Gibbs energy distribution
$P(m) \propto \lambda e^{-\lambda m}$	$P(\epsilon) \propto \lambda e^{-\lambda\epsilon}$
$1/\lambda$ is average wealth	$1/\lambda$ is average temperature

6.1.3 *The maximum entropy distribution*

In section 1.4 the second law of thermodynamics was introduced. The law states that the total entropy in a closed system tends to increase. The simple economy and ideal gas are closed systems. The second law implies that the equilibrium distribution, which we have seen is the exponential distribution, must be the distribution that has maximum entropy given the overall constraint on the total money in the economy (or the total energy of the gas). Let's check this. Consider the following entropy measure for the simple economy:

$$-\sum_{m=0}^{M} \mathrm{P}(m) \ln \mathrm{P}(m) \tag{6.1}$$

where $\mathrm{P}(m)$ is the probability that a randomly picked actor has money m. There are N actors in the economy and M dollars, both of which are conserved. Let n_m be the number of actors that hold m dollars. It is necessarily the case that:

$$\sum_{m=0}^{M} n_m = N$$

and

$$\sum_{m=0}^{M} n_m m = M$$

The probability that a randomly picked actor will have money m is $\mathrm{P}(m) = n_m/N$. If we substitute $n_m = \mathrm{P}(m)N$ into the above two equations we get two constraints on the probabilities:

$$\sum_{m=0}^{M} \mathrm{P}(m) = 1$$

which is the simple constraint that all probabilities must sum to one, and

$$\sum_{m=0}^{M} \mathrm{P}(m)m = \frac{M}{N}$$

which is the constraint that the probabilities must conform to the total wealth constraint.

The mathematical problem is to determine a formula for $\mathrm{P}(m)$ that meets the constraints and maximises the value of the entropy equation. This problem can be solved in a variety of ways, the details of which are unimportant. [3] But it turns out that the solution is indeed the Boltzmann-Gibbs (exponential) distribution $\mathrm{P}(m) \propto \lambda e^{-\lambda m}$. The exponential distribution of wealth

---

[3] The interested reader should consult Kapur (1989) and Kapur and Kesavan (1992).

is therefore the most *disorderly* distribution under the assumption that the only constraint on the system is conservation of money. Clearly, if the economic system were composed of more sophisticated agents such as 'economic demons', who, for example, formed coalitions or initiated joint plans in order to consciously change the income distribution, then new constraints on the probabilities would need to be considered, and the mathematical argument would change. But the fact that the majority of the empirical income distribution in capitalist economies is exponential suggests that such factors are not significant between individuals in the exponential regime of the income distribution.

In reality, unlike in the simple economy model, there are many schemes for money reallocation, for example limited redistribution of income via state taxes. But it is a surprising fact that such mechanisms do not affect the overall functional form of the income distribution. Markets appear to have a very robust tendency to maximise entropy, and generate highly unequal, predominately exponential income distributions.

We'll revisit the topic of income distribution in Chapter 9, where we'll discover that the full income distribution has lower entropy than the exponential distribution. So new causal factors, missing from this simple economy model, are at work, which place further constraints on the probabilities $\mathrm{P}(m)$. This implies that some kind of entropy-reducing demonic work is being performed to 'sort' money amongst different economic classes.

6.1.4 Random agents versus rational agents

It may be objected at this point that economic actors are clearly purposive and it is therefore essential to model individual rationality, even when considering macro-level phenomena, such as emergent income distributions. For instance, people do not exchange money according to random rules, and, depending on the amounts involved, often think very carefully about what they spend. But this objection confuses epistemology with ontology, a picture with reality. A random model need not imply that the causality it represents is random, only that it it is intrinsically difficult to model all the causality in perfect detail. The randomness is intended to represent *all* the many and varied rational (or otherwise) decisions of the economic actors.

The underlying assumption of the rational actor approach to economics is that macro phenomena are reducible to and determined by the mechanisms of individual rationality. Farjoun and Machover (1983) noted some time ago that the successful physical theory of statistical mechanics is in direct contradiction to this assumption. For example, classical statistical mechanics models the molecules of a gas as idealised, perfectly elastic billiard balls. This is of course a gross oversimplification of a molecule's structure and how it interacts with other molecules. Yet statistical mechanics can deduce empirically valid macro-phenomena. Quoting Khinchin (1949):

> Those general laws of mechanics which are used in statistical mechanics are necessary for any motions of material particles, no matter what are the forces causing such motions. It is a complete abstraction from the nature of these forces, that gives to statistical mechanics its specific features and contributes to its deductions all the necessary flexibility. ... the specific character of the systems studied in statistical mechanics consists mainly in the enormous number of degrees of freedom which these systems possess. Methodologically this means that the standpoint of statistical mechanics is determined not by the mechanical nature, but by the particle structure of matter. It almost seems as if the purpose of statistical mechanics is to observe how far reaching are the deductions made on the basis of the atomic structure of matter, irrespective of the nature of these atoms and the laws of their interaction. (Eng. trans. Dover, 1949, pp. 8–9).

The method of abstracting from the mechanics of individual rationality, and instead emphasising the particle nature of individuals, is valid because the number of degrees of freedom of economic reality is very large. We can picture individual decision making as a highly simplified random selection from possibilities constrained by overall macro-level principles, such as the conservation of money. At this level of abstraction, individual psychology can be modelled as extraneous noise.

Let's now consider a slightly more complex economy, in which the actors take time to produce different kinds of commodities, which are then exchanged against money.

6.2 A SIMPLE PROBABILISTIC MODEL OF A COMMODITY ECONOMY

6.2.1 The law of value

Marx, following Ricardo, held a labour theory of the economic value of reproducible commodities. According to Marx the value of a commodity is determined by the prevailing technical conditions of production and measured by the socially necessary labour-time required to produce it (Marx 1954). The value of a commodity is to be distinguished from its price, which is the amount of money it fetches in the marketplace. Although economic actors may differ in their subjective evaluations of the worth or 'value' of commodities there are emergent regularities in commodity economies that ensure that prices tend to 'gravitate' around labour values.

An important feature of Marx's theory of value is the strong distinction between value and price. Prices are what we see everyday in a market economy, but we never see values. In Marx's view the unknown and hidden real values of commodities constrain and shape the formation of commodity prices, whether we are aware of it or not, and despite the subjective evaluations we may form of the relative importance of the available goods and services. Prices are noisy and at any precise time are subject to multiple causes, not least the scarcity or abundance of goods, or the shifting tastes of the consuming public. Marx's theory of value is not intended as a direct explanation or prediction of particular prices on particular days, but abstracts from temporary or accidental conditions, and instead investigates a necessary determinant of price.

There is nothing unusual about this approach. In fact, the logical separation of different mechanisms that in practice mutually interact to cause an event to occur is a necessary part of scientific inquiry. For example, the law of gravity is a common and permanent factor that partially controls the movement of objects on earth. But the fact that books stay on shelves or planes fly does not invalidate the law; rather, the law explains the need and function of bookshelves and wings. And although the law of gravity cannot always be used to predict the trajectory of objects, it is nonetheless a real casual factor. Similarly, the law of value is a theory of a common and permanent factor that partially controls the movement of prices in commodity economies. The fact that a monopolistic firm may permanently over-charge for its services, or the price of non-reproducible goods, such as great works of art, appear to have no relationship to labour-time, does not invalidate the law of value.

It is important to develop theories of single mechanisms hypothetically considered to be working in isolation. Only then can we hope to predict actual events. Newton famously asserted, contrary to all appearances, that all bodies move at constant velocity unless an external force is applied. This is not an empirical statement, because apart from the odd special case, most bodies do not move at constant velocity. Simplification and abstraction is necessary in order to identify underlying, hidden causal mechanisms, particularly if the events that need to be explained, whether movement of bodies or movement of prices, are multiply determined by lots of different mechanisms working together.

In a theoretical simplification of capitalism often referred to as the 'simple commodity economy', Marx claims that prices will tend to 'correspond' to labour values. Only a few simple conditions need be met for this to occur:

> For prices at which commodities are exchanged to approximately correspond to their values, nothing more is necessary than 1) for the exchange of the various commodities to cease being purely accidental or only occasional; 2) so far as direct exchange of commodities is concerned, for these commodities to be produced on both sides in approximately sufficient quantities to meet mutual requirements, something learned from mutual experience in trading and therefore a natural outgrowth of continued trading; and 3) so far as selling is concerned, for no natural or artificial monopoly to enable either of the contracting sides to sell commodities above their value or to compel them to undersell. By accidental monopoly we mean a monopoly which a buyer or seller acquires through an accidental state of supply or demand.
>
> The assumption that the commodities of the various spheres of production are sold at their value merely implies, of course, that their value is the centre of gravity around which their prices fluctuate, and their continual rises and drops tend to equalise (Marx, 1972, p. 178).

The theory of the law of value motivates such statements. It is a fundamental building block of Marx's economics.

The law of value is intended to explain how the total labour of a society of commodity producers, who freely exchange their products in a marketplace, is divided and allocated to different branches of production via the market mechanism. The exchange of commodities at prices that deviate from values is the mechanism by which social labour-time is transferred from one sector of production to another. When prices equal values the division of labour has reached an equilibrium that satisfies social demand: 'the law of value is the law of equilibrium of the commodity economy' (Rubin 1973):

> [I]t is only through the 'value' of commodities that the working activity of separate independent producers leads to the productive unity which is called a social economy, to the interconnections and mutual conditioning of the labour of individual members of society. Value is the transmission belt which transfers the movement of working processes from one part of society to another, making that society a functioning whole (Rubin, 1928, p. 81).

In brief, the law of value is the process by which a simple commodity economy (i) reaches an equilibrium, in which (ii) prices correspond to labour values, and (iii) social labour is allocated to different branches of production according to social demand (where 'social demand' is understood to mean consumption requirements constrained by income).

We will investigate Marx's claim in some detail. The main result is that Marx's law of value does emerge as an unintended consequence of uncoordinated market activity. We will see how the law of value naturally emerges from 'behind the backs' of economic actors solely via money flows that place budget constraints on their local evaluations of commodity prices, which are otherwise subjective and unconstrained. The probabilistic model reveals particularly simple and satisfying dynamic relationships between values, prices, social labour-time and money.

It must be emphasised, however, that Marx did not think that prices correspond to labour values in capitalism. Instead, he thought there was a systematic deviation between labour values and profit-equalising 'prices of production'. But here we wish to exclude this complication and instead concentrate on a hypothetical case of the law of value operating in isolation.

6.2.2 The model

The model consists of a set of N economic actors (labelled $1,\dots,N$) that produce, consume and exchange a set of L commodity types (labelled $1,\dots,L$), a fixed amount of paper money M, which is distributed amongst the actors, a market mechanism that mediates commodity exchange.

For simplicity, we will assume that every actor specialises in the production of a single commodity at any one time. The current specialization of actor i is given by $A(i)$. All commodities are simple, and do not require other commodities for their manufacture. Each commodity requires the work of a single actor for its production. Constant returns to scale prevail and consequently there is no rationale for the existence of firms. Actors never cease production. A production column vector, $\mathbf{l} = (1/l_1,\dots,1/l_L)$, where $l_j > 0$, defines the rate at which an actor can produce each commodity type. For example, an actor that specialises in commodity type j produces at a rate of $1/l_j$ units per time step. Each l_j is the labour value of commodity j. The production vector is identical and fixed for all actors. Labour in the economy is therefore homogenous and is not subject to changes in technique. Once a commodity is produced it remains the property of the actor until consumed or exchanged. Each actor has an associated endowment vector that represents how much of each commodity is currently held.

Actors produce according to the following rule:

> Production update rule $\mathbf{P_1}$: (Deterministic). At the start of the simulation initialise the endowment vector for actor i to zero: $\mathbf{e}_i = \mathbf{0}$.
>
> Actor i subsequently generates one unit of commodity $A(i)$ every $l_{A(i)}$ time steps, and the appropriate element of the endowment vector, $\mathbf{e}_i[A(i)]$, is incremented by one.

Although no producer is more efficient than another a distinction between socially necessary labour-time and actual labour-time expended can be maintained. Overproduction of a commodity relative to the social demand implies that some of the labour-time expended was socially unnecessary.

Actor consumption

Every actor desires to consume all commodity types. This behaviour can be interpreted as subsistence or aspirational. A consumption column vector, $\mathbf{c} = (1/c_1,\dots,1/c_L)$, where $c_j \geq 0$, defines the desired rate of consumption events for all actors. For example, every actor desires to consume commodity j at a rate of $1/c_j$ units per time step. The consumption vector is identical and fixed for all actors and represents an economy with homogenous tastes that do not change. Note the asymmetry between production rates and consumption rates: an actor always meets its single production rate, but only conditionally meets its consumption rates. Actual consumption rates depend on the availability of commodities produced by other actors.

Actors consume according to the following rule:

> Consumption update rule $\mathbf{C_1}$: (Deterministic). At the start of the simulation initialise the consumption deficit vector for actor i to zero: $\mathbf{d}_i = \mathbf{0}$.
>
> Actor i subsequently generates one unit of consumption deficit for each commodity $j = 1, \ldots, L$ every c_j time steps, and the appropriate element of the deficit vector, $\mathbf{d}_i[j]$, is incremented by one.
>
> Each time step actor i consumes $\mathbf{o}_i = \min(\mathbf{e}_i, \mathbf{d}_i)$ commodities from its endowment to satisfy its current consumption deficit. A new endowment vector, $\mathbf{e}'_i = \mathbf{e}_i - \mathbf{o}_i$, and a new deficit vector, $\mathbf{d}'_i = \mathbf{d}_i - \mathbf{o}_i$ are formed.

Note that in each time step more than one commodity may be consumed, although only one commodity may be worked on. The assumption of universal and constant production and consumption vectors could be relaxed by introducing supply and demand noise due to heterogeneity of consumption tastes and production efficiency, but we won't pursue this extension.

The reproduction coefficient

The reproduction coefficient, $\eta = \sum_{j=1}^{L} l_j/c_j$, measures whether, given the 'social facts' of the production and consumption vectors, the economy may realise an overall social surplus, deficit or balance. A value of $\eta = 1$ implies the economy can achieve a state of simple reproduction (where total production equals total consumption), $\eta > 1$ implies an economy permanently in overall deficit (unrealised consumption capacity) and $\eta < 1$ implies an economy permanently in overall surplus (redundant production capacity). We will restrict our attention to economies with $\eta = 1$ that can theoretically achieve a balance between supply and demand but may over-and under-produce commodities due to a sub-optimal division of labour.

Money

Each actor i owns a sum of symbolic money, $m_i \geq 0$, which is used to purchase commodities for consumption. The total amount of money in the economy, $M = \sum_{i=1}^{N} m_i$, is conserved. The unit of measure of money is the 'coin', although it is an arbitrarily divisible unit. Coins are neither produced nor consumed by actors. Actors exchange money for commodities, and therefore gain money when they sell, and lose money when they buy. Complications due to changes in the money supply are ignored.

Subjective prices

Actors form subjective evaluations of commodity prices during bi-lateral exchange. Two requirements are placed on the evaluations: (i) a purchaser cannot offer more coins than they possess, and (ii) offer prices must not be fixed *a priori*. The second requirement is important because the law of value trivially does not hold in an economy of homogenous, *a priori* evaluators. For example, if every actor evaluated commodity j at 0 coins for all time then prices cannot converge to labour values. The law of value operates 'behind the backs' of economic actors because they adapt to changing local circumstances that are not of their own choosing but the result of global properties of the economic system.

To simulate adaptation we could select a machine learning algorithm, which has some psychological plausibility, that minimizes the consumption error. But this is an unnecessary level of detail at present. Instead, actors form selling and buying prices for each commodity according to:

> Price offer rule $\mathbf{O_1}$: (Stochastic). The price of commodity j according to actor i is $p_j^{(i)}$, and is randomly selected from the discrete interval $[0, m_i]$ according to a uniform distribution. The price is random but bound by the number of coins currently held.

The actors are adaptive in a weak sense: if they have less (resp. more) coins they probably will offer less (resp. more). Their changing circumstances are defined solely by how many coins they hold. The law of value, if it is to function, must therefore do so only via money flows, not by directly influencing or changing individual cost evaluations. $\mathbf{O_1}$ is one of many possible adaptive rules, but it is the simplest, and represents minimal theoretical commitment to the decision processes employed by actors in real economies. In addition, Gode and Sunder (1993) have shown that random traders with a budget constraint realise the same allocative efficiency as human actors under the same market discipline, so there is reason to believe that market structure plays a more important causal role than the individual rationality. Our aim is to concentrate on the structural determinations of the conditions under which evaluations take place, rather than the process of evaluation itself. Rule $\mathbf{O_1}$ assumes that, absent a decision theory, a range of possible decision outcomes are equally likely.

The market

Periodically actors meet in the marketplace. Trading behaviour continues until the market is cleared when for every commodity type there are either no buyers or no sellers. Commodities are bought and sold in single units. A cleared market does not imply that all needs are satisfied or all commodities sold.

> Market clearing rule $\mathbf{M_1}$: (Stochastic). Initialise the set of uncleared commodities to $C = \{j : 1 \leq j \leq L\}$.
>
> (1) Randomly select an uncleared commodity j from the set C according to a uniform distribution.
>
> (2) Form the set of candidate sellers S, which contains all actors with a desire to sell commodity j (i.e., $S = \{x : \mathbf{e}_x[j] > \mathbf{d}_x[j], 1 \leq x \leq N\}$). Select the seller s from S according to a uniform distribution.
>
> (3) Form the set of candidate buyers B, which contains all actors with a desire to buy commodity j (i.e., $B = \{x : \mathbf{d}_x[j] > \mathbf{e}_x[j], 1 \leq x \leq N\}$). Select the buyer b from B according to a uniform distribution.
>
> (4) If no seller or no buyer (i.e., $S = \emptyset \vee B = \emptyset$) then remove commodity j from C; otherwise, invoke market exchange rule $\mathbf{E_1}$ (see below).
>
> (5) Repeat until there are no remaining uncleared commodities (i.e., $C = \emptyset$).

Rule $\mathbf{M_1}$ matches buyers with sellers who then conditionally exchange coins for commodities according to:

> Market exchange rule $\mathbf{E_1}$: (Stochastic). Given a buyer b and seller s of commodity j with offer prices $p_j^{(b)}$ and $p_j^{(s)}$ respectively, determined by price offer rule $\mathbf{O_1}$, select the exchange price, x, from the discrete interval $[p_j^{(b)}, p_j^{(s)}]$ according to a uniform distribution. The exchange price is randomly selected to lie between the two offer prices.
>
> If the buyer has sufficient funds ($m_b \geq x$) then the transaction takes place. Actor b loses x coins and gains one unit of commodity j, and the appropriate element of its endowment vector, $\mathbf{e}_b[j]$, is incremented by one. Actor c gains x coins and loses one unit of commodity j, and the appropriate element of its endowment vector, $\mathbf{e}_s[j]$, is decremented by one.

Rules $\mathbf{M_1}$ and $\mathbf{E_1}$ do not represent a typical Walrasian market in which transactions take place at equilibrium after a process of extended price signalling or 'tatonnement'. Instead, transactions occur at disequilibria prices, commodities may go unsold, and the same commodity type may exchange for many different prices in the same market period. Further, commodities in oversupply may initially fail to sell only to find willing buyers at a later time, and commodities in undersupply may not necessarily realise a higher price. In sum, although the rules do implement short-term price signalling due to disequilibrium between supply and demand the detailed dynamics of this process are not straightforward, and can only be approximated by mathematical models that assume continuous price adjustment.

Division of labour

The set $A_j = \{i : 1 \leq i \leq N, A(i) = j\}$ contains those actors that specialise in the production of j. The set $D = \{A_j : j = 1, \ldots, L\}$ partitions the actors into production sectors and represents the total division of labour of the economy. The division of labour is dynamic because actors can change what they produce. Actors attempt to meet their consumption requirements but do not explicitly maximise wealth. They switch from one production sector to another according to the following rule:

> Sector-switching rule $\mathbf{S_1}$: (Stochastic). For actor i at the end of every nth period of length T time steps form the consumption error, defined as the Euclidean norm of the consumption deficit vector, $\|\mathbf{d}_i^{(n)}\|$. $\|\mathbf{d}_i^{(n)}\|$ is compared to the consumption

> error of the previous period $\|\mathbf{d}_i^{(n-1)}\|$. If $\|\mathbf{d}_i^{(n)}\| > \|\mathbf{d}_i^{(n-1)}\|$ then randomly select a new production sector from the available L according to a uniform distribution. In other words, if the consumption error has increased from the previous period then swap to a new sector.
>
> T is a constant multiple of the maximum consumption period, $\max(c_i)$, such that actors produce and have the opportunity to sell at least one commodity before sampling the consumption error and deciding whether to switch.

There are no switching costs. The result of all actors following rule $\mathbf{S_1}$ is to perform a parallel search over possible social divisions of labour. Dissatisfied actors randomly switch to new sectors in search of sufficient income to meet their consumption requirements.

Simulation rule

The cycle of production, consumption, exchange and reallocation of social labour proceeds according to the following rule:

> Simulation rule $\mathbf{R_1}$: Randomly construct production ($\mathbf{l}$) and consumption vectors ($\mathbf{c}$) for the economy, such that the reproduction coefficient $\eta = 1$. Allocate M/N coins to each of the N actors (the initial distribution does not affect the final outcome).
>
> (1) Increment the global time step.
>
> (2) For each actor invoke production rule $\mathbf{P_1}$.
>
> (3) For each actor invoke consumption rule $\mathbf{C_1}$.
>
> (4) Invoke market clearing rule $\mathbf{M_1}$.
>
> (5) For each actor invoke sector-switching rule $\mathbf{S_1}$.
>
> (6) Repeat.

The ruleset for the simple commodity economy

$$\mathbf{SCE} = \{\mathbf{R_1}, \mathbf{P_1}, \mathbf{C_1}, \mathbf{O_1}, \{\mathbf{M_1}, \mathbf{E_1}\}, \mathbf{S_1}, \}$$

defines the computational model. The implementation has five parameters: (i) the number of actors N, (ii) the number of commodities L, (iii) the amount of coins in the economy M, (iv) an upper bound, R, on the maximum possible consumption period, which is used to constrain the random construction of production and consumption vectors during initialisation, and (iv) a switching parameter C that is the constant multiple of the maximum consumption period required by sector-switching rule $\mathbf{S_1}$.

6.2.3 Simulation results

Computational models are suited to the detailed analysis of causal processes that are not amenable to straightforward mathematical treatment. The detailed supply and demand dynamics in this model are an example. But unlike mathematical proofs, which normally quantify over the whole parameter-space, the execution of a computational model is only a single sample of the parameter-space. It isn't practical to explore the entire parameter-space so the sampling process is biased toward subspaces that may be feasibly computed (for example, the time cost of the simulation rapidly increases with N), are realistic (for example, economies with a single coin are not considered) and conform to the requirements for the law of value to operate (for example, if the consumption period of a commodity j greatly exceeds the number of actors, i.e., if $R \gg N$, then the probability that a seller of j will find a buyer in the marketplace is low; hence exchange becomes occasional, failing a requirement for the law of value to operate). All simulation runs follow a similar pattern of initial non-equilibrium activity prior to settling down to stable averages and stationary distributions (appendix B contains further experimental details). We will measure the stationary distributions of the division of labour and market prices. But many other variables of interest could be tracked.

Division of labour

The distribution of actors in each sector of the economy settles to a normal distribution centred on a mean sector size. Figure 6.2 shows the stationary distributions of a typical sample. The equilibrium mean size of sector j is always approximately $N(l_j/c_j)$. Figure 6.3 reveals this relationship sampled over many runs.

Definition 1. A division of labour is *efficient* if for every commodity type the number of commodities produced equals the social demand.

Proposition 1. Let $a_j = |A_j|/N$ be the proportion of actors producing commodity j. Then $a_j = \frac{l_j}{c_j}$ $(j = 1, \ldots, L)$ is an efficient division of labour.

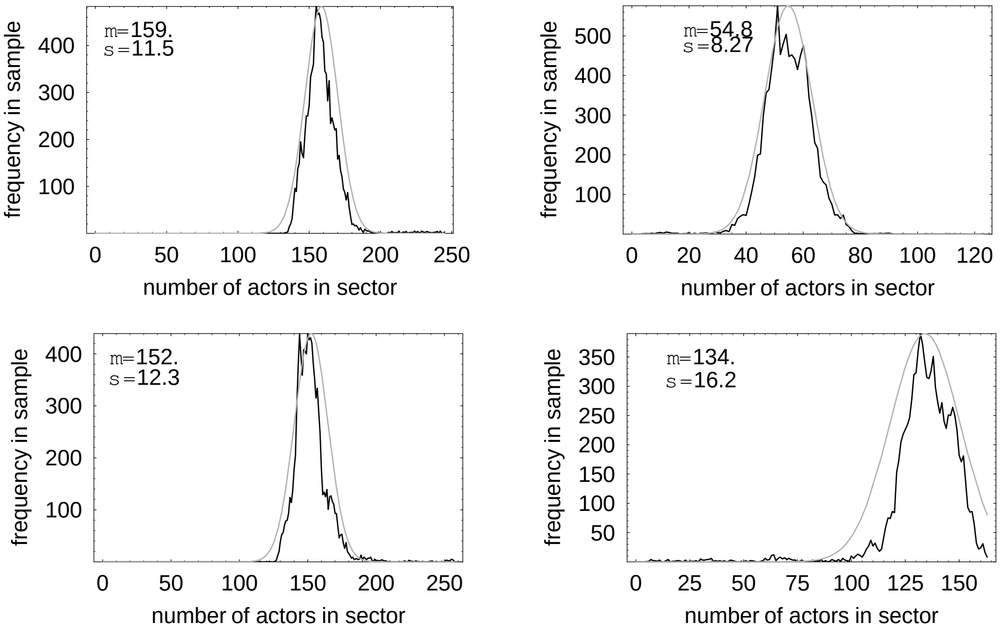


Figure 6.2: Stationary distributions of sector sizes with fitted normal distributions collected from a random sample of a 4-commodity economy with parameter settings N:500, L:4, M:2.5×10^5, R:25, C:2. The mean division of labour, $(159, 54.8, 152, 134)$, is close to the theoretical efficient division of labour, $N(\frac{l_i}{c_i}) = (152, 56, 146, 146)$.

Proof. The social demand for commodity j is $\frac{N}{c_j}$ units per unit time. When $a_j = \frac{l_j}{c_j}$ the number of units produced is $N\frac{a_j}{l_j} = \frac{N}{c_j}$ units per unit time, which equals the social demand. ☐

On average the division of labour is approximately efficient, but due to stochastic fluctuations perfect efficiency is not achieved. An efficient division of labour implies that the global consumption error is minimised and all actors meet their consumption requirements. Actual simulation runs only approximate maximum consumption, and unsold commodities and unsatisfied demands either stabilise or slowly accumulate over time. The results show that the **SCE** attains a (dynamic) equilibrium of the division of labour, and that the labour equilibrium is approximately efficient.

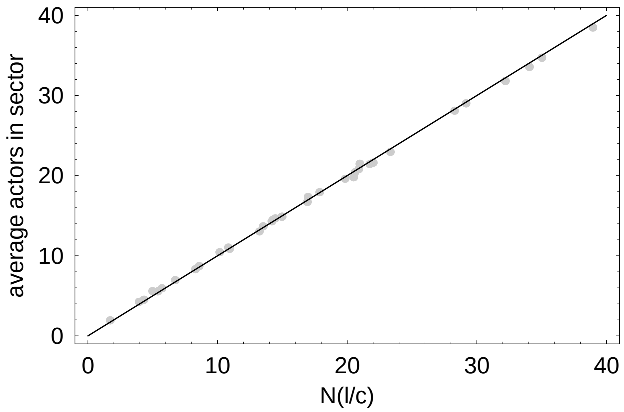


Figure 6.3: Relationship between mean sector size and $N\frac{l_i}{c_i}$ from 20 random samples of 3-commodity economies with parameter settings N:50, L:3, M:2500, R:25, C:2. The straight line represents the identity relationship $y = x$.

Objective prices

The stationary distributions of commodity prices can be approximately fitted by exponentials. Figure 6.4 shows the evolution of mean prices during a typical run and the associated stationary distributions. The price distributions have an exponential tail at the high end, but drop to zero at the low end, but the exponential distribution accurately models the price distributions over most of the price range. In equilibrium a single commodity type does not have a single price, but has a range of prices that occur with differing but fixed probabilities.

The law of value states that, in equilibrium, market prices ‘correspond’ to labour values. The Pearson correlation coefficient, r, between two vectors, $\mathbf{x}$ and $\mathbf{y}$ measures the linear relationship between them ($-1 \leq r \leq 1$). A value of -1.0 is a perfect negative (inverse) correlation, 0.0 is no correlation, and 1.0 is a perfect positive correlation. $r = 1.0$ implies that there is a single scalar constant, λ, such that $\mathbf{x} = \lambda\mathbf{y}$. We will check the correspondence between market prices and labour values by measuring their correlation.

Denote the average price of commodity j by $\langle p_j \rangle$. Figure 6.5 graphs representative time series of the correlation between the market price column vector $\mathbf{p} = (\langle p_1 \rangle, \ldots, \langle p_L \rangle)$ and the labour values column vector $\mathbf{v} =$

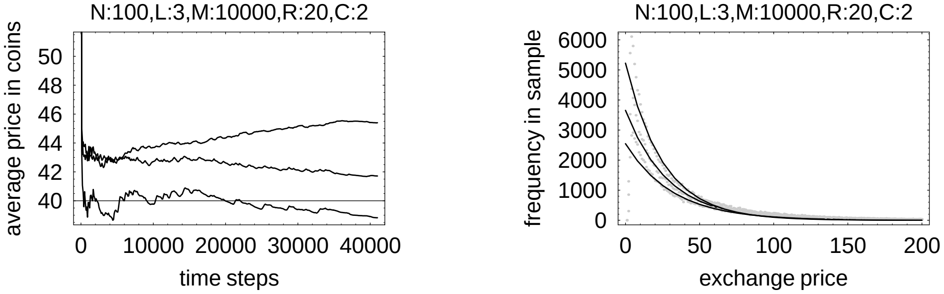


Figure 6.4: Evolution of mean commodity prices in a 3-commodity economy (left figure) and stationary distribution of commodity prices with fitted exponential distributions (right figure).

$(l_1, \ldots, l_L)$ (recall that l_j is the time period required to produce commodity j). We can now state the main simulation result of this chapter: the correlation between mean market prices and labour values approaches unity in equilibrium. Table 1, in the appendix, contains further experimental results that demonstrate the robustness of this result.

The results show that the **SCE** attains a (dynamic) equilibrium in which the mean equilibrium price of a commodity, measured over a sampling period, is proportional to the labour-time required to make it. Prices 'gravitate' around labour values and this equilibrium coexists with local and subjective pricing decisions constrained only by money endowments.

The equilibrium constant of proportionality, λ, between mean prices and labour values, such that $\mathbf{p} \approx \lambda \mathbf{v}$, must have dimensions *coins per unit labour-time*. λ summarises the causal relationship between expenditure of labour-time in production and the representation of that time in the market price of commodities. It measures how much labour-time money represents. Duménil (1983) and Foley (1982) emphasise the importance of this constant in Marxist economic theory. They define it in the context of a capitalist economy.

Definition 2. The *Monetary Expression of Labour-time* (MELT) is the ratio of the net product at current prices relative to the productive labour expended in an economy over a given period of time.

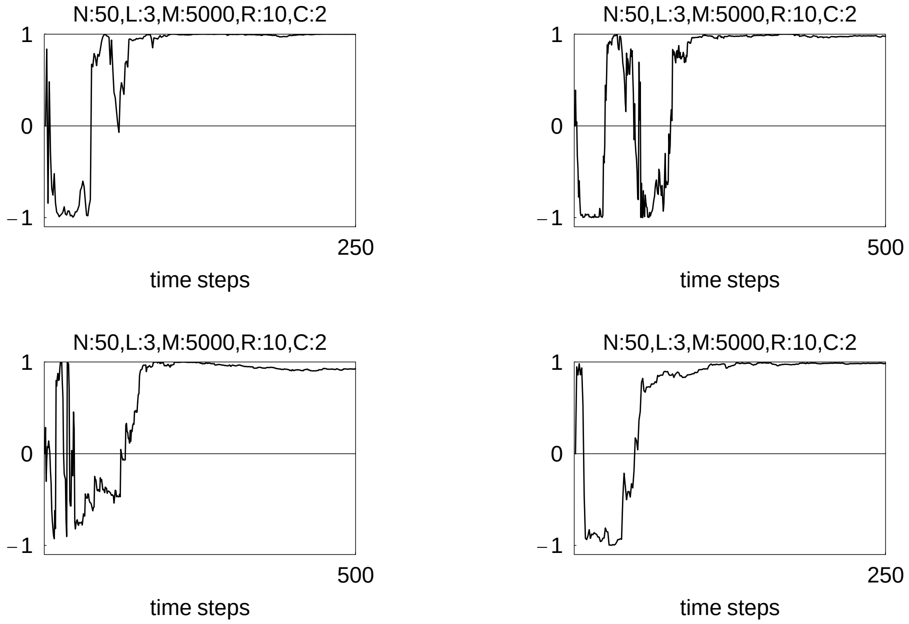


Figure 6.5: Evolution of vector correlation of mean prices and labour values over four samples of 3-commodity economies.

In a simple commodity economy there is no distinction between gross and net product and hence the MELT is the ratio of the product at current prices relative to the labour expended, which can be directly measured as:

$$\lambda = \frac{\gamma M}{\sum_{i=1}^{L} l_i v_i} \tag{6.2}$$

where γ is the proportion of the total money in the economy that on average exchanges per unit time (so γM is the average velocity of money), and v_j is the average exchange velocity of commodity j. The numerator in the definition is the rate of money exchange, the denominator is the rate of labour-time exchanged in the form of commodities, and the MELT is the ratio of the two, measured in coins per unit of labour-time. This definition translates into a computational rule to sample λ that executes per application of rule $\mathbf{R_1}$. The mean velocities of commodities and coins are calculated as historical averages. Figure 5 plots equilibrium mean prices, $\langle p_j \rangle$, against labour values multiplied by the MELT, λl_j, for a typical run of a 10-commodity economy. It demonstrates that the MELT is the constant of proportionality implied by the correlation results. The role of money as a representation of labour-time is particularly clear in this relationship.

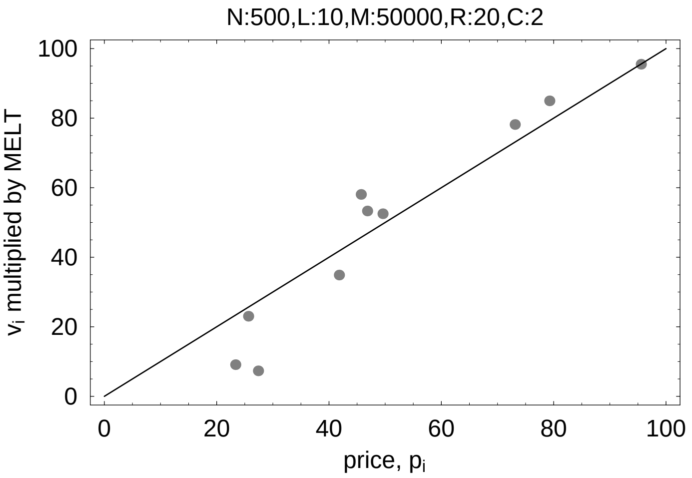


Figure 6.6: Stationary market prices and MELT transformed labour values in a 10-commodity economy with $r = 0.96$. The straight line represents the identity relationship $y = x$.

The definition of the MELT is not a causal theory of how the MELT is determined. The value of MELT will vary under different ‘institutional’ arrangements, such as how the market operates in detail, what kind of money and commodity throughput obtains, and so forth. Unlike the venerable quantity theory of money $MV = PT$ (where M is money, V is money velocity, P is the price level, and T the level of transactions), which is an accounting identity between market phenomena, the MELT abstracts a non-obvious causal relationship between non-market phenomena (production times) and market phenomena (prices).

6.2.4 Analysis

The results of the simulation experiment demonstrate that (i) a (dynamic) equilibrium is reached, in which (ii) mean prices are linearly related to labour values by a constant of proportionality called the monetary expression of labour-time (MELT), and (iii) social labour is allocated approximately efficiently. The computational model generates these regularities, but it does not provide an adequate explanation of them. The law of value *emerges* from dynamic interactions of the constituent parts of the **SCE**, but a theory is required to explain this emergence.

The qualitative theory of the law of value was most fully developed by Isaak Rubin in his 1928 book, *Essays on Marx's Theory of Value*. In what follows, the **SCE** is modelled by a system of ordinary differential equations that refer to the means of the variables of interest, thereby extending Rubin's theory.

The mathematical analysis aims to provide an intuitive explanation of the gross causal features of the computational model rather than provide definitive proofs of its properties or develop an accurate stochastic theory of the steady distributions. The mathematical model is a derivative and highly simplified analysis of the causal properties of the computational model. For example, discrete change is approximated by continuous change under the assumption that the size of discrete variables in the computational model is large compared to their change in magnitude per time step.

The labour equation

Let's think about what is happening when we run the simulation. We know that the rate money enters and leaves the market, or money velocity, is a proportion of the total money in the economy, which we will denote as γM $(0 \leq \gamma \leq 1)$. Assume that γ is fixed constant (an approximation). A money allocation column vector, $\mathbf{b}(t) = (b_1, \ldots, b_L)$, where $\sum_{j=1}^{L} b_j = 1$ and $0 \leq b_j \leq 1$, represents the instantaneous proportion of the total money flow received by each sector at time t. The sectoral income rate is therefore given by $b_j \gamma M$.

The labour allocation column vector, $\mathbf{a}(t) = (a_1, \ldots, a_L)$, where $a_j = |A_j|/N$ (see section 6.2.2), $\sum_{j=1}^{L} a_j = 1$ and $0 \leq a_j \leq 1$, represents the proportions of actors 'employed' in each sector at time t.

Use the mean price of a commodity to approximate its price distribution. Recall that the average price of commodity j is $\langle p_j \rangle$. The average cost of the universal commodity bundle, given current prices, is then $\sum_{j=1}^{L} \langle p_j \rangle / c_j$.

Actors switch sectors based on the consumption error, which is a function of the quantities of commodities received. To simplify the analysis we will use price signals, in the form of the mismatch between income and the average cost of the commodity bundle, as a proxy for the consumption error. This simplifying assumption holds for the remainder of the analysis.

Each sector has an ideal expenditure rate that represents the money that would need to be spent in order for the constituent actors to meet their desired consumption rates. The rate is a function of the number of actors in the sector and current prices, and is given by: $a_j N \sum_{k=1}^{L} \langle p_k \rangle / c_k$.

The sectoral income error, denoted ϕ_j, measured in coins per unit time, is the difference between the actual income rate and the ideal expenditure rate:

$$\phi_j(t) = b_j \gamma M - a_j N \sum_{k=1}^{L} \frac{\langle p_k \rangle}{c_k}$$

A value of $\phi_j > 0$ implies a sectoral 'profit' (the sector receives more income than its constituent actors require to purchase the commodity bundle), $\phi_j < 0$ implies a sectoral deficit (there is insufficient income for the actors employed in the sector to purchase the commodity bundle), and $\phi_j = 0$ implies sectoral income equals ideal expenditure.

Approximate the switching behaviour of actors by assuming that the rate of change of labour allocation (or sector size) is proportional to the sectoral income error:

$$\frac{d}{dt} a_j = \psi \phi_j(t) = \psi(b_j \gamma M - a_j N \sum_{k=1}^{L} \frac{\langle p_k \rangle}{c_k}) \tag{6.3}$$

where $\psi > 0$ is a reaction coefficient. It follows from the definition that $\phi_j < 0$ implies a net decrease in the sectoral population, and $\phi_j > 0$ a net increase, subject to the constraint $\sum_{j=1}^{L} a_j = 1$. Call (6.3) the *labour equation* because it defines how the allocation of labour to different sectors of production changes according to the money income received from the sale of commodities. The labour equation for the whole economy in vector notation is:

$$\dot{\mathbf{a}} = \psi(\gamma M \mathbf{b} - N(\mathbf{p} \cdot \mathbf{c})\mathbf{a}) \tag{6.4}$$

where $\mathbf{p} \cdot \mathbf{c}$ is the dot product of the average price vector and the consumption vector.

The production rate for commodity j is given by $a_j N / l_j$. Define the average price of a commodity to be the current sectoral income rate divided by the sectoral production rate:

$$\langle p_j \rangle = \frac{\gamma M}{N} \frac{b_j}{a_j} l_j \tag{6.5}$$

Hence, each $\langle p_j \rangle$ is a function of a_j and b_j.

The money equation

The labour equation describes how the division of labour changes depending on current incomes. But we do not yet have a model of how changes in incomes depend on the current division of labour.

A sector's income depends on the number of commodities produced. The maximum possible social consumption rate or 'social demand' for commodity j is N/c_j. The sectoral 'production error', denoted ξ_j, measured in units of commodity j per unit time, is the difference between supply and demand:

$$\xi_j(t) = \frac{a_j N}{l_j} - \frac{N}{c_j}$$

A value of $\xi_j > 0$ implies over-production, $\xi_j < 0$ implies under-production, and $\xi_j = 0$ implies supply equals social demand. It is assumed that market rule $\mathbf{M_1}$ operates such that it can be approximated by the expected relationship between supply, demand and price: commodities in over-supply have lower average prices than those in under-supply. This implies that the rate of change of sector income is negatively proportional to the production error:

$$\frac{d}{dt} b_j = -\omega \xi_j(t) = -\omega N \left(\frac{a_j}{l_j} - \frac{1}{c_j}\right) \tag{6.6}$$

where $\omega > 0$ is a reaction coefficient. It follows from the definition that $\xi_j < 0$ implies an increase in sectoral income, and $\xi_j > 0$ a net decrease, subject to the constraint $\sum_{j=1}^{L} b_j = 1$. Call (6.6) the *money equation* because it defines how the allocation of money to different sectors of production changes according to the over or under-production of commodities. The money equation for the whole economy in vector notation is:

$$\dot{\mathbf{b}} = -N\omega(\mathbf{A}\mathbf{l} - \mathbf{c}) \tag{6.7}$$

where $\mathbf{A}$ is the L by L diagonal matrix with element (i,i) equal to a_i and element (i,j) ($i \neq j$) zero.

Equilibrium

The $2L$ labour (6.4) and money (6.7) equations mutually interact and describe the evolution of the division of labour via the mechanism of market price changes. The causal schema is as follows: (i) an existing division of labour results in (ii) over and under-production of commodities that causes (iii) error-correcting price changes on the market due to supply and demand, which (iv) generate changes in sectoral incomes that (v) cause actors that cannot meet their consumption requirements to swap sectors, resulting in (vi) a new division of labour. Some mathematical results are now derived that show that the mutual interaction results in an equilibrium point at which prices equal labour values.

Definition 3. A *simple commodity system* is described by the following system of $2L$ coupled differential equations:

$$\dot{\mathbf{a}} = \psi(\gamma M\mathbf{b} - N(\mathbf{p}\cdot\mathbf{c})\mathbf{a}) \tag{6.8}$$

$$\dot{\mathbf{b}} = -\omega N(\mathbf{A}\mathbf{l} - \mathbf{c}) \tag{6.9}$$

and

$$\langle p_j \rangle = \frac{\gamma M}{N}\frac{b_j}{a_j} l_j$$

subject to the constraints

$$\begin{array}{ll}
\sum_{j=1}^{L} a_j = 1, & 0 \leq a_j \leq 1 \\
\sum_{j=1}^{L} b_j = 1, & 0 \leq b_j \leq 1 \\
\sum_{j=1}^{L} \frac{l_j}{c_j} = 1 = \eta & l_j, c_j > 0 \\
M, N > 0 & \omega, \psi > 0 \\
0 \leq \gamma \leq 1 &
\end{array}$$

Lemma 1 (Equilibrium point)**.** The simple commodity system has the unique equilibrium point

$$\mathbf{a}^* = (\frac{l_1}{c_1}, \frac{l_2}{c_2}, \ldots, \frac{l_L}{c_L}) = \mathbf{b}^* \tag{6.10}$$

(The proof is in appendix A.)

This lemma states that $\dot{a}_j = \dot{b}_j = 0$ (i.e., the system is at rest) when the proportion of actors employed in a sector equals the proportion of money received by the sector, and that proportion is l_j/c_j. This makes intuitive sense: every actor consumes the same consumption bundle, therefore, on average, they require the same income (otherwise actors move to different sectors and the system is not at rest). The lemma does not imply that every actor receives the same income in equilibrium, only that sectoral averages are equal. (In fact, the stationary income distribution in the **SCE** is highly unequal and approximately exponential).

Lemma 2 (Global stability)**.** The equilibrium point is globally asymptotically stable. (The proof is in appendix A.)

This lemma states that the system, regardless of its initial conditions, always approaches the equilibrium point. The simple commodity system is a feedback system that functions to minimise both income and production 'errors'. This formalises Rubin's assertion that '[a] given level of market prices, regulated by the law of value, presupposes a given distribution of social labour among the individual branches of production. ... Marx speaks of the "barometrical fluctuations of the market prices". This phenomenon must be supplemented. The fluctuations of market prices are in reality a barometer, an indicator of the process of distribution of social labour which takes place in the depths of the social economy. But it is a very unusual barometer; a barometer which not only indicates the weather, but also corrects it' (Rubin (1973, p. 78)). Lemma 2 explains why simulation runs tend to equilibrium.

Corollary 3 (Efficient division of labour)**.** The division of labour is efficient in equilibrium.

Proof. By lemma 1 the proportion of actors in sector j at equilibrium is $a_j = \frac{l_j}{c_j}$, which by proposition 1 is efficient. □

Corollary 3 is an explanation of why the simulation tends to an approximately efficient division of labour. The experimental results do not exhibit perfect efficiency because the **SCE** is non-deterministic and undergoes stochastic fluctuations in equilibrium.

Theorem 4 (The law of value)**.** Labour values are global attractors for average market prices.

$$\lim_{t \to \infty} \mathbf{p}(t) = \lambda \mathbf{v} \tag{6.11}$$

Proof. Substituting the equilibrium point, $a_j = l_j / c_j = b_j$, into (6.5) yields $\langle p_j \rangle = \lambda l_j$, which by lemma 2 is the globally asymptotically stable market price. □

At equilibrium the average price of a commodity is proportional to the labour-time required to make it. The constant of proportionality, $\lambda = \gamma M/N$, represents the monetary value of one unit of labour-time. Theorem 4 accounts for the observed correlations between prices and labour values.

In equilibrium actors receive equal mean incomes but are engaged in productive activity of unequal periods. Hence, commodities that take longer to produce sell for higher mean market prices. This is the fundamental reason why prices correspond to labour values at the equilibrium of the simple commodity economy.

Disequilibrium deviation of price from value

A key insight of Marx's theory of the law of value is that prices *refer* to amounts of labour time and *deviations* of prices from values are social *error signals* that function to redistribute labour. Only in the hypothetical situation of balanced supply and demand in which labour is efficiently distributed are prices proportional to labour values. We can analyse the deviation of price from value out of equilibrium by introducing the concept of labour commanded.

Definition 4. A commodity commands an amount of labour in exchange. The *labour commanded* by a commodity is its money price divided by the MELT, measured in units of labour-time. The mean labour commanded

$$\langle \varepsilon_j \rangle = \frac{\langle p_j \rangle}{\lambda} \tag{6.12}$$

represents how much social labour-time a commodity on average fetches in the marketplace.

If a commodity type commands an amount of social labour $\varepsilon_j < l_j$ then it is *undervalued*, if it commands amount $\varepsilon_j > l_i$ it is *overvalued*. The labour commanded is an objective property of the exchange, and is distinct from any subjective valuations of the utility of the transaction from the perspective of a particular actor. At equilibrium $\langle \varepsilon_j \rangle = l_j$ for all $j = 1, \ldots, L$ but otherwise commodities sell below value or above value, in accordance with the laws of supply and demand.

An act of exchange involves more than swapping of a commodity for an amount of money. It is also an exchange of a representation of an amount of social labour-time, measured by the labour commanded, for an amount of private labour-time actually expended in the production of the commodity. Normally this is not an exchange of equivalents.

If the global division of labour mismatches the social demand then labour associated with scarce commodities is rewarded with access to additional social labour-time, whereas labour associated with unwanted commodities is punished by a reduction of access. Out of equilibrium not all private labours are mutually equalised and not all private labours are socially necessary. But if the reallocation of labour resources is based on these monetary reward signals then the feedback loop completes and a division of labour emerges in which unnecessary private production is minimised and prices approach labour values. The dynamic relationship between labour embodied and labour commanded as regulator of the division of labour is apparent in the following relationship

$$\dot{a}_j = a_j \left(\frac{\langle \varepsilon_j \rangle}{l_j} - 1 \right) \psi \gamma M \tag{6.13}$$

which is derived in the appendix. The term in brackets is positive if the commodity type is overvalued (implying an increase in the sector size) and negative if the commodity type is undervalued (implying a decrease in the sector size). Equation (6.13) reveals the causal connection between labour allocation and prices that occurs under the surface of the simple commodity economy. It is a precise formulation of Rubin's observation that 'value is the transmission belt that transfers the movement of working processes from one part of society to another, making that society a functioning whole' (Rubin (1973, p. 81)) that summarises how the interaction of private commodity producers, using a monetary representation of the total social labour-time, spontaneously allocates labour to different branches of production according to social demand.

The precise price distribution will be sensitive to the particular price offer rule (or rules) employed by the actors. The more important point, therefore, is that in statistical equilibrium the same commodity type realises a range of different market prices, $p_k^{(1)}, p_k^{(2)}, \ldots$, each of which represents different transfers of social labour-time between buyer and seller. The role of the mismatch between labour embodied and labour commanded in regulating the division of labour is apparent 'on average' and is a property of the price distributions, not a property of individual transactions. Hence, a commodity type may be correctly valued in equilibrium while, at the same time, particular transactions may represent under or overvaluations of the commodity instance. The law of value states that, whatever the precise distribution of exchange prices, mean equilibrium prices are proportional to labour values.

6.2.5 Discussion

The choice of modelling symbolic money (e.g. paper or coins), which has nominal but no intrinsic value, rather than money in the form of a commodity such as gold, which has intrinsic value in virtue of the labour required for its production, differs from Marx's presentation but has the advantage of separating two definitions that may be easily conflated in his analysis of money (for a discussion, see Foley 1983): (i) the 'value of money', which is the inverse of the MELT and is the labour-time represented by the monetary unit (e.g. 1 hour of social labour-time is represented by 1 coin), and (ii) the 'value of the money commodity', which is the amount of social labour-time required for the production of a unit of the money commodity (e.g. 1 ounce of gold requires 1 hour of social labour-time for its production).

Roemer (1982, pp. 27–31) argues that in a simple commodity economy the only prices capable of reproducing the system are those proportional to embodied labour times. The derived prices satisfy the constraints of the economic situation represented as a linear programming problem. The deduction abstracts from market interactions that occur in historical time and disequilibrium supply and demand dynamics; hence, the mechanism by which such prices are reached is absent. The model is constraint-based rather than causal. The idea that labour values are *attractors* for prices in the simple commodity economy does not contradict this static result. A dynamic analysis, however, is a more stringent test of the conceptual integrity of the Marx-Rubin law of value, which is essentially concerned with how markets function to allocate social labour-time via error-correcting price signals. In static models, such as Roemer's, prices are nominal and lack a casual connection with the reallocation of labour. The mechanism of the law of value should not be reduced to its attractor.

Krause (1982) understands the importance of the dynamic coordination of concrete labours in market economies via the price mechanism. He contends that most modern formulations of the labour theory of value assume that concrete labours of different types are equivalently valued, an assumption he labels 'the dogma of homogenous labour' (Krause (1982, pp. 160–161)). According to Krause, the 'supposition of homogenous labour supplants any analysis of the specific coordination of concrete labours' (Krause (1982, p. 101)). The static methods employed by Krause, which represent the economic situation in terms of linear algebra, are sophisticated, and can quantify over complex production structures, in particular the production of commodities by means of others. In contrast, the dynamic approach taken here is relatively immature and models a simple production structure. Unlike static approaches, however, the dynamic approach can model the coordination of concrete labours, and this reveals a new dynamic relationship between homogenous and heterogenous labour.

Following Krause let α_{ij} be the reduction coefficient of concrete labour type i to j ($i \neq j$), such that 1 hour of labour of type i is equivalent to α_{ij} hours of labour type j, where the equivalence relation is induced by market exchanges. The assumption of homogenous labour is that $\alpha_{ij} = 1$ for all i and j. The reduction coefficients in the simple commodity system are:

$$\alpha_{ij} = \frac{\langle p_i \rangle / l_i}{\langle p_j \rangle / l_j} = \frac{b_i / a_i}{b_j / a_j}$$

Note that the assumption of homogenous labour is not made. Theorem 4 can be reformulated as

$$\lim_{t \to \infty} \frac{b_i}{a_i} = 1$$

and by the quotient rule for limits it follows that for all $i = 1 \ldots L$ and $j = 1 \ldots L$

$$\lim_{t \to \infty} \alpha_{ij} = 1 \tag{6.14}$$

The statement that labour values are attractors for prices is equivalent to the statement that homogenous reduction coefficients are attractors for heterogenous reduction coefficients. Krause writes: 'It is conceivable that certain assumptions about the mechanism of coordination could *produce* equal reduction coefficients. But the classical/contemporary labour theory of value does not formulate such assumptions, so the homogeneity is mere dogma' (Krause (1982, p. 101)). But it is inaccurate to state that the Marx-Rubin formulation of the law of value assumed homogenous labour without justification. The law of value is a dynamic theory of labour allocation based on the tendency of heterogenous labour to be homogenised via commodity exchange, and in this sense is very different from modern static formulations of it. The reduction coefficients are continuously calculated by a distributed computation that is implemented through the actions of the economic actors. Homogeneity emerges in the simple commodity economy under the assumption that economic actors have equal productive powers as members of the same species, strive for equal renumeration for their labour time, that is they consider themselves equal, and are free to realise their equality through unconstrained economic activity. Rubin states that the '*equalization of exchanged commodities* reflects the basic social characteristic of the commodity economy: *the equality of commodity producers*'. The **SCE** models this ideal situation by allowing identical actors to freely move between sectors of production in order to meet identical consumption requirements. In reality, things are not so simple, and in the context of tendencies to narrow the wage dispersion, Rubin (1973, ch. 15) discusses factors that prevent homogenisation.

6.2.6 Labour values as attractors for prices

The law of value is a phenomenon that emerges from the dynamic interactions of private commodity producers. In the model presented (i) labour values are global attractors for market prices, (ii) market prices are error signals that function to allocate the available social labour between sectors of production, and (iii) the tendency of prices to approach labour values is the monetary expression of the tendency to efficiently allocate social labour. The constant of proportionality of the linear relationship between labour values and market prices is the monetary expression of labour-time (MELT), which measures how many units of money represent one unit of social labour-time. The MELT summarises a non-obvious causal relationship between non-market phenomena (production times) and market phenomena (prices), and links the total available social labour-time to its monetary representation. The concept of labour commanded, which measures how much social labour-time a commodity fetches in the marketplace, is important for theorising how deviations of price from value are labour re-allocation ‘signals’. The labour commanded by a commodity normally mismatches the private labour-time expended in its construction, indirectly signalling whether the labour was socially necessary or not.

The law of value operates ‘behind the backs’ of actors via money flows that place income constraints on their local evaluations of commodity prices. The equilibrium of the simple commodity economy is a statistical equilibrium, in which a single commodity type may realise many different prices. In consequence, the regulating role of exchange value is a property of price distributions, not individual transactions. Further, the law of value can only emerge in broad models of economic systems that complete the feedback loop between production, consumption, exchange *and* reallocation of labour resources.

An actor engaged in free exchange derives personal benefit from transactions and the immediate apprehension of this fact motivates subjective theories of value. But an exchange has causal consequences beyond the immediate moment and the satisfactions of mutual commerce that derive from its embodiment within a system of generalised commodity production. Actors do not normally think money into existence although they do decide to spend more or less of what they have. Their income is a local representation of a global resource constraint not under their subjective control. Although money *exchanges* according to demands for use-values, and is normally accompanied by the satisfaction of desires, it *refers* to amounts of social labour-time. Local flows are easier to apprehend than global reference, which partially accounts for the relative neglect of objective theories of value.

CHAPTER 7

VALUE IN A CAPITALIST ECONOMY

Cottrell, Cockshott

7.1 THE PROBLEM OF PROFIT RATES

In our analysis so far we have looked at how the labour theory of value applies to an economy in which indidividuals exchange commodities, which they have themselves made, with other independent producers. The classical political economists Smith, Ricardo and Marx believed that things differed ’after the accumulation of stock’, that is to say once production was carried out by larger enterprises using significant amounts of capital.

Suppose it is the year 1800 and we have two capitalists Mr Miller producing flour and Mr Arkwright producing yarn. Flour production was considerably more automated than cotton manufacture. A large cotton mill might employ 500 people, a flour mill fewer. Let us suppose that Mr Miller has only 200 employees to Mr Arkwrights 500. The spinning of cotton and the grinding of flour both required large amounts of mechanical power. So both sorts of miller needed very substantial capital investment to dam rivers, divert the water through lades and over water wheels, and then to construct and equip the mill buildings. We will assume that the capital invested in the two mills was the same: £200,000. At that time in Britain a year’s labour created a value of about £200 of which the worker might be paid £100. The annual value added by Mr Millers employees would be £40,000 and by Mr Arkwrights £100,000. The implication of this is that the rate of profit that they would earn on their capital would be different (see Table 7.1).

Faced with this discrepancy the classica economists either had to accept that capitalists in different industries would earn different rates of profit,

Table 7.1: Two Millers

	Mr Miller	Mr Arkwright
Capital in mill	£200,000	£200,000
Employees	200	500
Value Added	£40,000	£100,000
less Wages	£20,000	£50,000
Profit	£20,000	£50,000
Rate of Profit	10%	25%

or they had to conclude that the labour theory of value had to be modified to adapt it to the new world of capitalist industry. In fact, they found the idea that capitalists in different industries would earn widely divergent profit rates implausible. They thought that the rate of profit in different industries should be the same, for were it not, capitalists would shift their funds into the industry earning the higher profit rate. A consequence of this was that they assumed that actual prices would diverge somewhat from labour values to allow capitalists in different industries to earn the same rate of profit. Among the classical economists the most thorough treatment of these corrections was given in Marx (1971). In the literature of Maxian economics the technique of applying these corrections was called the ”transformation problem” because it transformed of a set labour values into a set of prices at which profit rates would equalise.

During the 20th century a considerable body of mathematical literature (see for instance Sraffa (1960), Samuelson (1973), Steedman (1981)) grew up analysing how prices in a capitalist economy would be determined on the assumption of an equal rate of profit accross all branches of industry. This branch of economics came to be called the neo-Ricardian school.

But if were allowed that profits on stock caused prices to diverge from labour values, what was the point of the labour theory of value?

Could one not say that both capital and labour contributed jointly to the value of the product?

As a result the labour theory of value is now taught to students as something of a historical footnote, interesting but now considered obsolete. But this judgement is to harsh to great thinkers like Adam Smith and Karl Marx.

They were onto a basically sound intuition in emphasising the role of labour in creating value.

7.2 QUESTIONING THE PROFIT RATE

What theorists have been inclined to forget, is that the equalized rate of profit is not a fact. It is an assumption, one that is absolutely crucial to all neo-Ricardian theories. **?** was able to show that many examples used by Steedman to demonstrate the frailty of the Labour Theory of Value, fell apart and became economically meaningless given the slightest deviation from equal profit rates.

What is a fact, is that the distribution of the rate of profit in capitalist economies is quite wide, and broadly stable over time. Yes, there are forces working in the direction of equalization, but there are complementary forces working in the direction of dis- equalization; and the joint outcome of these forces seems to be an ”equilibrium” degree of dispersion of profit rates, with different capitals occupying different places in the distribution at different times[1].

The greater the equilibrium dispersion of profit rates, the worse are neo-Ricardian prices as approximations to actual prices – or even to their ”centers of gravity,” discounting the effects of short-run supply- demand disequilibria. On the other hand, on the maintained hypothesis of an equalized rate of profit, the greater the dispersion of the capital to labour ratio, the worse are labour-values as approximations to actual prices. Since both of these distributions are non-degenerate, the question of whether neo-Ridardian prices or labour- values offer the better systematic approximation to actual prices is an empirical one. The evidence to date shows, with remarkable consistency across data-sets drawn from different capitalist economies and different time periods, that the two approximations are roughly equally good (see Chapter 7.3). The labour theory of value is at least as good a predictor of actual prices as the neo-Ricardian theory.

It might be objected by some Marxist economists that treating the Labour Theory of Value and the Sraffian system as alternative theories of relative prices is to miss the point. Is it valid bracket Ricardo and Marx as proponents of the Labour Theory of Value when Marx was concerned to criticise Ricardo, not merely to second him?

---

[1] See Farjoun and Machover (1983)

There is some force in this objection, but it is overdone. True, Marx's primary object was not to develop a theory of relative prices. He wanted to lay bare the basis of profit in the capitalist exploitation of labour, to discern the ”laws of motion” of capitalism, and to demonstrate that capitalism is a historically transient mode of production, whose internal contradictions necessarily propel it in the direction of its supercession by socialism. From this standpoint, the Labour Theory of Value was but a stepping stone towards his theory of *surplus value* something quite foreign to Ricardo. And, it may be said, whatever is valid or salvageable from among the latter ambitions may be reconstructed without appeal to the Labour Theory of Value. This last claim we will tackle shortly. For the moment we want to point out that although a theory of relative prices was not Marx's central concern, as such, it does nonetheless play a key role in his work, and is a valid scientific question in its own right. One might add that Ricardo, too, placed the Labour Theory of Value in the service of an analysis of the ”laws of motion” of capitalism as he saw them – e.g. the progress towards the famous ”stationary state” via a falling rate of profit.

Marx's analysis of exploitation assumes that the prices of commodities in terms of money are in proportion to their labour-values. There is weak and a strong reading of this assumption. On the weak reading, it is just an expositional tactic for representing at the level of the individual factory and the individual worker, social relations that obtain between the class of workers and the class of capitalists. It projects onto the *individual* working day a division into surplus and necessary labour time that is in reality a relationship between parts of the *total social working day*. This is divided between time spent in industries producing workers' consumer goods and time spend producing goods used by the capitalists. The weak position would say that these conditions of projection need not hold empirically for the thesis about the social totality to be valid.

The strong position would state that the conditions of projection are more or less empirically valid, in the sense that there is such a strong correlation between the prices of commodities and their values that what is true at the social level is also true at the micro level.

Hence, although the principal concern of Marx in his famous chapter on the commodity (Marx (1954)) may have been the analysis of the *social form* of value, this does not indicate that he was unconcerned with the empirical relationship between price and value. Generally he held that movements in price reflected movements in value. This indeed was the specific form of representation of the category value (abstract social labour) in capitalist society. The essence of this form of representation was that there was a close correspondance between the structure of prices and the structure of values. Marx of course allows for disturbing elements – temporary imbalances of supply and demand, differing compositions of capital between branches, etc. – but the existence of these distorting factors no more invalidates the underlying hypothesis than the reality of air resistance invalidated Galileo's theory of falling bodies. The claim is that the underlying tendency will produce clear measurable effects, which can be distinguished from the effects of the disturbing factors.

What does one need to know in order to calculate labour-values?

The input-output structure of the economy[2], including intersectoral technical coefficients and direct labour coefficients. With this knowledge, one can invert the ”Leontief matrix” (or perform an iterative approximation of same) and derive the full set of labour-values. With the same information, and by means of the same computations, one can determine the vector of gross outputs required to support any given vector of final demand – a basic planning problem[3].

What does one need to know to calculate neo-Ricardian prices?

Basically the same: the full set of input-output coefficients, plus a distributional variable – either the (uniform) wage or the (uniform) rate of profit.

Is it in any way necessary to calculate labour-values as a step on the way to calculating neo-Ricardian prices?

No. This is one of Steedman (1981)'s key points, and of course he is right. In this sense there is no ”transformation problem”. If one's objective is to derive the set of Sraffian prices or ”prices of production,” one does not have to go via labour-values. That would be an awkward detour. And the question ”What is the correct mathematical relationship between labour-values and prices of production?” would seem to be of interest only if one has some prior commitment to labour-values. Why should one have any such commitment?

Labour-values seem to be analytically redundant.

But this argument loses its force if, as shown later, it turns out that labour-values and prices of production are about equally good as predic-

---

[2] This is discussed in more detail in Sections 7.3.1, 12.1 .

[3] We explain this in more detail in Section 7.3.1

tors of actual prices in capitalist economies. labour-values are a ”detour” only if one’s theoretical terminus is neo-Ricardian prices – but why should that be one’s theoretical terminus if one’s ultimate object is to analyze real economies and their laws of motion?

Suppose that labour-values and neo-Ricardian prices are, about equally good as predictors of actual prices. Are there then any grounds for preferring one theory over the other?

There are several, but here is one perspective to start with.

If, in the neo-Ricardian system, one asks, ”What determines prices (or price movements over time)?” the answer is, more or less, ”Everything.” (The full set of technical relationships and the profit-rate or wage.) This answer is strikingly uninformative.

But ask the same question of the Labour Theory of Value and you get a clear, informative answer: the systematic component of both cross-sectional and time-series relative price variation is primarily governed by the labour time socially necessary to produce the various commodities. On the grounds of parsimony in explanation (Occam’s razor), the Labour Theory of Value looks pretty good.

Why has the relative price of computing power fallen so dramatically over the last decade?

The testable explanatory hypothesis of the Labour Theory of Value is that technology in the computer industry has progressed in such a way as greatly to reduce the labour time necessary to make a computer of given specification, while also, of course, upping the specifications dramatically. By contrast, the neo-Ricardian answer – it would presumably go like this: the input-output structure of the economy has changed in such a way as to reduce the price for computers consistent with the computer industry earning the average rate of profit – seems not to offer any real explanatory purchase at all.

7.2.1 Does the Labour Theory of Value have a mechanism?

David Hume famously argued that we can never discover any necessity in causal connections between matters of fact. Necessity resides solely in the realm of mathematics and logic; among matters of fact there can at best be ”constant conjunctions”, brute empirical associations. Hume’s argument is notoriously difficult to refute, yet surely most scientists feel that there must be something wrong with it.

We expect of a good theory that it does more than produce predictions that happen to come out right most of the time. We expect the theory to specify some underlying mechanism responsible for the production of the effect in question. So: Empirical success apart, what is the mechanism of the Labour Theory of Value supposed to be?

Of the classical proponents of the Labour Theory of Value – Smith, Ricardo and Marx – only Adam Smith (whose version of the Labour Theory of Value was of course considered confused by the latter two) actually specified a mechanism.

In Smith, the pressure towards the exchange of commodity bundles containing equal quantitites of labour time resided in the subjective reckoning of the parties to the exchange. The beaver-hunter, seeing that his ”output” took twice as much labour to produce as that of the deer-hunter, refuses to part with the beaver for less than two deer. But unfortunately this mechanism would seem to operate, at best, only in the ”early and rude state of society which precedes both the accumulation of stock and the appropriation of land”. Capitalists don’t calculate the labour-content of their products, or of the commodities they purchase. Plus, even if they wanted to, it’s much more difficult to calculate the labour-content of a commodity produced via a complex division of labour.

Neither Ricardo nor Marx specified an alternative mechanism. Ricardo was perfectly confident that the Labour Theory of Value was right, but if you look for a definite mechanism in the Principles you will be disappointed. What purports to be an argument for the Labour Theory of Value appears on p. 25 of the Sraffa edition, but it is actually no more than an account of what will happen under certain circumstances *on the maintained hypothesis of the Labour Theory of Value*.

Marx, though he doesn’t give a mechanism as such, does offer an argument, in chapter 1 of Capital, I. It goes roughly like this.

(1) Commodity exchange should be conceived as an equation. To make sense of the ”exchange of equivalents” we must suppose that there is *something* present in equal quantities on both sides of the exchange.

(2) Labour time is the only acceptable candidate for this ”something”; since the use-values of commodities are incommensurable.

This argument has not persuaded many people. At least on the face of it, it seems to be full of holes. For instance:

(1) Why do we have to conceive of exchange as an equation, other than, trivially, of equal monetary magnitudes?

There doesn't seem to be anything compelling about this picture.

(2) Even if we do think of exchange in that way, and if we accept Marx's point about the incommensurability of disparate use-values, is labour time really the only candidate for the thing that is equated?

What about, say, energy-content?

(3) Besides, when we get to volume III of Capital, Marx admits that embodied labour time is *not* actually equated in commodity exchange under capitalism, even in ”long-run equilibrium,” so to speak.

There would seem to be two possibilities here. Either the confidence of Ricardo and Marx concerning the Labour Theory of Value was just misplaced. Their failure to come up with a convincing mechanism is fatal. Alternatively the intuition of Ricardo and Marx was sound, but outran their capacity to articulate a proper justification of the Labour Theory of Value: the job can, however, be done. We think the second interpretation is the right one.

Capitalists don't calculate labour-contents. But they do calculate profit rates, and act on those calculations, so there is a theoretical warrant for the idea of a tendency towards the equalization of the rate of profit. Under certain conditions: the rate of profit is ”small” and/or the dispersion of the value composition of capital across industries is limited, the equalization of profit rates will produce a tolerable approximation to relative prices = relative labour-values.

So we can produce a mechanism for the Labour Theory of Value, as approximation, after all – only it is ”parasitic” on the mechanism for generating Sraffian prices, which justifies the idea that the Labour Theory of Value is theoretically redundant!

But this is wrong. Farjoun and Machover (1983) showed that if the Labour Theory of Value is cast in a probabilistic form, one can derive a stochastic version of price–value proportionality without appealing to a uniform rate of profit. We develop this approach further in Chapter 6.

Thus Farjoun and Machover supply a definite mechanism supporting a stochastic version of the Labour Theory of Value, one that is not parasitic on the uniform- profit condition. This mechanism is statistically emergent. It is certainly not the direct result of agents' paying attention to the labour-content of commodities in the mode of Smith's hunters; and it would seem to be invisible to methodological individualism. On the other hand, the probability law in question clearly must be realized via the interactions of a multitude of capitalists and workers. The situation is analogous to statistical mechanics. The ideal gas laws, for instance, are statistically emergent from the interaction of millions of individual molecules.

7.2.2 On the specialness of labour

Suppose you grant the above, at least for the sake of argument. You may still wonder: But after all, what is really special about labour?

Couldn't you do the same sort of statistical number using oil-content, timber-content or what-have-you?

Why is the Labour Theory of Value of any more intrinsic significance than the Oil Theory of Value or the Timber Theory of Value?

Everybody knows that human labour is a special process and labour-power a very special commodity. But a certain sort of hard-nosed theorist is very unwilling to grant labour any special theoretical privilege. This attitude, although ultimately theoretically debilitating, is understandable. One doesn't want to be caught sneaking into one's basic economic theory a privileging of labour that is based on scientifically ”extraneous” ideological, political or humanitarian concerns. If the Labour Theory of Value acquires its validity only from, say, a standpoint of political-ideological sympathy with the labour movement, this seems like sufficient reason for rejecting it[4].

The thrust of the earlier chapters of this book has been to show that the granting of a special privilege to human labour time, and the commodity labour power, is not in the least the effect of the intrusion of extraneous factors into the realm of theory.

The economy is ”about” the production of goods that serve certain human purposes, by human beings, via their labour time. The stipulation that the goods serve human purposes – though those purposes may be quite various – is necessary to distinguish economic production from, e.g., the production by humans of carbon dioxide and other bodily wastes. This is not an ”ideological” statement, extraneous to science, since some such statement is absolutely required to prevent the ”economy” from vanishing as a

---

[4] Sraffa's very title, ”The Production of Commodities by Means of Commodities,” seems to bespeak such a concern, which is more explicit in some of his followers. From the abstract theorist's point of view, labour (or rather labour-power) is just one of the n commodities that jointly produce each other.

specific object of theoretical investigation. Otherwise how can one make a principled distinction between the economy, and all the other stuff going on in the biosphere with which, of course, the economy is intricately linked?

One could try delimiting the economy as the set of activities that earn, and participate in the determination of, the equalized rate of profit. But strictly speaking, this would be the empty set. Or ”the set of activities involving the allocation of scarce resources by and on behalf of humans.” But that is too broad to isolate the economy as such; and besides, it skirts the point that labour is the key ”scarce resource.”

Here are three quotations, one from each of the classical proponents of the Labour Theory of Value.

A. Smith:

> The real price of everything, what everything really costs to the man who wants to acquire it, is the toil and trouble of acquiring it. ... Labour was the first price, the original purchase-money that was paid for all things. It was not by gold or silver, but by labour, that all the wealth of the world was originally purchased...

B. Ricardo:

> Possessing utility, commodities derive their exchangeable value from two sources: from their scarcity, and from the quantity of labour required to obtain them.
>
> There are some commodities, the value of which is determined by their scarcity alone. No labour can increase the quantity of such goods, and therefore their value cannot be lowered by an increased supply. Some rare statues and pictures, scare books and coins, wines of a peculiar quality, which can be made only from grapes grown on a particular soil, of which there is a very limited quantity, are all of this description. Their value is wholly independent of the quantity of labour originally necessary to produce them, and varies with the varying wealth and inclinations of those who are desirous to possess them.
>
> These commodities, however, form a very small part of the mass of commodities daily exchanged in the market. *By far the greatest part of those goods which are the object of desire, are procured by labour; and they may be multiplied, not in one country alone, but in many, almost without any assignable limit, if we are disposed to bestow the labour necessary to obtain them.* (emphasis added)

C. Marx:

> Every child knows that any nation that stopped working, not for a year, but let us say, just for a few weeks, would perish. And every child knows, too, that the amounts of products corresponding to the differing amounts of needs demand differing and quantitatively determined amounts of society's aggregate labour. It is self-evident that this necessity of the distribution of social labour in specific proportions is certainly not abolished by the specific form of social production; it can only change its form of manifestation.

Can one in principle construct an X-theory of value (an XTV), substituting some other item in place of the labour of the Labour Theory of Value?

And if so, does that mean that the Labour Theory of Value has no special claim to privilege?

There is a first technical requirement: X must be a ’basic’ commodity, i.e. one which enters either directly or indirectly into the production of all others. But surely we can find some of these besides labour – oil, perhaps.

Now notice a second technical point. When calculating labour- values, it is necessary to value all inputs to production in terms of the amount of labour-time it takes to produce them – except for labour- power! Direct labour inputs must be ’valued’ at one hour per hour of labour performed. What would happen if one instead valued the direct labour input itself in terms of the labour-time required to reproduce the workers’ labour-power, when that is less than one hour per hour of labour performed? Then all ”labour-values” would go to zero. It can be seen more intuitively if you imagine calculating labour-values by an iterative method: start out by approximating the labour-value of each commodity by its *direct* labour-content; then adjust your first approximation by bringing into account the first-round labour-values of all the other inputs; and so on. The point is that if the direct labour input itself is revalued, in the second round, at its ’labour-content’, conceived as the value of labour-power, the second approximations will be smaller than the first; and “labour- values” will shrink every time round this loop until they disappear.

The same applies to oil: in order to prevent the “oil-values” of all commodities from going to zero, one has to attribute to each barrel of oil that enters the production process directly, a value of one barrel – and not the (smaller) amount of oil that it takes to *produce* a barrel of oil. In effect, one has to make a distinction analogous to that between labour and labour-power, e.g. between the combustion of oil, and a barrel of oil itself as“combustion-power”.

But is this distinction really significant for anything other than human labour?

As the possessor of a unit of human labour-power, it matters to me to what extent my labour-power is exercised in actual labour per unit time. I want it to be exercised to some degree, preferably in interesting ways, but enough is enough. Suppose somebody urges me to work more than this, saying “After all, your labour-power will still be reproduced. If you use up more calories labouring, you’ll be provided with more to eat.” This misses the point. I am not just concerned that my labour-power be reproduced (though this is important): I am independently concerned about the amount of work I do, since there are other things I like to do with my time.

Here, of course, there is no analogy with oil. A barrel of oil simply doesn’t care if it’s used up or not: it has nothing else to do. And neither does anybody else care, *except insofar as it is a non- reproducible resource that is liable to run out* (or at any rate become much more costly in terms of labour-time to extract, over a relevant time-horizon). A society that is capable of reproducing its stocks of fuels (”combustion-power”) over all relevant time-horizons has no additional reason to be concerned about the rate at which actual combustion is taking place per period. But a society that is capable of reproducing its stock of labour-power over all relevant time horizons does have an additional reason to pay attention to the rate at which actual labour is performed by its members per unit time, since this is of concern to all its members individually.

It may be helpful here to make a distinction between ”strongly producible goods” and ”weakly producible goods”. A strongly producible good is one that requires as ”ultimate” inputs only labour and natural resources the planetary supply of which is, at least for practical purposes, unlimited. A weakly producible good is one that requires as an input some natural resource whose supply is limited, and may pose a definite constraint over some economically-relevant time- horizon.

There is no possible justification for taking any strongly producible good X as the basis for an ”XTV”. The scarcity of such a good is strictly ”derived” –derived, that is, from the scarcity of labour-power and possibly of other weakly producible inputs.

On the other hand, it would be possible, in principle, to base an XTV on some weakly producible X other than labour. But notice a formal constraint on an X-content theory of value (as opposed to a non X-content theory, such as the ne-Ricardian or General Equilibrium systems). Since exchange ratios, the explananda of such a theory, are scalars, X- content must itself be a scalar. In other words, X must be homogeneous – or at least it must be possible to treat X as homogeneous for theoretical purposes, without departing too radically from reality. This requirement clearly rules out "land", i.e. one can't even begin to think of the land-content of a commodity as a scalar quantity, though it would seem not to rule out oil. Neither, of course, does it rule out labour. Yes, human labour-time is not truly homogeneous. But nonetheless human labour-power is an all-purpose resource, in the sense that anyone of average intelligence and dexterity can be trained to perform almost any of the tasks required in the economy.

Conclusions so far: An XTV is in principle possible for any X that

(1) is 'basic' in the technical sense,
(2) is only weakly producible and
(3) may be conceived as homogeneous as a tolerable first approximation.

There is an aphorism that Marxian economics is the economics of capitalism, while neoclassical economics is the economics of socialism.

You can see the general idea: Marxian categories are fine for exposing the injustices of capitalism and diagnosing its tendencies towards crisis, but if you can assume that the means of production are in the hands of the associated producers and the distribution of income is right, and you want to get down to some serious resource allocation, then what you want are the neoclassical marginal conditions. There may be a grain of truth in this. But there is more than a grain of truth in a proposition that is close to an inversion of it: One can see the rationale of the Labour Theory of Value most clearly by adopting the standpoint of a socialist planner.

It is fairly standard practice in domains such as investment and growth theory to frame the problem initially in terms of a 'command' system, to work out the optimal solution from this perspective, and then to claim that – Hey, presto! – a perfectly competitive system will duplicate the command optimum[5]. One may have some skepticism about this but nonetheless, perhaps one can say this: To the extent that capitalist economies approximate,

---

[5] Some of Solow's stuff is in this vein; and for a more recent example, see **?**.

loosely, stochastically, in certain dimensions, and relative to their peculiar distribution of income, to economic rationality, theoretical results relating to a planned economy may provide insights into the workings of capitalism. The Labour Theory of Value is a case in point.

Labour-power is a scarce resource: it is also a universal, all- purpose resource. And its scarcity is, unlike that of strongly producible items, direct and not derived[6].

The reason for this is that while labour-power *is* clearly reproducible, its reproduction takes place under ”special” conditions relative to the rest of the economic system. People can decide to put more resources into producing labour-power: they can have more children. But variations in the birth rate are, for the most part, not driven by the sorts of forces that drive other production decisions. In a capitalist economy, procreation is not a profit-oriented production process (and the ”output” is not the property of the possessors of the relevant means of production!). Similarly, in a planned economy, procreation does not fall within the sphere of planning of production. Even though the state may wish to encourage or discourage the having of children, and may have some impact on the birthrate, it can hardly plan this sector in anything like the way it can plan industry. This makes the production of labour-power the ”exceptional” sector of the economy, as Farjoun and Machover put it.

Furthermore, it is not just that the planners *can’t* plan the production of people like the production of steel: Why would they *want* to augment the population (hence relaxing the ”labour-power constraint” on the plan)?

There may be special instances where rapid population growth is in the interests of a state, but surely the general object of a plan is to maximize per capita production, not total production. And this objective will not, in general, be served by expanding the labour force via expansion of the population.

Thus, the very general fact that the planned economy is being planned for the benefit of human beings – and not for the benefit of oil, iron, electricity, or what have you – is not merely an extraneous “social” or “ideological” consideration, but rather connects directly with the issue of rational economic calculus. The focus on labour-per- unit-output as the appropri-

---

[6] Ricardo:“By far the greatest part of those goods which are the object of desire, are procured by labour; and they may be multiplied... almost without any assignable limit, if we are disposed to bestow the labour necessary to obtain them.”

ate measure of cost for each good is simply the converse of one's focus on output-per-unit-labour (strictly, per person, but if we assume that labour and number of people are positively linearly correlated, this amounts to the same thing) as the general maximand. Conservation of non-reproducible natural resources, while it may well be important, is in general only a means of ensuring that the maximization of output-per-unit-labour – minimization of labour-per-unit-output – is sustainable for future generations. We consider 'productive' natural resources here: conservation of other species is arguably a different matter – a moral imperative.

Further, consider the issue of full employment. Clearly, this was a priority under socialism. The minimization of the labour time required to produce things was also a priority. As a first approximation, the idea might be: "Use all the labour-power there is, but spread the resulting labour as thinly as possible over the things you are producing, so as to be able to produce as many things as possible."

This requirement creates another special feature of labour. Not only is labour scarce, leading to the need to economize it in any particular branch of production, but it should be fully used each period. This feature does not carry over to other resources.

A non-reproducible natural resource such as oil may be scarce, in the sense that its ultimate supply is finite, yet there is no requirement that it be "fully used" each period. Indeed, what would that mean in the case of oil? All we can say here is that there is no point in extracting more oil each period than one wants to use during the period; unless one has a specific reason for adding to stocks. In that sense the current flow output of oil should be "fully used". But of course this current flow output is endogenous: one produces just as much oil as one plans to use, and the planned usage in turn is determined by technology, in the form, let us say, of the oil-input to labour-input ratio in production, in conjunction with the amount of labour one plans to perform.

Thus, while it would in principle be possible to construct an oil theory of value in place of the Labour Theory of Value, there are several reasons why the Labour Theory of Value is of special significance to human societies. One can imagine circumstances in which the calculation of embodied oil-values might be desirable in a planning context. For instance if scarcity of oil is the most pressing constraint on the economy, and there is no possibility of substituting some alternative, producible via the application of labour, but these do not in fact obtain. Labour-content is clearly the best single, scalar measure of the cost to society of producing each sort of good. To the extent that relative prices under capitalism reflect, albeit in a highly imperfect and distorted fashion, social cost of production, one would expect to see the Labour Theory of Value borne out empirically – as indeed one does.

7.3 EMPIRICAL EVIDENCE FOR LABOUR THEORY OF VALUE

The view of orthodox economics in the West has been that the labour theory of value is ’discredited’. The labour theory of value has been replaced by the dominant marginalist price theory in university economics courses. But this discrediting has entirely been based on a-priori theoretical arguments. It has not been discredited by the the discovery of empirical evidence that was inconsistent with the theory. In science competing theories are supposed to be evaluated on the basis of their ability to explain observed data. Economics does not proceed in this way. The practical political implications of different economic theories are so great that it is very difficult for scientific objectivity to take hold. Whilst people build political parties on the basis of different economic theories, they dont fight in the same way over alternative theories of galactic evolution.

It was not until the 1980s that a serious scientific effort was made to test whether or not the labour theory of value actually held in practice. The pioneering work was done by Anwar Shaikh Shaikh (1984) and his collaborators **? ?** at the New School in New York. Following this, there is now a considerable body of econometric evidence in favour of the proposition that relative prices and relative labour values are highly correlated, or in other words, in favour of the law of value.

7.3.1 Method of calculation

The key to testing the labour theory of value has been the use of input-output tables. An input-output table is a way of showing the structural interaction of different industries. These tables are periodically constructed by government statistical departments for the leading economies of the world. The idea behind them can be grasped by looking at the example in Table 7.2. This shows in a very aggregate fashion the structure of an economy with 4 main industries labeled A,B, C, D. The columns corresponding to the industries show how much of the output of each other industry is used up by

Table 7.2: Example input output table

industry	A	B	C	D	final consumption
A		100	100	10	100
B	100				100
C		20			280
D	10		20		10
Wages	100	45	85	14	
Profits	100	35	95	16	
Sales	310	200	300	40	

a given industry. Thus industry A uses 100 from B and 10 from D. The numbers would refer to quantities of money, for now we can think of them as being billions of dollars. At the bottom we have rows showing the total amount of wages and profits earned in each industry and the total final sales of the industry. The final sales row is the sum of the wages, profits, and indirect inputs above.

It is possible to use input output tables to work out how many hours of labour went into producing the total output of each industry.

We start up by simply adding up the number of units of labour that were directly employed in each industry.

If we divide the directly utilised labour by the dollar value of the industry's output, we get an initial figure for the amount of labour in each dollar of the output. For industry A we see that 0.32 units of labour go directly into each dollar of output. Since we already know the number of dollars worth of A's output used by every other industry, we can use this to work out the amount of indirect labour used in each industry when it spends a dollar on the output of industry A.

This gives a second estimate for the labour used in each industry, which in turn gives us a better estimate for the number of units of labour per dollar output of all industries. We can repeat this process many times and as we do so, our estimates will converge on the true value. This process is illustrated in Table **??**. If the labour theory of value is empirically correct, then if you buy a dollar's worth of any product you would get back roughly the same

Table 7.3: Average percentage deviations between market prices and labour values for the USA over selected years. Figures extracted from (Shaikh 1998).

Year	Deviation
1947	10.5%
1958	9.0%
1962	9.2%
1967	10.2%
1972	7.1%
Average	9.2%

quantity of labour. In other words, the figures for labour/$ for each industry would be very similar as shown in the final line of Table **??**.

7.3.2 Results

Our example is very small and uses completely fabricated data. What happens when you look at a real economy?

Well for a start the tables are much larger, typically with around a hundred industries listed. But the same method can be applied, it just requires more computational effort. The work of calculation would have been daunting prior to the ready availability of computers for economic research. This may be why nobody seriously investigated the matter until the 1980s. But when Shaikh and others tried, they obtained results very similar to our toy example.

The general procedure in these studies has been to use data from national input–output tables to calculate the total labour content of the output of each industrial sector, and then to see how closely the aggregate money value of sales from each industry match their total labour content. Various different ways have been devised to measure the correspondence between the prices and the values. Shaikh (1984) explains the details of the process, and also offers a theoretical argument in favour of a logarithmic specification of the price–value regressions. Table 7.3 shows some results from Shaikh and his collaborators.

Table 7.4: **Comparing the correlation of prices to labour values in different countries (**Figures **?**).

Country	year	#industries	price/labour correlation
Japan	1995	85	98.6%
Sweden	2000	48	96.0%
USA	1987	47	97.1%
Greece	1970	35	94.2%
UK	1984	101	95.5%
Germany	1995	33	96.5%
France	1995	37	97.6%

As you can see, the average error you get when predicting United States prices using the labour theory of value is only about 9%. This has proven to be the case accross many industries and several decades.

An alternative way of measuring the similarity of prices to labour values is to draw a scatter plot relating the two and then try to fit a straight line to the data. If the labour theory of value is true, then the observations will tend to fall close to this line, and the line will pass through the origin. How close the observations are to the line is measured by what is termed the R^2 value of the data. If the $R^2 = 1$ then all points fall on the line and the line perfectly predicts the results. If the $R^2 = 0$ then the line is of no use at all in predicting the observations.

Studies—utilizing data from the United States, Sweden, Greece, Italy, Yugoslavia, Mexico and the UK—have produced remarkably consistent results, with strong correlations observed: R^2s of well over .90. It also seems to be the case from the literature that the larger the population of the country, the closer is the fit between observed prices and labour values, (Table 7.4). This may be an example of the way that statistical regularities become more apparent the larger the population on which the observations are performed.

Our presentation of how to calculate labour values from input output tables in section 7.3.1 said that you use the wages row of the input output table to estimate labour inputs to an industry. It could be argued that because this row is denominated in money, rather than in hours of labour, it is not really measuring labour inputs. It is possible to compensate for this by using data on hourly wage rates in the different industries. If we know the average hourly wage in an industry, we can translate that industries wage bill into actual hours worked.

The effects of doing this for the United Kingdom are shown in Table7.5.[7] In the published input–output tables, the labour input is expressed in £. Column (1) uses labour-value figures calculated on the assumption of a dummy wage-rate of £1 per hour for all industries. This is equivalent to assuming that any wage differentials across industries reflect differential rates of value-creation per clock hour. Column (2) is the same as (1) except for the exclusion of the oil industry, which is an outlier in the price–value regressions, presumably due to the high rent component (in the Ricardian sense) in oil extraction. Column (3) (which again excludes the oil industry) uses labour-value figures calculated using wages data from the *New Earnings Survey* to convert backwards from wages to hours for each industry—a correction relative to column (1) if (and only if) inter-industry wage differentials are the product of extraneous factors, and do not reflect differential rates of value-creation.

As can be seen from the equation (2) estimates, 'simple' labour values produce an R^2 of nearly 98% when the oil sector is excluded and the dummy uniform wage is adopted. The effect of adjusting for differentials in wage rates and using raw labour ours in calculating the values gives a lower correlation of just over 96%. This is consistent with the hypotheses that :

(1) Labour of higher skills produces more value per hour.

(2) Inter-industry wage differentials at least partly reflect such skill differentials.

This suggests that the use of money wage bills as a surrogates for labour inputs to industries is valid.

Alternative value bases: empirical evidence

However, the question arises as to whether one could produce equally good results using something other than labour time as the 'basis' of value. The empirical answer to this question seems to be negative, as shown in Table 7.6. For the purposes of these regressions we used the Leontief inverse

---

[7] For further details regarding these estimates, see Cockshott, Cottrell and Michaelson 1995.

Table 7.5: Price regressions for the UK in 1984

	(1)	(2)	(3)
constant	−0.055 (−2.04)	−0.034 (−1.79)	−0.046 (−2.00)
labour value	1.024 (46.55)	1.014 (63.38)	1.024 (51.20)
N	101	100	100
R^2	.955	.976	.964

Figures in parentheses are *t*-ratios. All variables in logarithmic form. *Data source*: Central Statistical Office (1988).

of the UK input–output tables (Central Statistical Office, 1988, Table 5) to calculate the total (direct plus indirect) electricity content, oil content and iron and steel content of the output of each industrial sector. Using the same methodology as in Table 7.5 (based on Shaikh, 1984), we then regressed aggregate price on these various 'values', both singly and in combination with labour values, in logarithmic form. The sample size is 100 for each of these regressions, the electricity industry being excluded from the equations including electricity-content, and similarly for oil and iron and steel.

From columns (6), (8) and (10) it can readily be seen than none of the alternatives, taken alone, performs anything like as well as labour. The highest R^2, at .682, is obtained for electricity content, as against .955 for labour in column (1) of Table 7.5. Columns (5), (7) and (9) show how the alternatives perform when entered alongside labour values, enabling us to address the question of whether the alternatives contain any independent information, or in other words offer any marginal predictive power over prices when labour content is given. Only oil content passes this test. From the *t*-ratios (in parentheses below the coefficient estimates) it can be seen that while labour content retains its statistical significance in all cases, electricity content and iron and steel content become statistically insignificant in the presence of labour content. The fact that oil content contains some independent

Table 7.6: Regressions of price on labour values and some alternative ‘value-bases’ for the UK.

	(5)	(6)	(7)	(8)	(9)	(10)
constant	−.056 (−2.06)	−0.169 (−2.425)	0.066 (3.15)	0.307 (3.16)	−0.067 −2.38	−0.263 (−2.47)
labour	1.030 (23.76)		0.904 (46.07)		1.048 (36.53)	
electricity	−0.009 (−0.19)	0.903 (14.60)				
oil			0.109 (7.43)	0.615 (13.29)		
iron and steel					−0.027 (−1.31)	0.445 (7.09)
Adjusted R^2	.953	.682	.984	.639	.954	.332

Figures in parentheses are t-ratios. All variables in logarithmic form. *Data source*: Central Statistical Office (1988).

Table 7.7: Regression of alternative value bases for Greece

Value Basis	R^2
Agriculture	0.174
Electricity	0.668
Oil	0.674
Chemicals	0.702
Labour	0.942

Data from Tsoufildis and Maniatis 2002

information regarding prices is presumably linked to the element of rent in the price of oil. The North Sea fields are not marginal, which means that the labour time taken to extract North Sea oil is less than the socially necessary amount (on a world scale). The price of oil being determined on the world market, UK oil will then sell at a price above that which corresponds to its particular labour content. Table 7.7 shows similar results are obtained when analysing the Greek economy.

Table 7.8 offers another perspective on this issue. It reports the coefficients of variation (standard deviation divided by the mean), across the 101 sectors in the UK input–output tables, for x-content per £'s worth of output, where x equals labour, electricity, oil, and iron and steel respectively. This is the basic information supplied by the input–output tables: in Tables 7.6 and 7.5 it is worked up into regression format,[8] but it is worth considering 'raw'. Clearly, to the extent that x is conserved in exchange, one will find a relatively small coefficient of variation for x-content per £ of sales. From the second column of Table 7.8 we see that the coefficient of variation is almost four times as large for electricity as for labour, with those for oil and iron and steel being greater still.

7.3.3 Are the results real?

One objection that has been made (Kliman 2002) to the observed correlations between market prices and labour values, is that they arise as a statistical artifact. What we are comparing is the aggregate selling price of, for example, all the iron and steel produced in the USA with the total labour that went into it, and similarly for all the other industries. What we see is that the value of an industry's sales proportionate to the direct and indirect labour it uses. It has been argued that this is simply because a large industry has both large sales and a large workforce, and small industries have small sales and small workforces. Thus the correlation we see is spurious, arising as a side effect of industry size.

The comparisons of labour values with oil and electricity values etc, tell us that the correlations between values and prices are something real rather not spurious. If industry size generated spurious correlations for labour it would do the same for other inputs, and oil or electricity values would be

---

[8] That is, x-content per £'s worth of output is multiplied by the total monetary value of output to yield total x-content, on which the total monetary value of output is then regressed, in log form.

Table 7.8: Coefficients of variation

for *x*-content per £ of output

x	Coefficient of variation	C.V. relative to labour
labour	0.189	1.00
electricity	0.698	3.69
oil	2.156	11.41
iron and steel	1.477	7.81

Source: Calculated from Central Statistical Office (1988, Table 5). Labour figures calculated recursively by authors.

strongly correlated with selling prices - which they are not. There is something special about labour.

The danger of spurious correlation is, in some contexts, real enough. Take for example a study of the association between alcohol consumption and violent crime. Suppose an investigator runs a regression with number of violent crimes as the dependent variable and amount of alcohol consumed as the independent variable, for a sample of cities of widely varying sizes. We would expect to find a significant positive coefficient on alcohol consumption, but this would be of no scientific interest: simply, larger cities would be expected to show both more crimes and more alcohol consumed. The obvious correction here is to scale both variables of interest by expressing them per capita, dividing by city population. If there is a still a significant positive association then this might be of sociological interest.

Correlations in which the units of observation are of different ‘sizes’ are not necessarily spurious, however. Consider a variant on the city size example. Suppose a researcher has the hypothesis that population is the principal factor governing the size of cities as measured by their land area, or in other words that variations in population density are second-order. One way of assessing this claim would be to regress city land area on population and see if the relationship between these variables is close to proportional. In this case one is well aware that both land area and population are measures of city ‘size’, and the object of the exercise is to see how closely they are related. Now suppose someone were to object to this hypothetical study as follows: ‘This is a case of spurious correlation. Of course, bigger cities will in general both occupy more land area and have larger populations. To overcome this problem you will have to deflate land area and population by a suitable measure of city size, say the number of residential units.’ The objection is misplaced. In the first case above, city size (population) was an independent ‘third factor’ that might plausibly induce an apparent correlation between crimes and alcohol consumption, while in the second case there is no such independent third factor in play.

The correlation of prices and values across industries is of the second sort: it forms part of an investigation into the closeness of two variables that are in themselves reasonable measures of the size of industries, namely the aggregate market price of their output and the labour time embodied in that output. There is no independent third factor that could plausibly induce a spurious correlation here. The notion of the ’size’ of an industry is rather vague, but in everyday terms it means how many people are employed in the industry. A large industry is one that employs lots of people. But the classical labour theory of value predicts that if an industry is large in this sense, then the value of its output will also be big. Which is just what we see.

Digression 7.1 Marx's Theory of Value and the Labour Theory of Value

Marx made a distinction between what he called concrete and abstract labour, identifying only the latter as a source of value. Whilst Marx was explicit in this, the distinction is strongly implicit in any Labour Theory of Value. Ricardo proposes that "commodities derive their exchangeable value from . . . the quantity of labour required to obtain them." To render this meaningful, we must be able, in principle if not in practice, to quantify the labour required to obtain any given commodity. But one can't add up hours of baking labour, spinning labour, mining labour, etc. (i.e. specific concrete labours), unless one conceives of these as just various instances of human labour in general (i.e. abstract labour). Marx was clearer and more explicit on this, to be sure, but we don't see the concrete labour/abstract labour distinction as something that Ricardo would have objected to; rather, he seems to have taken it for granted.

It is a serious mistake, however, to go on to say that abstract, socially-necessary labour-time is something that is manifest or measurable only in the market prices of commodities. This is to render the Labour Theory of Value empirically vacuous. If the Labour Theory of Value is to have any empirical content, one must suppose that although one cannot *identify* the actual clocked labour-content of any given commodity with its abstract, socially-necessary labour-content, nonetheless market competition ensures that these two magnitudes do not diverge to an arbitrary extent. And if one is dealing with large collections of specific commodities, it is reasonable to take clocked labour-content as a measure of Marx's "substance of value."

More generally, what distinguishes Marx's Theory of Value from the Labour Theory of Value?

Marx's theory is a Labour Theory of Value set in a particular theoretical and political context; it is a Labour Theory of Value developed into a theory of exploitation and a critique of capitalism, something foreign to both Smith and Ricardo. To achieve this development, Marx had to distinguish very clearly between labour the activity and labour-power the commodity: that is the key conceptual difference with respect to Ricardo. Marx's theory of value is also in a sense the Labour Theory of Value generalized. That is, the exchange of commodities at prices roughly proportional to socially-necessary labour content is conceived by Marx as the specific manifestation, under capitalism, of the "necessity of the distribution of social labour in specific proportions" in order to satisfy the conditions of reproduction of any economic formation.

Digression 7.2 Complexity of computing labour values

The computational complexity of iteratively determining labour values as described in Table **??** is relatively low, significantly lower than the process of computing a strict matrix inverse which is the normal way the problem is specified in the literature. Naive matrix inversion has complexity N^3 but optimal versions exist with complexity $N^{2.38}$ (see Numerical Recipies Software (1988) page 104).

The iterative approximation method has complexity kN^2 where k is the number of iterations required to get an acceptably accurate answer. The answer converges rapidly so acceptable results are obtained with $k < 10$.

If fact disaggregated input output matrices are typically sparse with most elements being zero which allows further significant speedups by compacting the data to elide the zero elements. The resulting complexity is of order kNM where M is the mean number of direct inputs that go to make an output. For fully disaggregated tables M grows much slower than N, so the overall complexity is significantly less than N^2.

CHAPTER 8

FARJOUN AND MACHOVER'S THEORY OF PRICE

Cockshott

8.1 FARJOUN AND MACHOVER'S STATISTICAL MECHANICS APPROACH

When Shaik and other empirical investigators started to produce their results[1] after 1984, these results were a surprise to economists who noticed them. However, the year before, a remarkable book 'The Laws of Chaos' (Farjoun and Machover 1983) had appeared. The book, by two physicists, argued that economists were mistaken in trying to construct purely deterministic theories. The authors pointed out that since Boltzman physicists had been able to make useful predictions about the aggregate behaviour of systems which, at a small scale appear random and chaotic.

At a small scale the movements of molecules in a gas or a liquid are random, and this random movement is even visible, as Einstein pointed out in 1905, in the form of Brownian motion - the jiggling about of small particles like pollen grains in water observed under the microscope. But at a large scale these random motions even out, allowing useful generalisations: the gas laws, the laws of thermodynamics. Farjoun and Machover avered that economists were stuck with an early 19th century model of causality. If this was dropped then quite different modes of reasoning about the economy would become possible. Dispensing almost completely with orthodox economic theory the authors derived a series of interesting generalisations about capitalist economies. One of these was a prediction that market prices would be closely correlated with labour values.

---

[1] See section 7.3.2.

In science, predictions always seem more convincing than postdictions. The fact that Farjoun and Machover's theoretical results were rapidly confirmed by empirical research, lends their results weight, the more so when one considers that their predictions ran counter to recieved opinion in economics. We can not hope to give a full account of their theory here. Instead we will offer a simplified account, missing most of the mathematical rigour, but which should still give an intuitive understanding of the mechanism they proposed for the operation of the law of value.

Consider all of the commodities sold by firms in one country over the course of a week. These will constitute a vast array of different goods and services, some expensive and some cheap. Some will require a lot of labour to make, some a little. Suppose that the law of value holds and prices of commodities are closely proportional to their labour content. How should we measure this?

Farjoun and Machover introduce a random variable Ψ which stands for the average price of an hour's worth of embodied labour. The idea is that we express all of the national production of different goods: A380 airbuses, chocolate digestive biscuits, disposable nappies etc in terms of their labour content. We then divide this up into units of one hour each, and imagine that we randomly select an hours worth from this huge aggregate. We then look at how much that hours worth sells for in money terms.

They predicted that if one were to graph the frequency of occurence of different values of Ψ that one observed over a sufficiently large sample of commodities then the distribution would look like Figure 8.1. They predict that it should take the form of a bell curve or what statisticians call a normal distribution.

A normal distribution, as its name implies, is one of the most common sort of distribution that one comes accross. Lots of observations take this form. For example if one plotted the heights of 10 year old boys in the city of London, one would get a normal distribution. If you plot the actual weights of a sufficiently large sample of point coins you will get a normal distribution. If you plot the number of photons arriving per second in a telescope from some distant star you will get a normal distribution. In fact, wherever the feature you are measuring is the result of summing up a large number of random, independently operating causal processes, the distribution you get when you plot it will be the familiar bell curve of the normal distribution.

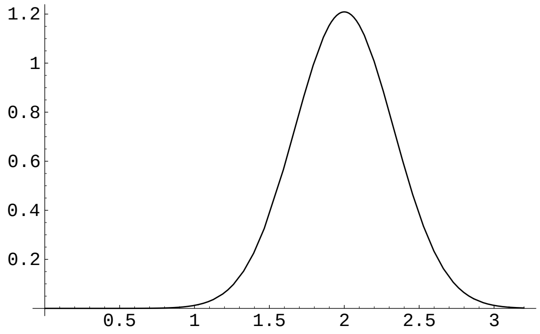


Figure 8.1:

Farjoun and Machovers predicted form of Ψ the relation between labour value and price

A normal distribution $\mathcal{N}(\mu,\sigma)$, is characterised by two numbers its mean (μ) or average, which goes through the peak of the distribution, and its standard deviation (σ) which describes how wide the bell curve is. Farjoun and Machover predicted that the plot of prices to labour values would have a mean of 2 and a standard deviation of less than $\frac{1}{3}$. How did they arrive at this conclusion?

First why did they say that one would expect the mean to be 2, in other-words that one would expect the average price of a commodity to be twice its labour value?

Well this is partly a matter of the unit of measurement they chose. As soon as you try to construct a theory of prices you are confronted with the question of the unit of account. When we want to measure distance we can do it in meters, which in their turn are defined in terms of a constant of nature - the wavelength of a particular type of light. This gives us a standard that is unvarying across space and time. But when it comes to measuring price, what do we use?

If we use money, should it be dollars, euros or yen?

If we stick to a single national currency, how do we account for inflation?

To get round this Farjoun and Machover use a technique favoured by Adam Smith and Maynard Keynes. They use the average hourly wage as their unit of account. Smith, as we have seen, held that the real price of any commodity was the amount of labour that it would command. Farjoun and Machover are slightly more precise and say that the real price of a commodity that took one hour's labour to produce is the number of hours of labour at the average hourly wage that it would command. Suppose a 4 kilo cod took an hour of direct and indirect labour to bring to the market. Suppose further that this cod sold for £15 and the average hourly wage was £6. In Farjoun and Machovers terminology then

$$\Psi_{cod} = \frac{\frac{\text{£}15/4kilo}{}\text{£}6/1hour}{1hour/4kilo} = \frac{\frac{15/4kilo}{}6/1hour}{1hour/4kilo} = \frac{\frac{15hour/4kilo}{}6}{1hour/4kilo} = \frac{15hour/6}{1hour} = \frac{15}{6} = 2\frac{1}{2}$$

We would expect Ψ to be > 1 since the selling price of any commodity can, as Smith showed, be decomposed into a part that pays wages and a part that pays profit[2]. The sale price goes to pay wages, profit and raw material costs. But the raw material costs likewise decomopose into wages, profits and a residuum of raw material costs. As you push the process back more and more stages one finds that the residual fraction of raw material costs tends to zero, so one can, to a good approximation, say the entire selling price goes ultimately to pay wages and profits[3]. Since Farjoun and Machover believed that in most capitalist countries value added was split 50/50 between wages and profits, it follows that the average price of the product of an hours labour will be twice the average wage for an hour's labour.

That explains why they expect the mean of $\Psi = \frac{\textbf{Price}}{\textbf{labour content}}$ to be 2. Why then do they settle on a standard deviation of $\frac{1}{3}$?

The argument here is very simple. They say that it is very rare for commodities to be sold so cheaply that the selling price would be insufficient to

---

[2] Smith also allowed for a part to pay rent, but Farjoun and Machover ignore this as being less significant than in the 18th century.

[3] Marx objected to this saying that the residual element of raw materials costs never quite reached zero. As a mathematical objection this is not very serious since the residual raw material cost exponentially approaches zero as a limit. As a sociological objection it has some weight since capitalist production presupposes the existence of capitalists who own raw materials and means of production and hire labour. If the raw materials and means of production were not in the hands of capital, then the workers would simply produce on their own account and there would be no division into wages and profits. Accepting this sociological point, Smith's mathematical approximation was reasonable.

pay the direct and indirect wages needed to make it. The cutoff point here is a value of $\Psi = 1$. Below this, the production of the commodity would be unviable, as not even wage costs would be met. For the sake of argument they assume that there is only one chance in a thousand of a commodity selling this cheaply relative to its cost of production.

By consulting a table of the normal distribution, one finds that the likelyhood of events 3 standard deviations away from the mean is about 1/1000, hence they derive that $\sigma = \frac{\mu - 1}{3}$, so for a μ of 2, then σ must equal $\frac{1}{3}$.

How do these predictions stack up against real data. Using data for the United Kingdom in 1984, the year after their book was published, we calculate[4] that Ψ can be pretty well approximated by a distribution with $\mu = 1.46$ and $\sigma = 0.151$.

At first sight this appears significantly different from the precdition they gave. But the difference is almost entirely due to the fact that in the UK in 1984, value added was split between profits and wages in the ratio one to two instead of the equal split assumed by Farjoun and Machover. The full form of their predition was that if e is the ratio of aggregate profit to aggregate wages, then $\Psi \approx \mathcal{N}(\mu, \sigma)$ with $\mu = 1 + e$ and $\sigma \leq \frac{e}{3}$. If we substitute the relevant value of e for the UK in 1984 into the equations, we find an almost exact fit.

An interesting consequence of their theory is that it predicts that the correspondence between prices and labour values will be closer when the share of profit in national income declines. If the share of profit in the national income declines, then relative market prices can be expected to approximate more closely to relative labour values. Profits allow room for prices to have a lower signal to noise ratio.

The distribution of Ψ is random, or entropic. One can calculate the entropy of a normally distributed random variable using an amended form of Shannons formula. Shannon gave the entropy of a signal as

$$\sum_i -p_i \log_2(p_i)$$

where i takes on a set of discrete values corresponding to recognisably different quantisations of the signal. A Normal distribution $\mathcal{N}(\mu, \sigma)$ is a Prob-

---

[4] Result derived from Cockshott and Cottrell 1998, with slight adjustment to bring the definition of Ψ used in that paper in line with the definition used by Farjoun and Machover.

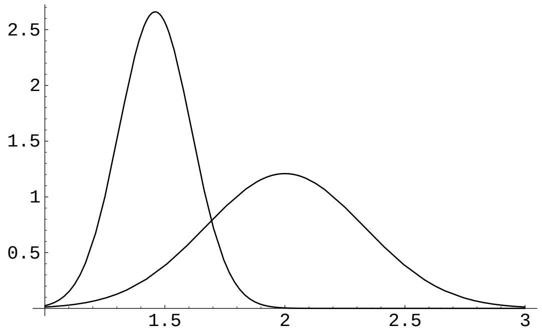


Figure 8.2:

Farjoun and Machovers predicted Ψ (right) compared with a measured Ψ for the UK in 1984, (left).

ability Density Function (PDF). It is a function over the reals such that

$$P(a,b) = \int_a^b \mathcal{N}(\mu,\sigma)(x)dx$$

specifies the probability that x will be in the interval between $a..b$. If we substitute this into the Shannon formula and numerically integrate, we can compute the entropy of a normal distribution with a given standard deviation.

We find is that normal distributions with a small standard deviation have a low entropy and ones with a large standard deviation have a large entropy. Figure 8.2 shows the distribution of Ψ predicted by Farjoun and Machover, compared with a normal distribution with the mean and standard deviation observed for the UK in 1984. The entropy of wider bell curve on the right is about 7.1 bits, whereas that on the left is about 5.9 bits.

From the standpoint of the thermodynamic approach to the economy, Ψ's entropy $H(\Psi)$, measures the disorder of price with respect to value. From the standpoint of information theory, $H(\Psi)$ measures how much information there is in the deviation of prices from values. Our computed values for $H(\Psi)$ tell us that the market price of a commodity gives around 6 bits of information distinct from the information provided by its value. This raises the question : what about the rest? How much of the information in prices comes from labour values?

8.2 INFORMATION CONTENT OF PRICES

The random variable $\Psi = \frac{\pi}{\lambda}$ gives the ratio of a price π to its labour value λ. It thus assumes that we know the value of a commodity as well as its price. Strictly speaking $H(\Psi)$ is a *conditional* entropy.

A conditional entropy written as $H(A\,|B)$ or the entropy of A given B, is defined on two random variables: A, B, and is the disorder of A with respect to B. In our case we have $H(\pi\,|\lambda)$, or the entropy of price given value. The information shared by both A and B, which is called their *mutual* information, is given by $H(A) - H(A\,|B)$. We want to know the mutual information of prices and values $H(\pi) - H(\Psi)$. This will tell us how much information is common to both price and value.

To work it out we need some estimate of $H(\pi)$ the information content of prices. To do this accurately we would need to apply Shannons entropy formula to all prices so that $H(\pi) = \sum -p(\pi)\log_2(p(\pi))$.

This would involve knowing the probability distribution of prices. We would have to know how frequent prices of £1.00 were, how frequent prices of £2.00 were, etc, which must be done for all possible prices going from the lowest price at which a commodity can be bought - say 1 penny, up to the largest observed price, perhaps something like £1,000,000,000 for a large warship[5]. Although in principle this could be worked out, we don't have access to the data on real commodity prices required to get an answer, so we will use an alternative approach, based on coding theory, which will give us a rough estimate of the information content of prices.

Although prices can range from pennies to billions, a price in the billions will not be quoted down to the last penny. A shipyard are selling an aircraft carrier to the Navy need only quote to the nearest £million. If you are buying a cooker in the price range £100 to £500, you only look at the pounds and ignore the pennies. In general prices need not be quoted to more than 3 significant figures, the rest is just noise or a convention like the last 99 on a

---

[5] To be compatible with the definition of Ψ we would have to weight the probabilities of each price with the amount of labour embodied in that price, but we need not be overly concerned with this technicality

£34.99 pair of shoes. What we also need to know is the order of magnitude of the price: are the units, pennies, pounds, tens of pounds etc. This implies that for most purposes prices can be written in so called scientific notation as something like 1.47E3 to represent £1,470.00.

In a number with the format `x.xxEy` there are 4 digits that carry all the information. But 4 decimal digits can be encoded in just over 13 bits of information, so we can give a rough bound on the information content of a price as $H(\pi) < 14$. This implies that the mutual information shared between the price and value of a randomly selected commodity is likely to be $< 14 - H(\Psi)$ or roughly 6 to 7 bits. We reasonably assume that the shared bits of information will tend to be the leading bits of the price.

We have given an outline of Farjoun and Machovers arguments let us now look at them more rigorously.

CHAPTER 9

A PROBABILISTIC MODEL OF THE SOCIAL RELATIONS OF CAPITALISM

Wright

9.1 INTRODUCTION

The dominant social relation of production within capitalism is that between capitalists and workers. A small class of capitalists employ a large class of workers organized within firms of various sizes that produce goods and services for sale in the marketplace. Under normal circumstances capitalist owners of firms collect revenue and workers receive a share of the revenue in the form of wages.

Over the last hundred years or more the number and type of material objects and services processed by capitalist economies have significantly changed, but the social relations of production have not. Marx (1954) proposed the distinction between the forces of production and the social relations of production to convey this idea. The existence of a social relationship between a class of capitalists and a class of workers mediated by wages and profits is an invariant feature of capitalism, whereas the types of objects and activities subsumed under this social relationship is not.

The social relations of production constitute an abstract, but nevertheless real, enduring social architecture that constrains and enables the space of possible economic interactions. These social constraints are distinct from any natural or technical constraints, such as those due to scarcities or current production techniques. Many economic models describe relations of utility between economic actors and scarce commodity types (i.e., actor to object relations studied under the rubric of neo-classical economics), or theorise relations of technical dependence between material inputs and outputs. But here we want to do something different and entirely abstract from these relations. Instead we'll examine relations of social dependence mediated by economic value. The basic parts of the economic model developed in this chapter are therefore quite simple, consisting solely of economic actors and money. The aim is to concentrate as far as possible on the economic consequences of the social relations of production alone, that is on the enduring social architecture, rather than particular and perhaps transitory economic mechanisms, such as particular markets, commodity types and industries. As the worker-capitalist social relation is dominant in developed capitalism the model abstracts from land, rent, states and banking.

In what follows we describe a dynamic, computational model of the social architecture of capitalism. It uses a small set of assumptions about capitalist property relations, but, when we simulate it on a computer, we find that it replicates some of the most important empirical features of modern capitalism. The computer serves as a logical testbed, and simulation allows us to explore the complex consequences of our simple assumptions. It allows us to say if important large scale features of a capitalist economy follow necessarily from its most basic social relationships.

The features of capitalism that we want to recreate are:

(1) The structural division of society into a small employing class and a large employed class (see section 9.3.1).

(2) The class distribution of income both between the employing and employed class (see section 9.3.2), and also the distribution of individual incomes.

(3) The distribution of sizes of capitals/firms with a small number of large firms and a large number of small firms (see section 9.3.4).

(4) The way in which the growth rates of firms cluster around the mean growth rate (see section 9.3.5).

(5) The rate at which firms die or go bankrupt (see section 9.3.6).

(6) The distribution of GDP growth rates and recessions (see section 9.3.7 and 9.3.8).

For each of these criteria we will examine the predicted statistical structures derived from the computational model and compare these to what is known about the statistical properties of the corresponding real-world data. Our aim is to see if a formal model of the social relations of capitalism can predict what we know about the statistical properties of capitalist economies.

This way we'll begin to understand what features of capitalism are necessary consequences of the way economic activity is socially organised. For instance, we will see that extreme income inequality is a necessary feature of capitalist social relations. This does not mean that we should accept this as a natural feature of economic life. There can be many kinds of political response to this scientific fact: accept the necessity of extreme income inequality (pro-capitalist), try to alleviate it within the current social relations (reformist), or accept the necessity of changing the social relations that give rise to it (anti-capitalist). But whatever the favoured political response the economic model we develop in this chapter indicates there are powerful and enduring market forces that continually generate income inequality, whatever the subjective intentions of politicians.

9.2 A DYNAMIC MODEL OF THE SOCIAL RELATIONS OF PRODUCTION

The elements of the model are a set of N economic actors each of whom has a sum of cash at their disposal. This sum may fall to zero, but in the model we assume that nobody actually gets into debt. We do not concern ourselves with the process by which the state issues money therefore the total money in the economy is a constant. We assume all transactions are in cash, there are no cheques, credit cards, etc in use. Each actor is either an employee, an employer or is unemployed. So the model consists merely of a set of people, each of whom has a sum of money. The simulation keeps track of who their employer is, if any.

All the actors in the economy are naturally partitioned into three mutually exclusive classes: an employing or capitalist class, if they employ one or more other actors, an employee or working class, if they have an employer, and an unemployed class, if they are neither an employee or employer. We assume an actor cannot belong to more than one class, but an actor may change classes over time.

The structure of a firm is simply an employer and their employees. Firm ownership is limited to a single capitalist employer: there are no stocks or joint ownership.

Although the total number of actors is fixed this can be interpreted as a stable workforce in which individuals enter and exit the workforce at the same rate. An actor, therefore, represents an abstract role in the economy, rather than a specific individual. At each instant of simulated time t the model has a state S_t. The evolution through time of this state, $S_t \rightarrow S_{t+1}$, is determined by a set of predominately random transition rules, which are applied at each time step. Processes that involve subjective indeterminacy (e.g., deciding to act in a given period) or elements of chance (e.g., finding a buyer in the marketplace) are modelled by selection from a bounded set according to a given probability distribution. Often the chosen distribution is uniform in accordance with Bernoulli's Principle of Insufficient Reason, which states that in the absence of knowledge to the contrary assume all outcomes are equally likely.[1]

The model considers a pure capitalist economy in isolation from non-capitalist sectors. The assumption of a finite set of actors differs from Marx's assumption of the existence of a latent reserve army of potential workers in the non-capitalist sector (e.g., domestic and subsistence agricultural workers) that may enter the capitalist sector and regulate the wage at a conventional level (Foley and Michl 1999).

Next we shall describe the rules that control how the actors interact with each other.

9.2.1 *The active actor*

Each actor in the economy performs actions on average at the same rate, which is modelled by allowing each an equal chance to act in a given time period. Note however that an actor may act multiple times in a given period, or not at all. The following rule selects an *active actor* who subsequently has the opportunity to perform economic actions. The unit of time is interpreted as a single month of real time, and therefore each actor is active on average once each month.

Actor selection rule : (Stochastic).

(1) Randomly select an actor *a* according to a uniform probability distribution.

9.2.2 *Employee hiring*

The labour market is modelled in a simple manner. All unemployed actors seek employment, and all employers hire if they have sufficient *ex ante*

---

[1] More generally, each uniform distribution can be considered as a default functional parameter of the model, which may be replaced with a different distribution that has empirical support.

funds to pay the average wage. The wage interval, $\omega = [w_1, w_2]$, is a fixed, exogenous parameter to the model. Wages are randomly chosen from the wage interval according to a uniform distribution; hence the average wage is $\langle w \rangle = (w_1 + w_2)/2$.

Hiring of employees by firms is controlled by a hiring rule:

Hiring rule : (Stochastic).

(1) If actor a is unemployed then:
  (a) Form the set of potential employers, H, consisting of all non-employees.
  (b) Select an employer, $c \in H$, according to a probability function that weights potential employers by their wealth.
  (c) If c's cash holdings m_c exceeds the average wage, then c hires a.

The hiring rule allows all non-workers to potentially hire employees, including hiring by other unemployed individuals to form new firms, but the chances of hiring favour those employers with greater wealth, a stochastic bias that represents the tendency of firm growth to depend on accumulation of capital out of current profits (Kalecki 1954). But the stochastic nature of the rule reflects the innumerable concrete reasons why particular firms are willing and able to hire more workers than others. Note that the rule does not imply that workers know the money holdings of potential employers, only that the wealthy firms probably hire more people.

9.2.3 Expenditure on goods and services

Each actor spends its income on goods and services produced by firms. But the particular purchases of an individual actor are not modelled. Instead, they are aggregated into a single amount that represents the actor's total expenditure for the month. The total expenditure can represent multiple small purchases, a single large purchase, or a fraction of a purchase amortized over several months: the interpretation is deliberately flexible. Absent a theory of consumption patterns the only relevant information is that expenditure is constrained by the amount of money an actor has. For simplicity assume that the amount spent is bounded by the actor's coin endowment on a randomly selected day. A *consumer actor* is selected to spend its income but the spent income is not immediately transferred to firms. Instead, it is added to a pool of market value that represents the currently available sum of consumer expenditures, which firms compete for.

Expenditure rule : (Stochastic).

(1) Randomly select a consumer b other than the current actor a according to a uniform distribution.

(2) Randomly select an expenditure amount, m, according to a uniform distribution, from the budget set by b's cash holding.

(3) Transfer the m cash from b to the available pool of market value, V.

This rule controls the expenditure of all consumers, whether workers, capitalists or unemployed. Clearly, a rich actor is more likely to spend more.

Different classes spend for different reasons, in particular workers normally spend their incomes on consumption goods, whereas capitalists not only consume but invest. The payment of wages is treated separately, and therefore capitalist expenditure is interpreted as expenditure on non-wage goods, such as capital goods or personal consumption. The expenditure rule is also implicitly a saving rule as in a given period the probability of an actor spending all its wealth is low.

9.2.4 Interaction between firms and the market

To simplify matters assume that all means of production are controlled by capitalist owners and therefore individual actors are unable to produce. Self-employment is ignored in this model: productive work resulting in saleable goods or services is performed only by actors within firms.

Each firm produces some collection of use-values that it attempts to sell in the marketplace. But individual commodity types and sales are not modelled. Instead, the total volume of a firm's sales in a given period are disaggregated into *market samples*, which are transfers of money from marketplace to seller, representing multiple separate transactions, or fractions of a single large transaction. At this level of abstraction the mapping from market samples to actual material exchanges is ignored and assumed to be arbitrary.

Under normal circumstances a firm expects that a worker's labour adds a value to the product that is bound from below by the wage. A firm's markup on costs reflects this value expectation, which may or may not be validated in the market. Obviously, there are multiple and particular reasons why a worker adds more or less value to the firm's total product, most of which are difficult to measure, as partially reflected in the large variety of contested and negotiable compensation schemes.

We will model the relationship between concrete labour and value-added by assuming that a firm randomly samples the market once for every employee. The firm samples per employee to reflect the fact that each worker potentially adds value, but samples randomly to reflect contingency and subsume the range of possibilities, from slackers to Stakhanovites, or from replaceable administrators to irreplaceable film stars. This is a weak formulation of the law of value (Marx 1954, Rubin 1973, Wright 2003b), which implies that, absent profit-equalizing mechanisms and rents, there is a statistical tendency for the value of a firm's product to be linearly related to the amount of social labour-time expended on the product.

Each firm therefore samples the market to gain revenue for every worker employed. In an idealised freely competitive economy there is a tendency for particular production advantages to be regularly adopted by competing firms, including the removal of scarcities due to employment of particular kinds of skilled labour. We can therefore assume that the determinants of the value-added per worker are statistically uniform across firms. The statistical variation can be interpreted as representing transient differences in the productivity of different concrete labours.

Although different workers may be more or less productive the value realised from their labour is constrained by the overall level of demand in the market. The value-added by an active worker to the firm's product is represented by a transfer of money from the current available market value V. The actual value received in money-form depends on the prevailing market conditions, and mismatches between value and exchange-value, or more plainly, costs and revenue, determine whether firms are rewarded with profits for performing socially-necessary labour.

The revenue received from the market is the legal property of the capitalist owner. Capitalist owners therefore accrue revenue via market sales that represent the social utility of the efforts of their workers. All these abstractions are expressed in the following market sample rule:

Market sample rule : (Stochastic).

(1) If *a* is not unemployed then:

(a) Randomly select a revenue amount m from the interval $[0,V]$ according to a uniform distribution (V is reduced by m.)

(b) If actor a is an employee then transfer m coins to the employer (hence the employers cash is increased by m.)

(c) Alternatively, if actor a is a capitalist owner, then transfer m coins to actor a (hence the employers cash is also increased by m).

In either case the transferred coins are counted as firm revenue. In the first case we are modelling the way that a worker contributes to the firm's income, in the second we are modelling the way that capitalist owners also contribute to firm revenue by their work. We are assuming that the expected contribution of an employee or employer to the firms revenue will be the same. This, of course, applies only to the expected contribution: the individual contributions of actors will vary randomly. In a real economy higher motivation might make non-absent employers contribute more per day than their employees. But we ignore this for simplicity.

The money received may represent value embodied in many different kinds of products and services that are sold in arbitrary amounts to arbitrary numbers of buyers. The market sample rule abstracts from the details of individual market transactions and may be interpreted as modelling the aggregate effect of a dynamic random graph that links sellers to buyers in each market period. The stochastic nature of the rule subsumes innumerable reasons why particular firms enjoy particular revenues: the only constraints are that revenue received is determined by the available value in the marketplace, and that a firm with more employees will on average sample the market on more occasions than a firm with fewer employees, a bias justified by the law of value.

A firm may enjoy a sequence of high value samples of the market, which can be interpreted as the result of a competitive advantage, for example, highly productive workers or advanced capital equipment. However, each sample is independent, hence we abstract from the possibility that the value-added by workers in the same firm is correlated over a time period.

9.2.5 Employee firing

If the revenue received by a firm is insufficient to pay the wage bill then the employer must reduce costs and fire employees. This is captured by the following firing rule:

Firing rule : (Deterministic).

(1) If actor a is an employer, then determine the number of workers to fire, u, according to the rule that no workers are fired if the *ex ante* wage bill is payable from the firm's current money holdings (the wage bill is calculated from the average wage and the number of employees). Otherwise, the firm's workforce is reduced to a size such that the wage bill is payable.

(2) Select the u actors from the set of employees, according to a uniform distribution, and fire them.

In this model there are no skill differences therefore each actor is identical. It does not matter which particular workers are fired, simply the amount, and so the particular individuals to fire are chosen randomly. Note the asymmetry between hiring and firing: hiring occurs one individual at a time at a frequency determined by the number of unemployed actors, whereas firing may occur in bulk at a frequency determined by the number of firms. Just as new firms may form when two actors enter an employee-employer relationship, existing firms may cease trading when all employees are fired and the capitalist owner enters the unemployed class.

9.2.6 Wage payment

Employers pay wages according to the following rule, which implements the transfer of value from capitalist to worker.

Wage payment rule : (Stochastic).

(1) For each actor e that a employs

    (a) transfer w in cash from a to e, where w is selected from the discrete interval $[w_1, w_2]$ according to a uniform distribution. (If employer a has insufficient funds to pay w then w is selected from the employer's current cash holdings according to a uniform distribution.)

In reality wages are not subject to monthly stochastic fluctuations. A more elaborate model would introduce wage contracts between employer and employee that fix the individual employee's wage for the duration of employment. But in the aggregate, for example in terms of the total wage bill on average payable by a firm, or wage and profit shares in national income, the existence of monthly fluctuations in individual wages is not significant, and allows a considerable simplification of the model.

9.2.7 Historical time

Finally, the above rules are combined and repeatedly executed to simulate the functioning of the economy over time. The following simulation rule orders the possible economic actions:

> Simulation rule: Allocate M/N in cash to each of the N actors; that is we set all actors to have equal wealth at the start. Also set all actors to be initially unemployed.
>
> (1) Execute the actor selection rule to select the active actor a.
>
> (2) Execute the hiring rule.
>
> (3) Execute the expenditure rule that augments the available market value with new expenditure.
>
> (4) If a is associated with a firm, execute the market sample rule that transfers m in cash from the market to the firm owner.
>
> (5) Execute firing rule.
>
> (6) Execute wage payment rule.

The application of this simulation rule can generate a variety of events. For example, if the active actor is unemployed it may get hired by an existing firm, or with lower probability form a new small firm with another unemployed actor. An employed active actor will generate a market sample for its employer, which generates revenue bound by the available market value, itself a function of the stochastic spending patterns of other actors. If the active actor is a capitalist owner of a firm it may decide to fire employees if current revenues do not cover the expected wage bill. If all employees are fired then the firm ceases trading. Otherwise, the wage bill is paid, augmenting the spending power of the working class, which on the next cycle will affect the available market value that firms compete for, and so on.

A period of one month is defined as the *N* applications of the simulation rule. This means that on average wages are paid once per simulated month.

<u>One month rule </u>:

(1) Execute the simulation rule.

(2) Repeat *N* times.

The rule is executed *N* times to allow each of the *N* actors an opportunity to act. But clearly this does not guarantee that each actor will in fact act within the month: some actors may act more than once, others not at all. This introduces a degree of causal slack that is intended to model the fact that in real economies events do not occur with strict regularity. In addition, the repeated random selection of active actors during a simulated month breaks any symmetries that might be introduced if actors are selected in a regular order. In reality, economic actions occur both in order and in parallel and this causal chaos is modelled by noisy selection.

A period of one year, which is the accounting period, is defined as 12 applications of the one month rule. The model is therefore given a notional time scale loosely linked to real time via the empirical fact that on average wages are paid once each month.

The set of rules discussed above, and three parameters – the total cash in the economy *M*, the total number of actors, *N*, and the fixed wage interval ω – constitute a dynamic, computational model of the social architecture (SA) of capitalist production.

9.3 RESULTS

Over the last 20 years it has become apparent that very simple computational models can generate complex behaviours (Wolfram 2002). The rules of the computational model described here are also simple, yet the dynamic behaviour they generate is rich and complex.

The total number of coins, *M*, and the total number of actors, *N*, on condition that $M >> N$, appear to act as scaling parameters and do not affect the relative dynamics, unlike the wage interval parameter. The computational rules do not refer to absolute numbers of coins or actors, hence a doubling

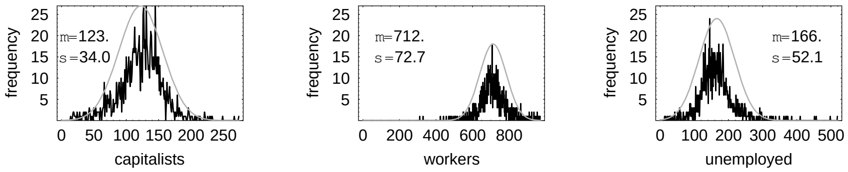


Figure 9.1: Class distributions: histograms of the number of actors in each economic class with a constant bin size of 1. The smooth lines are fitted normal distributions. On average approximately 71.2% of the population are workers, 12.3% are capitalists employing one worker or more, and the remaining 16.6% are unemployed.

of both leaves wealth per actor unchanged. Similarly, increasing the number of coins scales the overall wealth and income levels, all other things being equal. As opposed to this, if the number of actors is very small the model behaves qualitatively differently. But real economies are composed of millions of people, so we do not examine such edge cases.

The computational rules refer to the absolute wage, and hence changes to the wage parameter affect the emergent dynamics. In all reported results, $N = 1000$ and $M = 100000$, so that the average wealth in the economy is 100 coins. On *a posteriori* grounds the wage interval is set to $\omega = [10, 90]$; hence, the minimum wage is 10 coins, the average wage 50 coins, half the mean wealth in the economy, and the highest possible wage never exceeds the mean wealth. This results in an almost equal split of national wealth between the two classes, and is designed to be in agreement with the general predictions of Farjoun and Machover.

When we start the simulation running, it very rapidly organises itself into a stochastic equilibrium. In this equilibrium, whilst individual economic variables fluctuate, the probability distributions of these variables do not change over time. The simulation does not settle to a motionless equilibrium but converges to a dynamic equilibrium of ceaseless motion and change.

Unless stated otherwise the model was allowed to run for 100 simulated years.

9.3.1 Class distribution

The social stratification generated by capitalist economies is a complex phenomenon with systematic causal relations to the dominant social relations of production. In reality the social relations of production are more complex than the relations in the SA model (actors may receive combinations of wage and property income and therefore belong to more than one economic class, some actors are self-employed, others receive the majority of their income from rent, many people work for governments rather than private enterprises, and so forth). In consequence, some work is required to map empirical data on social stratification to the more basic categories employed here. It is equally clear, however, that the class of capitalists is numerically small, whereas the class of workers, that is those actors who predominately rely on wage income for their subsistence, constitute the vast majority of the population. The SA model should reflect this empirical fact.

Figure 9.1 is a group of histograms showing class sizes generated by the model collected over the duration of the simulation. The number of workers, capitalists and unemployed are normally distributed. The normal distributions summarise a dynamic process of individual social mobility, where actors move between classes during their imputed lifetimes, occurring within a stable partition of the population into two main classes – a small employing class and a larger employed class. Fluctuations in class sizes are evidently mean-reverting, reflecting stable and persistent class sizes, given the pre-specified and constant wage interval. The unemployment rate is higher than is usually reported in modern economies, but published measures of unemployment typically under-report actual unemployment. For example many people who might work but are not eligable for unemployment benefits are not counted, whereas here all non-employed actors are considered unemployed. In addition, there is no concept of self-employment. In conclusion, the SA model self-organises into a realistic partition of the working population into a minority of employers and a majority of employees.

9.3.2 Class distribution of income

GDP (which we label X) is the sum of revenues received by firms during a single year. Firms pay the total wage bill, W, from this revenue. Hence the total value of domestic output is divided into a share that workers receive as wages, $X_w = \frac{W}{X}$, and the remainder that capitalists receive as profit,

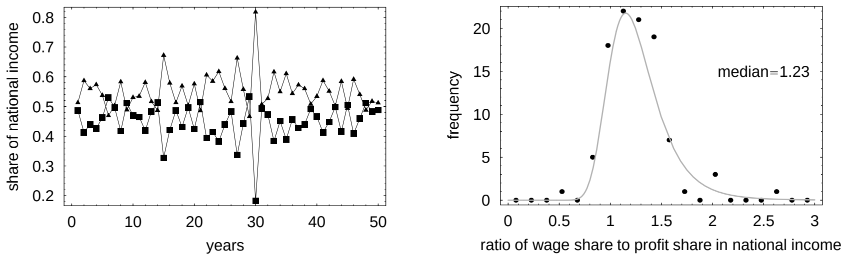


Figure 9.2: Wage and profit shares in national income. The LHS graph is a representative time series of the fluctuating shares in national income. GDP, denoted X, is the sum of revenues received by firms during a single year. The solid triangles are the wage share, X_w, which represents the sum total of wages paid to the working class, W, divided by GDP, $X_w = \frac{W}{X}$. The solid squares are the profit share, X_p, which represents the sum total of profits received by the capitalist class divided by GDP, $X_p = 1 - \frac{W}{X}$. The wage share fluctuates around a mean of 0.55 and the profit share fluctuates around a mean of 0.45. The RHS graph is a histogram of the ratio $\frac{W}{1-W}$. The smooth line is a fitted probability distribution of a ratio of two normal variates, which indicates that fluctuations of shares in national income are normally distributed around long-term stable means.

$X_p = 1 - X_w$. Advanced capitalist countries publish national income accounts that allow wage and profit shares to be calculated, which reveal some characteristic features. Shares in national income have remained fairly stable during the twentieth century, despite undergoing yearly fluctuations. For example, the profit share, normally lower than the wage share, is between 0.25 to 0.4 of GDP, although it occasionally can be as high as 0.5 (source: the calculations of Foley and Michl (1999) for the US, UK and Japan spanning a period of over 100 years; other authors place the wage share nearer to $\frac{1}{2}$, for example on average 0.54 between 1929 and 1941 for the USA (Kalecki 1954) and similar in chapters 3 and 8 of Farjoun and Machover (1983)).

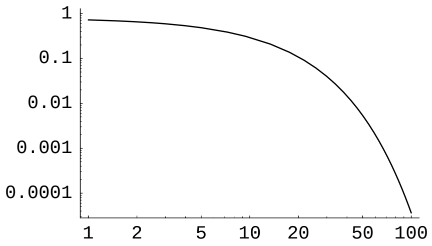
In the model we compute the wage share as $X_w = \frac{W}{X}$ where W is the total wages paid during the year, and X is the total firm income during the year. Figure 9.3.2 is a plot of the shares in national income generated by the model. The profit share is generally lower than the wage share, and the yearly fluctuations are normally distributed about long-term stable values. Ignoring differences of definition, and for the purposes of a rough and ready comparison, the model generates an average profit share of 0.45, which compares well to the empirical data. The model therefore reproduces the empirical situation of fluctuations about a long-term stable mean, and additionally the profit and wage shares have realistic values, although it is an open question whether suitably de-trended fluctuations are normally distributed in capitalist economies.

9.3.3 Disaggregated income distributions

The income shares produced by the model can be disaggregated and measured at the level of individuals in order to understand income differentiation within classes.

The empirical income distribution is characterised by a highly unequal distribution of income, in which a very small number of households receive a disproportionate amount of the total (e.g., using wealth as an indicator of income, in 1996 the top 1% of individuals in the US owned 40% of the total wealth (Levy and Solomon 1997)). The higher, property-income, regime of the income distribution can be fitted to a Pareto (or power) distribution (Levy and Solomon 1997, Matteo et al. n.d., Levy and Solomon n.d., Dragulescu 2003, Nirei and Souma 2003a, Souma 2000, Nirei and Souma 2003b), whereas the lower, or wage-income, regime, which represents the vast majority of the population, is normally fitted to a lognormal distribution (Souma 2000, Montroll and Shlesinger 1983, Badger 1980), but recently some researchers report that an exponential (Boltzmann-Gibbs) distribution better describes the empirical data (Nirei and Souma 2003b, Dragulescu 2003, Dragulescu and Yakovenko 2002). Plotting the income distribution as a complementary cumulative distribution function (ccdf) in log-log scale reveals a characteristic ‘knee’ shape at the transition between the two regimes (Matteo et al. n.d., Dragulescu 2003, Dragulescu and Yakovenko 2002, Souma 2000, Nirei and Souma 2003b). The functional form of the income distribution is stable over many years, although the parameters seem to fluctuate within narrow bounds. For example, for property-income, the power-law, $P(x) \propto x^{-(\alpha+1)}$, has a value $\alpha = 1.3$ for the UK in 1970 Levy and Solomon (n.d.), $\alpha = [1.1, 1.3]$ for Australia between 1993 and 1997 Matteo et al. (n.d.), $\alpha = 1.7$ for US in 1998 Dragulescu (2003), on average $\alpha = 1.0$ for post-war Japan Nirei and Souma (2003a), and $\alpha = [0.5, 1.5]$ for

Digression 9.1 Boltzmann-Gibbs distribution



The Boltzmann-Gibbs distribution describes the probability distributions of energies of particles in a thermodynamic system. It has the general form $P(\varepsilon) = Ce^{-\frac{\varepsilon}{T}}$, where ε denotes the energy of a particle. The graph above shows the shape of the Boltzmann-Gibbs distribution on a log-log plot. The Boltzmann-Gibbs distribution is a particular example of a *negative exponential* distribution.

This distribution arises as a consequence of the fact that in a closed system of particles the total energy must be conserved but random energy exchanges between these particles cause the energy to be spread through the population in a particular pattern. The probability of an individual particle successively gaining additional energy from a sequence of exchanges is quite low. So we would expect to see most particles having low energy (the most likely cases), but a small number with a disproportionately large amount of energy (the exceptional cases). The plot shows that the probability of a particle with the highest energy (100) is very low.

Dragulescu and Yakovenko (2000, 2002) have argued that since money is conserved in the exchange of commodities the distribution of money should follow a similar functional form. This is approximately true for the lower, predominately employee regime of the income distribution of capitalism. We will see in Chapter 10 that the assumption of conservation of money can only be held to a limited extent in a capitalist economy with modern banks.

US and Japan between 1960 and 1999 Nirei and Souma (2003b). In sum, the income distribution is asymptotically a power-law with shape parameter $\alpha \approx 1.0$, and this regime normally characterises the top 1% to 5% of incomes.

The two-parameter lognormal distribution

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(\frac{-(\log\frac{x}{m})^2}{2\sigma^2}) \tag{9.1}$$

where m is the median, and $\beta = 1/\sqrt{2\sigma^2}$ is the Gibrat index, can describe the remaining 95% or so of incomes. For example, for post-war Japan, the Gibrat index ranges between approximately $\beta = 2.25$ and $\beta = 3.0$ (Souma 2000). In contrast, if the lower income range is fitted to an exponential law

$$P(x) \propto \lambda \exp^{\lambda x} \tag{9.2}$$

then by analogy with a perfect gas, from which the Boltzmann-Gibbs law originates, λ is interpreted as an average economic 'temperature', which should be close to the average wealth in the economy, adjusting for the effects of the Pareto tail.

The SA model is in close qualitative and quantitative agreement with all these empirical facts. It also explains why there are two major income regimes, and provides a candidate explanation of why the distribution of low incomes is sometimes identified as either lognormal or exponential.

Figure 9.3 is a plot of the stationary income ccdf generated by the model. It reproduces the characteristic 'knee' shape found in empirical income distributions. The 'knee' is formed by the transition from the lower regime, consisting mainly of the wealth of the working class and owners of small firms, to the higher regime, consisting mainly of the wealth of the capitalist class. The knee occurs at around $P(M \geq m) = 0.1$, which means the power-law regime holds for at most 10% of incomes.

Figure 9.4 splits the income distribution according to class. The capitalist distribution has a long tail, qualitatively different from the worker distribution, which is clustered around the average wage.

Figure 9.5 is a plot of the lower regime of the income distribution in log-linear scale fitted to a lognormal distribution with Gibrat index $\beta = 1.42$.

Figure 9.6 is a plot of the property-income regime in log-log scale. The straight line fit indicates that higher incomes asymptotically approach a power-law distribution of the form $P(x) \propto x^{-(\alpha+1)}$, with $\alpha = 1.3$. The two income regimes are consequences of the two major sources of income in capitalist societies, that is wages and profits, and the overall income distribution is a mixture of two qualitatively different distributions. The lower

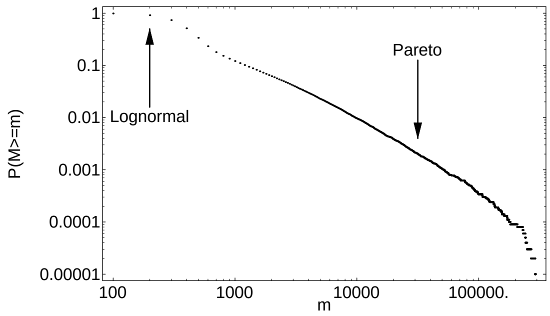


Figure 9.3: The complete income distribution plotted as a ccdf in log-log scale. The data is binned at a constant size of 1. Note the characteristic ‘knee’ shape, a feature found in empirical distributions. The transition from the lognormal to the Pareto regime occurs between $P(x) = 0.1$ and $P(x) = 0.01$, which means that under 10% of incomes follow the Pareto law.

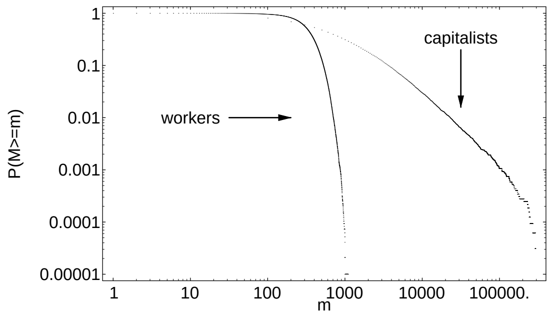
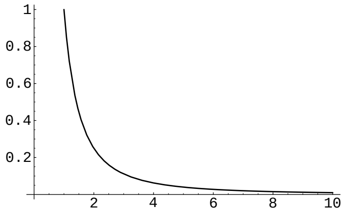


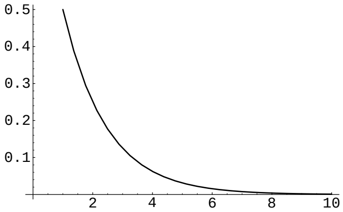
Figure 9.4: The class components of the income distribution plotted as ccdfs in log-log scale. Note the long tail of the capitalist income distribution. Worker income is clustered around the average wage.

Digression 9.2 Power law distributions

Many man-made and naturally occurring phenomena, including city sizes, incomes, word frequencies, and earthquake magnitudes, are distributed according to a power-law distribution. A power-law implies that small occurrences are extremely common, whereas large instances are extremely rare. The Boltzmann-Gibbs distribution also has this characteristic, that large energy values are very rare. So how do the two distributions differ?



Power Law Distribution x^{-2}



Negative Exponential Distribution 2^{-x}

Look at the two graphs above. Negative Exponential distributions, e.g. the Boltzmann-Gibbs, fall off more sharply compared to a power law distribution.

A power law distribution has a longer 'tail' to the right. If personal wealth is governed by a power law it means that there will be more very rich people than there would be if wealth was governed by a negative exponential distribution.

A power law has the general form of probability density function $P[X = x] = Cx^{-a}$, which is a formula that gives the probability that a person's income is exactly £20,000 (e.g., $P[X = 20000]$). The Pareto distribution is the cumulative distribution function corresponding to a power law. It is generally written as $P[X > x] = x^{-k}$, for example the probability that a person's income is greater than £20,000. It is related to the power law distribution by the formula $a = k + 1$.

regime is fitted better by a lognormal distribution rather than an exponential. The lognormal distribution, in this model, is not the result of stochastic multiplicative process, which is the explanation often proposed, but results from a mixture of normally distributed wage incomes and the profit-income of small firm owners. It is an open question whether the lognormal distribution found in empirical data can be similarly explained by the combined effect of income from employment and the income of small employers.

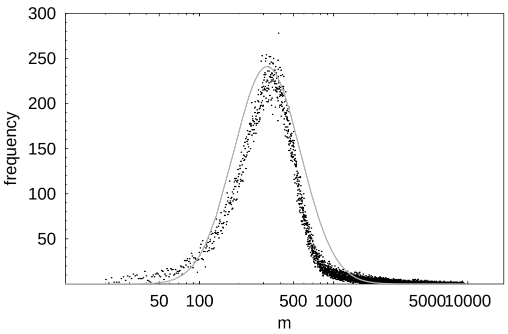


Figure 9.5: The lower regime of the income distribution plotted in log-linear scale. The solid line is a fit to a lognormal distribution. The approximately lognormal distribution results from a mixture of wage income and small employer income.

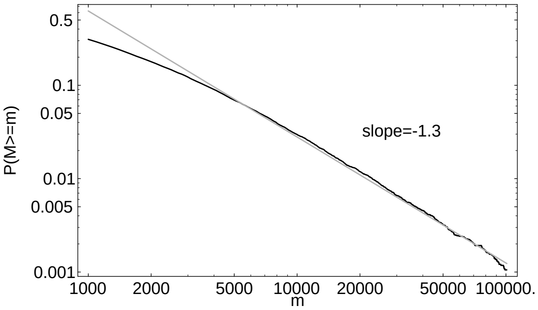


Figure 9.6: The power law regime of the income distribution plotted as a ccdf in log-log scale. The straight line is a fit to the power (Pareto) law, $P(x) \propto x^{-(\alpha+1)}$, where $\alpha = 1.3$.

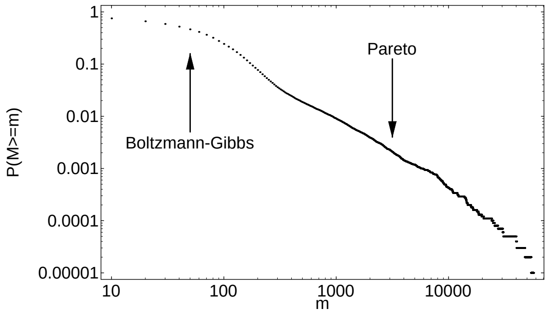


Figure 9.7: The complete money distribution plotted as a ccdf in log-log scale. The transition from the Boltzmann-Gibbs to Pareto regime occurs in the middle of the ccdf. The data is binned at a constant size of 1.

At first glance it appears that the model contradicts empirical evidence that the lower income regime is exponentially distributed. But if the stationary distribution of money holdings (i.e., instantaneous wealth) is measured, rather than income, a different picture emerges, which may help explain the lack of consensus in empirical studies. Wealth in our model can be measured by the total money held by each actor at the end of the year.

Figures 9.7 to 9.10 are plots of the stationary money ccdf generated by the model. Figure 9.7 reproduces the characteristic 'knee' shape found in empirical income distributions. But in this case the lower regime is characterised by an exponential (or Boltzmann-Gibbs) distribution. The transition between regimes occurs approximately in the middle of the ccdf corresponding to a situation in which the total wealth in the economy is distributed approximately evenly between the classes. Figure 9.9 plots the workers' money distribution in log-linear scale. The straight line fit reveals an exponential distribution of the form $P(x) = \lambda \exp^{\lambda x}$, where $\lambda = 0.017$, which is reasonably close to the average wealth in the economy, $\lambda = \frac{M}{N} = 0.01$ (Dragulescu and Yakovenko 2000). Figure 9.10 plots the capitalists' money distribution in log-log scale. The straight line fit reveals a power-law distribution with similar exponent to that of income.

The higher income and wealth regimes are qualitatively identical, but the lower income and wealth regimes are qualitatively distinct. Measuring the lower end of income yields a lognormal distribution, whereas measuring

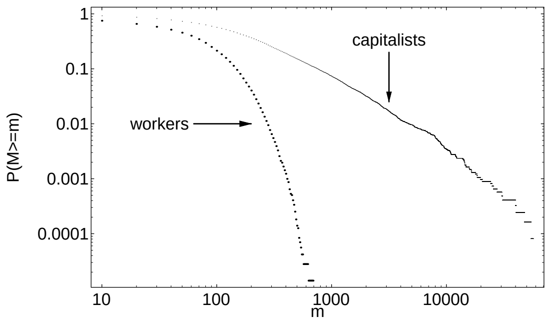


Figure 9.8: The class components of the money distribution plotted as ccdfs in log-log scale. Note the long tail of the capitalist money distribution.

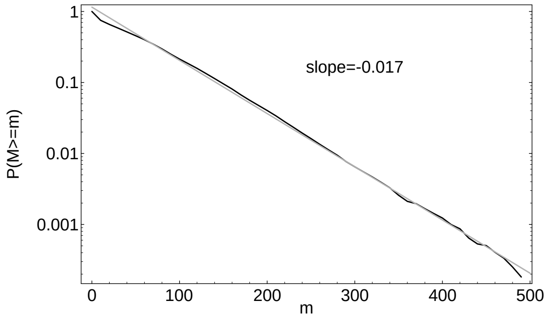


Figure 9.9: A section of the workers' money ccdf plotted in linear-log scale. The straight line is a fit to the exponential (Boltzmann-Gibbs) law, $P(x) = \lambda e^{\lambda x}$, where $\lambda = 0.017$.

the lower end of wealth yields an exponential. Income depends solely on monies received during an accounting period, whereas wealth depends on both income and spending patterns. The differences between the empirical studies could be due to differences in whether the measures employed are predominately income measures or wealth measures.

The lognormal and power-law fits are only approximations to the true distributions, and we do not embark on a full analysis of the income distribution here. However, a few brief points can be made. A popular explanation of the power-law tail of the income distribution is that it arises from

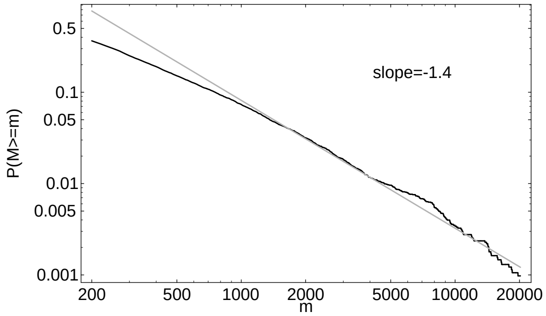


Figure 9.10: A section of the capitalists' money ccdf plotted in log-log scale. The straight line is a fit to the power (Pareto) law, $P(x) \propto x^{-(\alpha+1)}$, where $\alpha = 1.4$.

an underlying stochastic multiplicative process, often thought to model the geometric growth of capital invested in financial markets (Nirei and Souma 2003b,a, Reed 2000, 2001, Levy and Solomon n.d., 1997, Bouchaud and Mezard n.d.). The importance of financial markets in determining capital flows and hence capitalist income is undeniable. But the model developed here shows that an income power-law can arise from industrial capital invested in firms, absent financial markets that support capital reallocation between industries or between capitalists. Capitalist income, in this model, is not derived from investment in portfolios that provide a return, but is composed of the sum of values added via the employment of productive workers.

It is remarkable that the model's simple rules generate detailed income distributions in close agreement with reality. It seems very likely, therefore, that the fundamental reason for the observed income distribution in capitalism is due to the way firm revenue is distributed: as wages to workers, and profits to capitalist owners. There are two major ways of getting money in capitalism: by working, or by employing. Hence there are two, qualitatively distinct income regimes, the negative exponential for the majority, and the Pareto for the few.

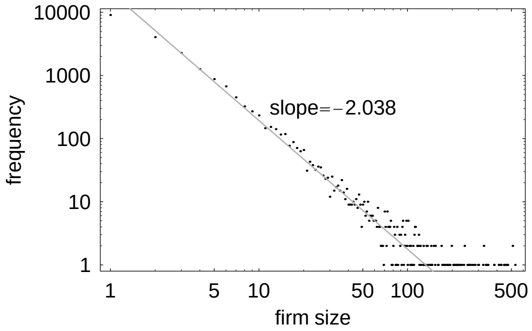


Figure 9.11: Firm size distribution: histogram of firm sizes by employees in log-log scale with a constant bin size of 1. The straight line is an ordinary least squares regression of the data and represents a power-law distribution $P(x) \propto x^{-(\alpha+1)}$ with exponent $\alpha = 1.038$ for data collected over 15 simulated years. Axtell (2001) reports $\alpha = 1.059$ from data of approximately 5.5 million U.S. firms in 1997. The special case $\alpha = 1$ is known as the Zipf distribution.

9.3.4 Firm size distribution

Axtell (2001) analysed US Census Bureau data for US firms trading between 1988 and 1997 and found that the firm size distribution followed a special case of a power-law known as Zipf's law, and this relationship persisted from year to year despite the continual birth and demise of firms and other major economic changes. During this period the number of reported firms increased from 4.9 million to 5.5 million. Gaffeo et al. (2003) found that the size distribution of firms in the G7 group over the period 1987-2000 also followed a power-law, but only in limited cases was the power-law actually Zipf. Fujiwara et al. (n.d.) found that the Zipf law characterised the size distribution of about 260,000 large firms from 45 European countries during the years 1992–2001. A Zipf law implies that a majority of small firms coexist with a decreasing number of disproportionately large firms.

Firm sizes in the SA model are measured according to the number of employees they have. The model also replicates the empirical firm size distribution. Figure 9.11 is a histogram of firm sizes. The straight line is a fit to the power-law:

$$P(x) \propto x^{-(\alpha+1)} \tag{9.3}$$

For data collected over a relatively short time period, such as 15 simulated years, α approaches 1.0. The special case $\alpha = 1.0$ is Zipf, and hence the firm size distribution generated by the model is consistent with the empirical data. Data collected over shorter periods follows a power-law with exponent that deviates from 1.

The largest US firm in 1997 had approximately 10^6 employees from a total reported workforce of about 10^7 individuals (Axtell 2001). Therefore, the largest firm size should not exceed about $\frac{1}{10}$th of the total workforce. Figure 2 shows that, with low but non-zero probability, a single firm can employ over half the workforce, representing a monopolisation of a significant proportion of the economy by a single firm, a clearly unrealistic occurrence. A possible reason for the over-monopolisation of the economy is the assumption that firms have a single capitalist owner, which conflates capital concentration with firm ownership. In reality, large firms normally have multiple owners and individual capitalists own multiple firms. Further, there are many technical reasons why particular firms do not grow beyond a certain size that are ignored in this model. A final point is that the probability of monopoly within the period of observation decreases with the number of actors; hence, if the simulation were run with $N = 10^7$ actors (which is not possible due to insufficient computational resources) then it is unlikely that a single firm would employ half the workforce. Gaffeo et al. (2003) note that firms are distributed more equally during recessions than during expansions, which accounts for the yearly deviations from Zipf. We have not tested this relationship in the SA model.

9.3.5 Firm growth

Stanley et al. (1996) and Amaral, Buldyrev, Havlin, Leschhorn, Maass and Salinger (1997) analyzed the log growth rates of publicly traded US manufacturing firms in the period 1974 – 93 and found that growth rates, when aggregated across all sectors, appear to robustly follow a Laplace (double exponential) form. This holds true whether growth rates are measured by sales or number of employees. More precisely, if the annual growth rate is $r = \ln(\frac{s_{t+1}}{s_t})$, where s_t is the size of a firm in year t, then for all years the probability density of r is consistent with an exponential decay:

$$f(r) \propto e^{-\left|\frac{r-\alpha}{\beta}\right|} \tag{9.4}$$

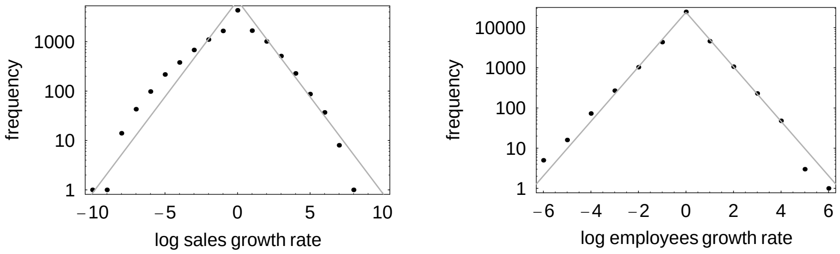


Figure 9.12: Firm size growth rate distribution: histogram of the log growth rates of firms per simulated year in linear-log scale with a constant bin size of 1. The LHS graph shows growth rates of firm sales. The RHS graph shows growth rates of employees. The solid lines are OLS regressions of the data and represent a Laplace (double-exponential) distribution $P(x) \propto e^{-|(x-\alpha)/\beta|}$. Many researchers report that log growth rates of sales and employees of US and Italian firms follow a Laplace distribution (Lee et al. 1998, Bottazzi and Secchi 2003, Stanley et al. 1996, Amaral, Buldyrev, Havlin, Leschhorn, Maass and Salinger 1997, Amaral, Buldyrev, Havlin, Maass, Salinger, Stanley and Stanley 1997, Amaral et al. 2001, Fabritiis et al. n.d.).

with some deviation from the Laplace distribution at high and low growth rates resulting in slightly ‘fatter wings’ (Lee et al. 1998, Amaral, Buldyrev, Havlin, Maass, Salinger, Stanley and Stanley 1997, Amaral et al. 2001). Bottazzi and Secchi (2003) replicate these findings and report a Laplace growth distribution for Italian manufacturing firms during the period 1989–96.

We can measure firm growth in the SA model in terms of the change in the number of employees or in sales from year to year. The model generates log annual growth rates for firms that are consistent with a Laplace distribution, whether growth is measured in terms of sales or number of employees. Figure 9.3.5 plots log growth rates in log-linear scale and reveals the characteristic ‘tent’ shape signature of a symmetric exponential decay. In the SA model there is no net growth in population or in monetary stocks, this means that the mean rate of growth of a firm will be zero. There is some deviation from a Laplace distribution for firms with shrinking sales, which may be due to noise or represent some non-accidental property.

The replication of the empirical Laplace growth distribution suggests that the social relations of production may play an important role in constraining the dynamics of firm growth. Lee et al. (1998), Fabritiis et al. (n.d.), Amaral et al. (2001), Amaral, Buldyrev, Havlin, Maass, Salinger, Stanley and Stanley (1997), Stanley et al. (1996) note that the standard deviation (std) of growth rates decreases as a power law with size, that is, $\ln\sigma(r) \sim -\beta \ln r$, where $\beta \approx 0.15$. The SA model does not replicate this finding given the specified wage interval. In fact, $\ln\sigma(r)$ appears to increase as a power law with size, although the data is quite noisy. However, the exponent of the power law is sensitive to the wage parameter, and it is possible to replicate the empirical relationship at lower average wages. Explanations of the relationship between growth variation and size assume that firms have internal structures such that increased size lessens market risk (Amaral et al. 2001, Amaral, Buldyrev, Havlin, Maass, Salinger, Stanley and Stanley 1997), which contrasts with the simple firm structure employed in this model. Axtell (1999), for example, presents an actor-based model of the life-cycle of firms that replicates the Zipf size distribution, Laplace growth rates and power-law scaling of the std of growth. In Axtell's model firms have a richer internal structure compared to firms in this model.

9.3.6 Firm deaths

Cook and Ormerod (2003) report that the distribution of US firm deaths per year during the period 1989 to 1997 is closely approximated by a lognormal distribution, and note that the number of deaths varies little from year to year with no clear connection to recession or growth.

We can measure the number of firm deaths per month in the simulation. A firm dies if it fires all its employees. Demises per month are measured rather than per year in order to avoid bucketing the data. Figure 9.3.6 is a histogram of firm deaths per month with a fitted lognormal distribution. It shows that the model generates a distribution of firm deaths that is approximated by a lognormal distribution and is therefore consistent with empirical findings.

According to Cook and Ormerod the average number of firms in the US during the period 1989 to 1997 was 5.73 million, of which on average 611,000 died each year. So roughly 10% of firms die each year. In the simulation on average 18 firms die each month and therefore on average 216 firms die each year, a figure in excess of the 123 firms that exist on average.

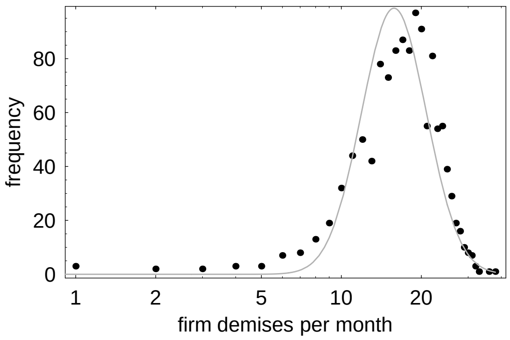


Figure 9.13: Firm deaths distribution: Histogram of firm deaths per simulated month in log-linear scale with a constant bin size of 1. The solid line is a fit to the lognormal distribution. Cook and Ormerod (2003) report that the distribution of US firm deaths per year during the period 1989 to 1997 is closely approximated by a lognormal distribution.

So although the distribution of firm deaths is consistent with empirical data, the rate at which firms are born and die is much higher than in reality. This is not too surprising when it is considered that the model entirely abstracts from the material nature of the goods and services processed by the economy and any persistent demand for them. In this model firms compete by playing a game of chance that models the unpredictability of a competitive economy. But the complete absence of the material side of the economy results in an unrealistic level of volatility in market interactions. The SA model must therefore be extended to include causal relations between the social architecture and the forces of production. Clearly there is a limit to what may be deduced from consideration of the social relations of production alone.

9.3.7 GDP growth

Gross Domestic Product (GDP) measures the value of gross production at current prices, including consumption and gross investment. Lee et al. (1998) and Canning et al. (1998) analyse the GDP of 152 countries during the period 1950–52 and find that the distribution of GDP log growth rates is consistent with a Laplace distribution, and therefore conclude that firm growth and GDP growth are subject to the same laws (Lee et al. 1998).

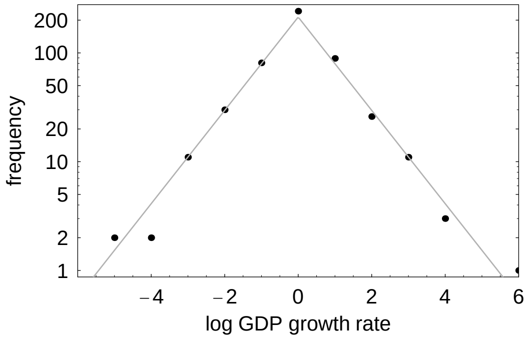


Figure 9.14: Rescaled GDP growth rate distribution: histogram of the log growth rate of GDP in linear-log scale with a constant bin size of 1. The solid lines are OLS regressions of the data and represent a Laplace (double-exponential) distribution $P(x) \propto e^{-|(x-\alpha)/\beta|}$. Lee et al. (1998) report that log GDP growth rates of 152 countries during the period 1950–92 follow a Laplace distribution.

The GDP in the SA model is measured using total firm income. Growth rates are measured year on simulated year. Empirical measurements of GDP must be detrended to remove the effects of inflation but this is unnecessary when measuring GDP in the model due to the assumption of a fixed amount of money.

Figure 9.3.7 plots log GDP growth rate for the simulated economy in log-linear scale. The data is noisy but consistent with a Laplace distribution when sampled over a period of 100 years so for clarity figure 9.3.7 contains data from an extended run of 500 years. The characteristic tent shape indicates that the SA model is consistent with the Laplace distribution of GDP growth.

Gatti et al. (2003) present an actor-based model of the life-cycle of firms that replicates the Zipf size distribution and Laplace growth rates of firms and aggregate output (GDP). They show that the power-law of firm size implies that growth is Laplace distributed and also that small micro-shocks can aggregate into macro-shocks to generate recessions. Firms in their model are not disaggregated into employees and employers and market shocks are exogenous, whereas in this model firms are composed of individuals and are subject to endogenous shocks that are the consequence of the competi-

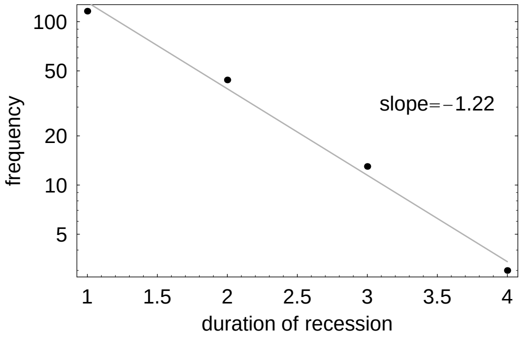


Figure 9.15: Recession duration distribution: Histogram of the frequency of the duration of recessions in log-linear scale with a constant bin size of 1. The solid line is a fit to an exponential distribution, $f(d) \propto \lambda \exp^{-\lambda d}$, with exponent $\lambda = 1.22$, representing an average recession duration of 1.22 simulated years.

tion for a finite amount of available market value, itself a product of income flows.

9.3.8 Duration of recessions

Wright (2003a), reinterpreting empirical data presented by Ormerod and Mounfield (2001), concludes that the frequency of the duration of economic recessions, where a recession is defined as a period of shrinking GDP, follows an exponential law for 17 Western economies over the period 1871–1994. Recessions tend not to last longer than 6 years, the majority of recessions last 1 year, and for the US the longest recession has been only 4 years (Ormerod 2002).

The SA model, in which recession begins when the GDP falls and ends when it ceases to fall, is in close agreement with these empirical findings. Figure 9.3.8 is a histogram of the frequency of the duration of recessions collected over a period of 500 simulated years. The functional form of the frequency of duration of recessions is exponential, $f(d) \propto \lambda \exp^{-\lambda d}$, with $\lambda = 1.22$, which compares to a value of $\lambda = 0.94$ for the empirical data (Wright 2003a). The value of λ is the average duration of a recession. Also, the duration of recessions in the model ranges from 1 to 4 simulated years.

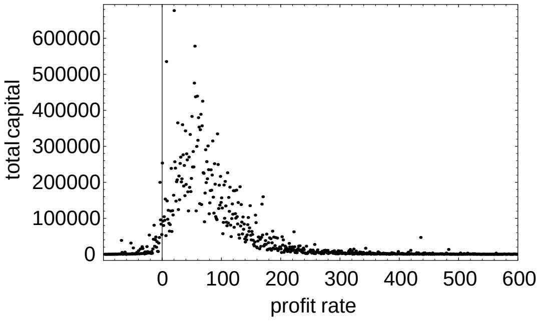


Figure 9.16: Capital-weighted rate-of-profit distribution: Histogram of amount of capital invested that generated a given percentage profit rate within a simulated year. The data is collected over the duration of the simulation and binned at a constant size of 1. The average profit rate is 80.5% and the median profit rate is 64% (on average 1 coin invested returns 1.8 coins). Wells (2001) measured the profit rate distribution of over 100,000 UK firms trading in 1981 and found that the distribution was right-skewed.

Ausloos et al. (2004) subsequently analysed a more comprehensive set of GDP data and concluded that overall the distribution of recessions follows a power-law, not an exponential law, although the matter is not entirely settled. Ormerod and Mounfield argue that economic management often prevents recessions lasting more than one year, but if they do last longer, then subjective expectations of growth become depressed and recessions may then occur on all scales of duration, resulting in a power-law. They propose that the distribution is not determined by a common set of causal factors for all durations, but instead there is a ‘breakdown of scaling’ for recessions of short duration. The SA model does not include the subjective expectations of economic actors, and therefore it is an open question whether the introduction of expectations to the model could more closely replicate the empirical data.

9.3.9 Rate-of-profit distribution

Farjoun and Machover (1983) propose that the proportion of industrial capital (out of the total capital invested in the economy) that finds itself in any given rate-of-profit bracket will be approximated by a gamma distribution by analogy with the distribution of kinetic energy in a gas at equilibrium. The gamma distribution is a right-skewed distribution. Wells (2001) examined the distribution of profit rates defined in a variety of ways of over 100,000 UK firms trading in 1981 and found right-skewness to be prevalent, but did not investigate their functional form.

In reality capitalist owners of firms invest in both variable (wages) and constant capital (investment in commodity inputs to the production process and relatively long-lasting means of production) (Okishio 1990) and the rate-of-profit is calculated on the total capital invested. The SA model abstracts from the forces of production and hence capitalist owners invest only in variable capital (i.e. expenditures on wages). Capitalists also spend income in the marketplace and this expenditure could be interpreted as either consumption or investment in constant capital, but to theoretically ground the latter interpretation the model would need to be extended to include a determination of the distribution of ratios of constant to variable capital across firms. Rather than introduce the material side of the economy, which properly belongs to future substantive extensions of the model, the rate-of-profit in the simulation is calculated on variable capital alone. Hence rate-of-profit measures will exceed those found empirically.

The rate-of-profit distribution in the model is measured according to:

> Profit rate measure: After each year calculate the profit rate for each firm trading at the close of the year. The profit rate, p_i, of firm i is defined as
>
> $$p_i = 100\left(\frac{r_i}{w_i} - 1\right) \tag{9.5}$$
>
> where r_i is the total revenue received during the year and w_i is the total wages paid during the year.

Figure 9.17 graphs the amount of capital that returned a given profit within a year. Consistent with empirical research the distribution is highly right-skewed. Wells (2001) reports that if the rate-of-profit is weighted according to number of firms, rather than capital invested, the distribution is also right-skewed and very similar in overall character, although less noisy. Figure 11 graphs the firm-weighted distribution from the simulation. It is also right-skewed, like the capital-weighted distribution, but considerably less noisy.

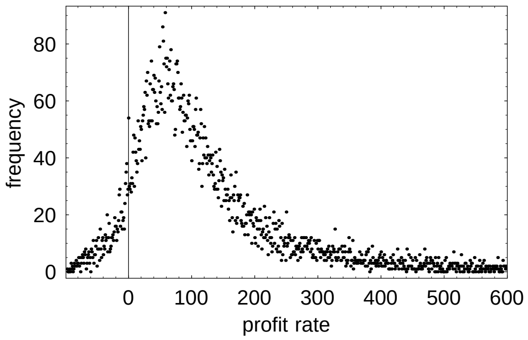


Figure 9.17: Firm-weighted rate-of-profit distribution: Histogram of number of firms that generated a given percentage profit rate within a simulated year. The data is collected over the duration of the simulation and binned at a constant size of 1. Wells (2001) measured the firm-weighted profit rate distribution of over 100,000 UK firms trading in 1981 and found that, similar to the capital-weighted rate-of-profit, the distribution was right-skewed, although less noisy.

9.4 A NOTE ON METHODOLOGY

The empirical coverage of the SA model is broad although the model can be formally stated in a small number of simple economic rules that control the dynamics. The model compresses and connects a large number of empirical facts within a single causal framework. Our aim in this chapter was to show how the social relations of production peculiar to capitalism, that is how humans relate to each other as workers and capitalists in order to produce the things they need, has a pervasive and determinate effect on many of the macro-level properties of capitalism. We can extend this modelling approach in many ways, and there are many aspects of the simulation that we could further measure and analyse, and so this chapter represents only a starting point.

The enormous benefit of exploring computational models of phenomena is that the complex dynamic consequences of a set of causal rules can be automatically and correctly deduced by running a computer program that performs a computational deduction. In this case the deduction is from micro-economic social relations to emergent, macro-economic phenomena. But the reasons why a set of causal rules necessarily generate the observed dynamic consequences may initially be opaque precisely because a computer simulation is required to perform the deduction. This is why computational modelling is not an alternative to deductive mathematical modelling but is connected to it. To give just one example, within the parameter space explored, the SA model generates fluctuations in national income about long-term stable means. But it requires a mathematical deduction to understand why this necessarily occurs. The computational model demonstrates that in principle such a deduction may be produced and its basic elements and assumptions will correspond to those of the computational model. Of course, a deductive proof may be more or less difficult to construct even if known to be possible. So one use of computational modelling is to more easily identify candidate theories, which may then be further analysed to generate explanations in the form of mathematical deductions or natural language explanations, the aim being to understand why the dynamic consequences are logically necessary. An example of the potential of this approach is the deduction of a candidate functional form for the distribution of industrial profit, which for the interested reader is discussed in appendix E. Contrast this situation to a purely deductive approach, in which the investigator may only explore candidate theories that are directly amenable to mathematical deduction. This methodology is unnecessarily restrictive, particularly if the system presents difficult analytic challenges.

The fact that the empirical distributions considered emerge from the social relations of production alone suggests that some of the striking phenomena of a capitalist economy depend not so much on specifics but on very general and highly abstract structural features of that system. In consequence, existing theories may be looking in the wrong place for economic explanations, or at least introducing redundant considerations. Given this possibility, it is worth making a few comments to contrast the approach taken in this paper to standard approaches.

The basic elements of this model differ from standard economic models. Standard competitive equilibrium models, or neoclassical models, normally take as their starting point an ontology of rational actors that maximise self-interest in a market for scarce resources (Debreu 1959). Attention is focussed on determining the equilibrium exchange ratios of commodity types, which are solutions to a set of simultaneous, static constraints. Historical time is absent, so equilibrium states are logically rather than causally derived, and typically money is not modelled. Neo-Ricardian models, in contrast, take as their starting point an ontology of technical production relations between commodity types that define the available material transformations that economic actors may perform. The production of commodities by means of commodities (Sraffa 1960) results in a surplus product that is distributed to capitalists and workers (Pasinetti 1977). Despite many essential differences, there are some important similarities between neo-classical and neo-Ricardian models. For example, prices in neo-Ricardian models are also exchange ratios determined by solutions to static, simultaneous constraints. Similarly, historical time is absent, so there is no causal explanation of how or why a particular configuration of the economy arose. Money only plays a nominal not a causal role. There are clear differences between, on the one hand, neo-classical and neo-Ricardian ontologies, and, on the other, the basic ontology of the model developed here. Most obvious is that commodity types and rational actors are absent. Instead, the model emphasises precisely those elements of economic reality that neo-classical and neo-Ricardian theories tend to ignore, specifically actor-to-actor relations mediated by money, which unfold in historical time, and result in dynamic, not static, equilibria. At a high level of abstraction, and at the risk of over simplification, neo-classical models theorise scarcity constraints, neo-Ricardian models theorise technical-production constraints, whereas this model theorises the dynamic consequences of social constraints, which are historically contingent facts about the way in which economic production is socially organised.

There is a large and longstanding literature on the failings of standard general equilibrium theory to describe economic reality. But there are deep and enduring sociological reasons why standard economic theory is resistant to criticism and persists largely unchanged. Nonetheless, the model we have described in this chapter constitutes constructive proof that the standard ontology is redundant for forming explanations of the macroeconomic phenomena we have surveyed. This is not to deny that some other, perhaps more concrete issues, may require consideration of purposive activity for their explanation and hence the introduction of rational actors, or require consideration of technical production constraints and hence the introduction of commodity types. Rather, the claim is that, for the empirical aggregates considered, there is no need to perform the standard reduction of political economy to psychology and the technical conditions of production, and further, that the dominant causal factors at work are not to be found at the level of individual behaviour, nor are they to be found at the level of technical-production constraints, but are found at the level of the social relations of production, which constitute an abstract, but nevertheless real, social architecture that constrains the possible actions that purposive individuals may choose between, whether optimally or otherwise. This is why the actors in this model probabilistically choose between possible economic actions constrained only by their class status and current money endowments, an approach that is closer to the Classical conception of political economy of Smith, Ricardo and Marx, in which individuals are considered to be representatives of economic classes that have definite relations to each other in the process of production. The social architecture, in particular the wage-capital social relation, dominates individuals, who, although free to make local economic decisions, do so in a social environment neither of their own choosing or control.

As we discussed in Chapter 6 the method of abstracting from the mechanics of individual rationality, and instead emphasising the particle nature of individuals, is valid because the number of degrees of freedom of economic reality is very large. This allows individual rationality to be modelled as a highly simplified stochastic selection from possibilities determined by an overriding social architecture. The quasi-psychological motives that supposedly drive individual actors in the rational actor approach can be ignored because in a large ensemble of such individuals they hardly matter.

9.5 ESSENTIAL AND INESSENTIAL PROPERTIES OF CAPITALISM

Our aim was to begin to understand the economic consequences of the social relations of production considered in isolation and develop a model that included money and historical time as essential elements. The theoretical motivation for the approach is grounded in Marx's distinction between the invariant social relations of production and the varying forces of production. Capitalism does change over time, but the existence of the social role of worker and capitalist are an unchanging and defining feature of it.

The model of the social relations of production replicates some important empirical features of modern capitalism, such as (i) the tendency toward capital concentration resulting in a highly unequal income distribution characterised by a lognormal distribution with a Pareto tail, (ii) the Zipf or power-law distribution of firm sizes, (iii) the Laplace distribution of firm size and GDP growth, (iv) the exponential distribution of recession durations, (v) the lognormal distribution of firm deaths, and (vi) the gamma-like rate-of-profit distribution. Also, the model naturally generates groups of capitalists, workers and unemployed in realistic proportions, and business cycle phenomena, including fluctuating wage and profit shares in national income. The good qualitative and in many cases quantitative fit between model and empirical phenomena suggests that the theory presented here captures some essential features of capitalist economies, demonstrates the causal importance of the social relations of production, and provides a basis for more concrete and elaborated models.

A final and important implication is that the computational deduction outlined in this paper implies that some of the features of economic reality that cause political conflict, such as extreme income inequality and recessions, are necessary consequences of the social relations of production and hence enduring and essential properties of capitalism, rather than accidental, exogenous or transitory.

CHAPTER 10

MONEY AND THE FORM OF VALUE

Cockshott

We have argued in previous chapters that the price of commodities tends to be fairly closely proportional to their labour values. In Section 8.2 we showed that the six or seven leading bits of information in a price derive from labour value. The question therefore naturally arises as to why prices are expressed in £ or Euro not in hours of labour.

10.1 THE FORMAL PROPERTIES OF EXCHANGE

One of Marx's criticisms of his predecessor Ricardo was that the latter had identified labour as the source of value, but he had not given an explanation of why social labour should be made manifest in money prices. In contrast to his predecessor, Marx focussed from the start, on the representation of labour in exchange value.

The first chapter of his Capital is concerned with this process of representation. It is a somewhat difficult chapter to read, but it is considered by some economists[1] to be essential to understanding Marx's whole conceptualisation of capitalism. It is a relatively formal text but not in the sense that we would now describe a scientific or mathematical text as being formal. Instead of mathematics or modern formal logic, it uses Hegelian logic to analyse the form assumed by value.

Since the mid 19th century the study of formal systems has advanced tremendously in its scope and the tools available for constructing formalisms have multiplied. In this chapter we want to construct an analysis of the value

---

[1] For example see Rubin (1973).

form using modern conceptual tools. The possibility of doing this is predicated on the fact that value and money are in the strict sense formal systems. They are systems of symbols whose time evolution is governed by formal rules analogous to the term-rewrite rules used in certain branches of computing or logic. These are programatic rules that tell you how to validly transform one algebraic formula into another. The ensemble of technical and accountancy practices of modern society can be thought of as rules operating on a vast ‘formula’ : its commercial/monetary records.

We attempt to identify what these rules and explain their necessity.

This approach is, at abstract level, similar to that pursued by Marx in that he tried to elaborate what he saw as the logically necessary consequences of the basic social forms of capitalist society : the commodity, money, and capital. He remarks “A commodity appears, at first sight, a very trivial thing, and easily understood. Its analysis shows that it is, in reality, a very queer thing, abounding in metaphysical subtleties and theological niceties[2].” He also refers the commodity having a ‘mystical’ character. We believe that these mystical attributes associated with money and commodities stems from their position within information structures. The idea of an information structure dates from the second half of the 20th century. We are now used to thinking in terms of information structures, formal languages, generative grammers etc. What was once mystical, becomes formalised and automated in the software procedures of that giant automaton : the inter-bank computer network. It is perhaps no accident that Marx’s Capital opens with an analysis of commodity exchange and and the circulation of capital that uses a formal apparattus very similar to that of generative grammars.

He introduces the notion of the circuit of capital ss $M \rightarrow C \rightarrow M'$ where M stands for a quantitity of money, M' stands for an augmented sum of money, and C for commodities purchased with the initial money. If we rewrite this as

$$M \rightarrow C$$

$$C \rightarrow M'$$

then we have in Chomsky (1956)’s terminology, a derivational phrase structured grammar that will produce the phrases:

M ,
C ,

---

[2] Marx (1954) Chap. 1. sec. 4

M' ,
C'
M'' ,
C'' ,
M''' ,
...

modelling the process of growth of a capital. Marx can almost be seen as anticipating the sort of formal analytical tool that has become commonplace in the sciences since the mid 20th century.

Our emphasis is therefore on money as an information structure and on the supporting technologies that permit this information structure to operate.

We then go on to look at money and the historical process by which labour came to be represented as money issued by the state. The conclusions that we come to regarding money differ somewhat from those of Marx, being influenced by the modern chartalist theorists: see Wray (2004), Ingham (2004), Knapp (1973). Readers interested in a modern presentation of Marx's theory of money should consult Itoh and Lapavitsas (1999) or Foley (1983). It should be pointed out that in Marx's day gold and silver coins still circulated in England, and this day to day reality undoubtedly influenced him and other contemporary economists in their understanding of money.

10.1.1 *Legally independent owners - economic subjects*

Commodity exchange presupposes the existence of economic subjects. An economic subject is an abstract category that encompasses both people and social organisations that engage in trade. The reason why economic subjects exist is two fold:

a The units of production in a society are not self sufficient.

b There exists no overall system of social direction of labour.

In a capitalist economy unit of production takes the form of an enterprise. An enterprise is a technical unit of production and, at the same time, an economic subject. It can own things and it can buy and sell things. Such economic subjects are ultimately the result of technology and a form of social division of labour. Capitalist production is social. The whole of society is involved in the division of labour. Enterprises don't produce for their own needs, they produce for society. This contrasts with the ”natural economy” of the peasant household, where the units of production and consumption coincide.

Technology forces each enterprise produce just a few types of goods, while consuming many. Enterprises consume goods produced by others which necessitates a circulation of products among the enterprises. But, as these are all subjects owning property, how can circulation take place, without a loss of property?

The only possible way is the exchange mechanism. If products are exchanged as equivalents, then there is no loss of property. An enterprise exchanges something of no use to it for something that it needs. It is not interested in what its products are used for, but in the equivalents that it can get by selling them.

The category of economic subject, is reflected juridically in the form of abstract legal personalities: Pashukanis (1989). Here is seems that it is the attributes of a person that are projected onto firms. It may be better to look at it the other way round - that the properties of humans as legal personalities - able to own property - derive from the needs of the enterprise system. Historically most enterprises were sole proprietorships, and the rights of the sole proprietor shaped the concepts of capitalist law. But these sole proprietors were faces for units of production. It was the reproduction of these units of production by trade that necessitated that their representatives could own and dispose of property. These requirements informed our whole contemporary outlook on what are 'natural' or 'human' rights.

> This juridical relation, which thus expresses itself in a contract, whether such a contract he part of a developed legal system or not, is a relation between two wills, and is but a reflex of the real economic relation between the two. It is this economic relation that determines the subject matter comprised in each such juridical act. The persons exist for one another merely as representatives and therefore, as owners of, commodities. In the course of our investigation we shall find, in general, that the characters who appear on the economic stage are hut the personification of the economic relations that exist between them." (Marx, Capital I, L&W, 1970, p 84)

If one thinks back to previous societies, one realises that people have not always been legal personalities, bearers of inherent rights. To the framers of the US constitution certain rights might have appeared self evident, but they were self evident only as the rights of white property owners. Black slaves and white indentured labourers were equally ’self evidently’ not the bearers of such rights. Going back further, members of a hunter-gatherer tribe or subsistence farming family were not economic subjects in the modern sense. The constitution of people as economic subjects is associated with the onset of commodity production and the establishment of money. Today, capitalist enterprises are forced by their technology to be inter-dependent. But this is not the aboriginal condition. The aboriginal condition is virtual self sufficiency, the self sufficient household economy or the self sufficient village community. The existence of commodities and money can not originally have sprung from the demands of reproduction, even if they play that role today.

10.1.2 Lydia

Orthodox economic theory portrays money as arising out of barter, with one commodity being set aside as a means of exchange. Typically this commodity is said to be coin or silver, and standardised units of which then serve as the unit of account. Coinage is explained as a public spirited effort of the state to mass produce standardised weights of precious metal. This fable projects the monetary practices of Victorian Britain back onto the early history of money.

Prior to the issue of coinage, particular commodities acted as a unit of account in the payment of taxes and for extended systems of barter. Polanyi et al. (1957) describes this process in early Mesopotamia where trade transactions were entered into accounts in units of shekles, these being equivalently quantities of barley or silver. But it does not follow that payments were actually made in barley or silver. Rather the common unit of account allowed the mutual settling of debts in barter transactions. For international trade weights of metal seem to have been used in the settlement of payments.

There is a very imporatant difference between coins and quantities of silver denominated in some standard of weight. Coins ‘pass by tale’ that is to say that the value of a purse of coins is determined by counting them, multiplying by their denominations and summing the total. Precious metal, on the other hand, is valued by weight. These weights have to be verified on each transaction if one party is not to be defrauded. But the hypothesis that coins arose just as a means of providing standard weights of gold or silver does not fit well with the numismatic evidence.

Early coinage was far from being a standarised weight of gold or silver. The first coins were issued by Lydia in the 7th century BC. These are a standard weight the *stater*, roughly 220 grams, but, rather than being of pure gold, were made of electrum an alloy of gold, silver and copper: Bolin (1958). The addition of copper meant that they still looked golden, instead of the whitish look that a simple gold/silver alloy would have had. If they were supposed to be standard ingots of pure gold, then the Lydian state was defrauding its public. Also if their purpose was to facilitate commodity exchange in the markets of the Kingdom, why were they so heavy and valuable?

Why were they worth a month's subsistence[3]?

An alternative explanation is that they were used for the payment of taxes due to the Crown. With this we have a theory of the origins of money that ties it in to the development of class society and the state. This theory is currently advanced particularly by writers such as Wray (2004), Ingham (2004), Forstater (2003) , building on an tradition established by Knapp (1973), Innes (1913). According to this State or *Chartalist* theory, the state calls money into being by requiring that taxes be paid in money. At an earlier stage of development - for example early Egypt, the state levied taxes in the form of a duty to labour or a duty to provide agricultural produce. By specifying that taxes must be paid in coin, and at the same time offering coin as payment for labour performed for the Crown, the state enforced the currency of its coinage. The early Lydian stater were far to costly for day to day transactions but a month's subsistence would be a reasonable minimal unit of annual tax.

If the Crown imposes on its citizens a duty to pay tax in coin of the realm, then these citizens must either work directly for the state - building roads, acting as soldiers etc, or, they must produce commodities to sell to those who do serve in the army, build roads etc. In this conception, it is the coercive power of the state that accelerates the penetration commodity production into the social organism. Adam Smith called money the 'power to command the labour of others'. This power, in aboriginal form, belongs to the state. By issuing coins stamped with a royal emblem, the state delegates this command over labour to those who hold the coins. Possession of the coins indicates either that one has personally discharged one's obligations to the state as a soldier etc, or has indirectly discharged one's obligations by

---

[3] Carradice and Price (1988)

providing services to the soldiery. This ties money in to the production and appropriation of a surplus product. States were the first appropriators of a surplus. It is this state power to command a surplus product, that through the commuting of taxes in kind to taxes in money, forces an initially self-sufficient peasantry to produce for the market and eventually gives birth to civil, or bourgeois society.

In this process the use of precious metal is incidental. Wray emphasises that in Britain up to the 19th century the predominant form of state money was actually the tally stick not the gold coin:

> Originally, the money liability was always in terms of a unit of account as represented by a certain number of grains of wheat or barley. In fact, all the early money units were weight units for grain–the mina, the shekel, the lira, the pound. Once the state has imposed the tax liability, the taxed population has got to get hold of something the state will accept in payment of taxes. This can be anything the state wishes: It can be clay tablets, hazel-wood tallies, iron bars, or precious metal coins. This, in turn, means the state can buy whatever is offered for sale merely by issuing that thing it accepts in payment of taxes. If the state issues a hazel-wood tally, with a notch to indicate it is worth 20 pounds, then it will be worth 20 pounds in purchases made by the state so long as the state accepts that same hazel-wood stick in payment of taxes at a value of 20 pounds. And that stick will circulate as a medium of exchange at a value of 20 pounds even among those with no tax liability so long others need it to pay taxes. The matching of those with tallies but no taxes with those who have tax liabilities but no tallies is accomplished by bankers–who have always been the agents of government precisely to accomplish such matching.
>
> ...
>
> A tally was simply ”a stick of squared hazel-wood, notched in a certain manner to indicate the amount of the purchase or debt”, with the name of the debtor and the date of the transaction written on two opposite sides of the stick. (Innes (1913), p. 394) After notching, the stick was split down the middle in such a way that the notches were cut in half. The split was stopped about an inch from the base, with the longer piece (called the stock, from which our term ”capital stock” derives) retained by the creditor, with the ”stub” (a term still used as in ”ticket stub”) held by the debtor. The two pieces of the tally would be matched later (most significantly at the time of settlement)

> to verify the amount of the debt. Importantly, governments spent by raising a ”tallia divenda” on the exchequer, issuing tallies for payment for goods and services delivered to the court (after 1670, wooden tallies were supplemented by paper ”orders of the exchequer”, although tallies were still held in the English House of Commons until 1834).
>
> Wray (2004)

A monetary system sets up a binary relation [4] associating with each juridical subject an integer number. Each historical form of money represents a step in the development of the technologies of record which support this binary relation.

Coins maintain the relation by possession. The number associated with each individual is encoded in the coins they carry. Coin, however is an imperfect technology of record as it can only record positive numbers. You can not have £ – 50 in your pocket. Coins and paper money are both a token based method of record keeping. They are *abacic*, in that they correspond to abacus based systems of calculation[5]. A change of state in the system of record is achieved by the physical movement of tokens.

Tallys, double entry account books, decks of punched cards or computerised relational databases are more sophisticated monetary technologies able to associate with a legal person either a credit state or a debit state. Tallys are a specialised token system of record. The other technologies are *algorithmic* in the sense described in section **??**. A change of state is achieved by the writing down or recording of symbols.

A key concern of all monetary technologies is their integrity of record. They must provide some protection against falsification. It is in this light that the use of precious metal for coins should be seen. States have always enacted severe penalties for the fraudulent issue of coin. But penalties would be ineffective if the issue of fraudulent coin is made too easy. Beyond legal prohibitions on forgery, state coin had two distinct protection mechanisms.

(1) The coin is made by stamping from a master, one of the basic copying technologies described in Section 3.3.3. Unless one has access to the master it is difficult to make accurate copies of the coin. Reasonably

---

[4] We use the term relation in the strict logical sense of set of tuples defining the extent of logical predicate. The predicate in the case of money has the form x `is credited with` y: $x \in$Juridical Subjects, $y \in$ integers.

[5] See section **??**. This shows the basic historical materialist postulate of the controling influence of the forces of production : machinery for calculation in this case.

good copies may however pass without notice. To do this one has to replicate the master, which can in principle be done by taking an impression of the coin, using this to make a mould and from that cast a new die. Until the invention of iron casting, this process was technically infeasible, since dies made from softer castable metals like bronze would not have the toughness required to stamp out coin. Note that there are 3 copying stages between the coin used as a model and the new forged coins. Errors in copying accumulate exponentially so it is very difficult to get forgeries of acceptable quality.

The remaining forgery techniques were to hand carve a new die, or to use an existing coin to make negative moulds from which coin could be cast rather than stamped. These are relatively expensive processes and would not be worth while for the production of low denomination coinage. For high denomination coinage they would be feasible.

(2) Whilst low denomination coins were made from copper or copper alloys, and protected against forgery by the method above, high denomination coins required additional protection. This could be done by forging them from expensive materials like gold and silver. Provided that the nominal value of the coins was not hugely in excessive of the value of the metal they contained, this, in conjunction with the inherent difficulties of accurate copying, reduced the profits to be made from forgery.

The use of gold or silver is not essential to money tokens, as is shown by their abandonment in favour of the use of paper money printed using sophisticated techniques that make it difficult to copy. The use of bullion was a low-tech anti-forgery expedient.

As the state commutes taxes in kind to money taxes it moves from the direct appropriation of the surplus product to its indirect appropriation, mediated by the money symbol. In levying a money tax, the state symbolically asserts its right to a portion of of society’s labour. When it spends the tax money purchasing goods and labour, it performs a real appropriation of a surplus product. Civil society then acts as an intermediary transferring labour from those who paid the tax, to those who provide the actual services to the state.

The state, of course, predates societies in which commodity production is general and it has a primordial power to appropriate part of society’s labour time. In the early empires of Mesopotamia and Egypt, or the later Inca empire this appropriation was performed directly. All peasants had a

Digression 10.1 British monetary policy in Africa.

The essential interdependence of state and money is particularly clear in the history of empires. On conquering Africa, the Europeans face the problem that

> if the subsistence base was capable of supporting the population entirely, colonial subjects would not be compelled to offer their labor-power for sale. Colonial governments thus required alternative means for compelling the population to work for wages. The historical record is clear that one very important method for accomplishing this was to impose a tax and require that the tax obligation be settled in colonial currency. This method had the benefit of not only forcing people to work for wages, but also of creating a value for the colonial currency and monetizing the colony. In addition, this method could be used to force the population to produce cash crops for sale. What the population had to do to obtain the currency was entirely at the discretion of the colonial government, since it was the sole source of the colonial currency. (Forstater 2003)

duty to provide either time or crops to the state. Some of the crops would be consumed by priests or state officials, another portion would be stockpiled against drought and redistributed to the working population in times of scarcity. This form of economy was termed *redistributive* by Polanyi. Such a system requires the development of information technology - systems of writing down and recording numbers. Thus the Mesopotamian civilisations developed cuneiform numbers and later, developed writing. The Incas developed quipu, a numerical notation based on knotted strings.

Such systems of record had to :

- keep track of physical stocks of crops held by the state or temples;
- keep track of the deliveries made by individuals and groups subject to tribute deliveries;
- track the tribute obligations of such groups

These require a the development of a recording technology, standardised systems of measurement and a reliable arithmetic technology. The state had to be able to associate numbers with tax-payers and types of products. It had to be able to measure the grain delivered. It had to be able to add up tribute delivered by groups to know what total it had in stock - hence a reliable technique for adding large numbers was needed. In order to determine if

If a man has hired a boatman he shall pay him 6 **gur** grain a year.
If a man has hired an ox he shall give its owner 4 **gur** grain for the hire of the rear ox and 3 **gur** for the hire of a front ox.
If a man has hired a farmer he shall give him 8 **gur** grain a year. If a man has hired an ox-herd he shall give him 6 **gur** grain a year.

Figure 10.1: Code of Hammurabi, cited in Postgate 1992.

a group had met their tribute obligations, a technique of subtraction was required, taking away their deliveries from their obligations.

The Sumerian civilisations developed a sophisticated system of written numerals, using a place notation similar to that we use today. The key difference was the number base. Our place notation, deriving originally from India, uses base 10, the Sumerian's used base 60. Place notation is concise and allows large numbers to be readily manipulated. It was also a written notation, lending itself to the recording of tables of tax deliveries. Without this technology for recording and processing information the social complexity of the early empires would not have been feasible. In all but the simplest social systems, social relations are embodied in information technology. Without a technique for recording debts, the social relation of creditor/debtor can not persist. Without a means of measuring land and recording ownership, the relation of landlord to tenant cannot exist.

Different subjects of the empire would deliver different crops depending on their circumstances. Some might deliver barley, some dates, some dried fish, or a mixture of such products might be delivered. It is thus necessary to determine if a farmer delivering a basket of dates and three *gur* of barley has met his tax obligations. The solution was to define the tax obligation in terms of barley and for the state to then define how much fish, dates etc would be required to meet this obligation in terms of barley. The standard volumetric unit of barley, the *gur*, about 300litres, then became the unit in which deliveries of other products were measured. The gur of barley had an equivalent in silver the *shekel*, defined as silver to the weight of 240 grains of barley. It appears that this then became the basis for a purely accounting based monetary system. The shekel/gur was never issued as a coin, it existed only as entries in accounting records on clay tablets. This notional quantity of barley then acted as a generalised way of measuring values and obligations. From regulating obligations to the state, it moved to being the

1 **gur** barley	for 1 shekel silver
3 litres best oil	for 1 shekel silver
1.2 litres vegetable oil	for 1 shekel silver
1.5 litres pig fat	for 1 shekel silver
40 litres of bitumen	for 1 shekel silver
6 minas wool	for 1 shekel silver
2 **gur** salt	for 1 shekel silver

Figure 10.2: Opening section of Esnunna Law Code, cited in Postgate 1992

unit in which credit relations between private individuals were expressed. Such a system of credit based accounting was only possible thanks to there being a literate and numerate class of scribes. The place based number system and algorithmic calculation underlay it. If you are to become proficient in a place based number system you need to spend childhood years learning by rote your tables. You have to learn to memorise the addition, subtraction and multiplication tables. This is a hard enough task using base 10. With a base 60 number system it would probably have been more difficult. A naive estimate indicates that the size of the tables to be learnt is 36 times as great as for our school children. This almost certainly overestimates the task, however since the Babylonian number system is better seen as an alternating base 10, base 6 system. This gives rise to patterns that can be more easily learnt than would be the case in a pure base 60 system. Notwithstanding, to operate an accounting based monetary system required an expensively educated class that was lacking in the petty kingdoms and city states who first introduced coinage. Coins allowed monetary relations to operate in societies which lacked this class of numerate scribes.

10.1.3 Money space

Any monetary system must maintain a binary relation between juridical subjects and sums of money. We can represent such a binary relation as a table like:

Subject	Shekels
Alande	7
Tunde	12
Eve	200
Rachel	18
Ogun	23
TOTAL	260

There needs to be some form of persistent store that can hold the state of this relation through time - clay tablets served well, as do modern computer disks, but coins also work. Coins are self registering and self accumulating. The physical presence of coins in a purse records a number. The possession of the purse associates it with a juridical subject. The state of the monetary system is then encoded in the totality of such records - the totality of current account tablets, the totality of current account database relations or the totality of purses in peoples pockets.

Next you need a mechanism by which the system of records can be updated when a *transaction* occurs. Transactions are how the state of a monetary system evolves through time. They are atomic, indivisible events. A basic transaction must update two peoples records, leaving the totals unaffected. A payment of 13 from Eve to Alande yields a new state of the system:

Subject	Shekels
Alande	20
Tunde	12
Eve	183
Rachel	18
Ogun	23
TOTAL	260

There are several ways such a transaction can be performed:

(1) A system based on coins or other portable tokens like banknotes achieves the transaction by hand-over. The physical conservation of the coins ensures the atomic, conservative character of the transaction.

(2) If the relation is stored on some erasable and re-writable medium - like the old ”slate” used by shopkeepers to keep track of credit to customers, or a modern magnetic disk, then one simply

(a) adds 13 to the total listed for Alande, and then

(b) rubs out Alande's total, and then

(c) writes down the new total - 20.

(d) subtracts 13 from Eve's total, and then

(e) rubs out Eve's total, and then

(f) writes down Eve's reduced total - 183.

Although this sounds simple, when computer disks are used, a great deal of trouble has to be gone through to ensure the atomicity of such transactions. One has to deal with the possibility that there may be a computer failure half way through the process. That might leave Alande credited with an extra 7 units, whilst Eve had nothing debited from her account.

One solution to this is what is to use *before-looks*. The computer writes the previous values of Alande and Eves accounts to a special disk file - the before-look file, prior to altering either of them. After they have both been updated, the before-look file is deleted. If a failure occurs midway through, then the before-look file will survive intact. When the database starts up again it uses the before-look to roll back the transaction to its start. This brings the records back into a consistent state.

(3) In a system based on permanent records - paper, clay tablets, one has to add a new record detailing the transaction:

Payer	Payee	amount
Eve	Alande	13

Each transaction requires the storage of a new record. The balance associated with each individual now has to be obtained by adding up all of the extant transaction records.

Modern banking databases may supplement the mechanisms of before-look files with what is termed a transaction log. This simply lists all of the transactions in a file as they arrive, much as would be done in pen and ink accounting. At a later stage, perhaps at the end of the banking day, these transaction logs are run through the computers to perform atomic updates on the master relational database.

A consequence of moving from an account based monetary system to coinage, was decentralisation. Accounting requires the records of transactions to be concentrated in a few accounting centers: palaces, bank clearing houses etc. Coins can be dispersed around the population at large. They were a flexible, low-tech, decentralised monetary technology that allowed monetary relations and commodity economy to spread rapidly. According to Ingham, the Macedonian and later Roman empires were important vehicles for this spread.

A disadvantage of coins however, is that they can only record positive numbers. They can not record the situation of having a negative amount of money - a situation of debt. Debt required the existence of supplementary documents, recording the existence of debt. Some such debts arose from private transactions. Other debts were due to the state - the obligation to pay taxes. A person's total position with respect to the state is now encoded in two distinct forms - the coins that they hold, and the tax obligation written down in the tax collectors records. We can thus extend our previous type of relations with an extra table, that of tax obligation, and an extra row to represent the state. We imagine our example monetary system to be in Nigeria in 1905 after its conquest by Lugard and incorporation into the British Empire.

Step 0 Let us imagine that

Agent	Coin	Tax Obligation
State	0	0
Alande	0	0
Tunde	0	0
Femi	0	0
TOTAL	0	0

represents the state of the system at the start of the year.

Step 1 The first step is for the state to mint coin.

Agent	Coin	Tax Obligation
State	9	0
Alande	0	0
Tunde	0	0
Femi	0	0
TOTAL	9	0

Steps 2 and 3 The state then employs Femi in the Royal West African Frontier Force for some months and pays him 7 coins (step 2). It then announces that everyone will have to pay a poll tax of 2 coins (step 3). So the state of the system can now be described by the relation:

Agent	Coin	Tax Obligation
State	2	6
Alande	0	-2
Tunde	0	-2
Femi	7	-2
TOTAL	9	0

Alande and Tunde hear that they must pay a poll tax in the new coin. They are also told that if they fail to pay the district commissioner will force them to stand in the open staring at the burning sun all day, and then have them publicly flogged. They are understandably keen to get hold of coins.

Step 4 They offer to sell Femi food so that they can get hold of these coins. Femi buys his food and we have the situation:

Agent	Coin	Tax Obligation
State	2	6
Alande	3	-2
Tunde	2	-2
Femi	2	-2
TOTAL	9	0

Step 5 Finally the day of reckoning[6] arrives. Taxes are due. Coin is accepted by the collectors in cancellation of tax debts due on that day. This gives us a situation described by:

Agent	Coin	Tax Obligation
State	8	0
Alande	1	0
Tunde	0	0
Femi	0	0
TOTAL	9	0

Femi has lost all his money and is available for hire again, Tunde has sold part of his crop and covered his debt. Alande has sold somewhat more of her crop, but is left with a coin. She is now in a position to continue operating as a trader. Looking at things from the standpoint of the new colonial monetary economy she is richer than she started, but in reality she has given up food which is really useful and is left with a copper disc of limited practical use. In real terms she has been impoverished. Set against this material impoverishment there is a social advance. Holding the King's coin, she partakes indirectly in the power and authority of the King. With coin she can command the labour of her fellows. She buys kola nuts for her stall. By itself this purchase looks a very emblem of reciprocity, free and voluntary exchange. But behind it, driving it, is coercion and fear of the tax collector.

We can now identify the basic circuit of money

`King → Lackey → Subject → Subject ... Subject →King`

We can also identify the basic primitive operations that describe or change the state of a monetary system. These are the *signature*[7] of money:

(1) *Holding(agent x → money)* This function specifies the holding of money by agent *x*. The social system will have various ways of encoding this holding.

---

[6] The notion of the day of reckoning, reflects the apotheosis of the state. God as the supreme tax collector, an imaginary projection of the God-Kings of the Roman, Hellenistic and earlier empires.

[7] We borrow the term signature from its usage in type theory where it describes the collection of basic operations supported by a type: See for example Goguen and Meseguer (1982).

(2) *Pay(agent x, agent y, money m)* a payment by x to y of an amount of money m. This operation follows the constraint that the total holdings of the two agents x, y does not alter, and that

$Holding(x)_{pre} = Holding(x)_{post} + m$, where the subscripts indicate the situation prior to and after the operation. This is a conservative operation.

(3) *Mint(money m)* This operation increases the holding of money by the state by m, thus

$Holding(state)_{pre} = Holding(state)_{post} - m$ This is a non-conservative operation.

Should taxation be considered a distinct operation in this signature? No, because from the standpoint of the state of the money system, taxation is just another payment. Its special enforced character is invisible in the space of money. It is only when we look at a more comprehensive space, commodity/money space that tax payments stand out as special.

Seigneurage

In a monetary economy the state has two was of gaining access to real labour resources.

a It can levy a tax in money and spend the money buying labour or commodities.

b It can simply mint and spend the money. This process is termed seigneurage.

Taxation and seigneurage are mutually inter-dependent. Unless there is an initial minting of money, no tax in money can be levied. On the other hand if no taxes are levied, then the money will be valueless and the state will be unable to appropriate real resources with its coin.

In a natural economy the appropriation of resources by the state is direct and constrained by its political ability to coerce property owners into handing over goods, and also to coerce subjects into performing labour services. With the invention of money, the appropriation of a surplus splits into two - a symbolic appropriation of coins as tax goes alongside a real appropriation, by purchase, of labour time and commodities. The real appropriation appears as something equitable and voluntary. The coercive aspect of the process occurs entirely in the realm of symbols - rendering unto Caesar that which is Caesar's.

Coercion remains bounded by political ability. Taxation meets resistance whether it occurs in money or kind. But because there is a split between the real and symbolic domains, seigneurage can act as a wedge to force them apart. A state can, within limits, appropriate more labour than it can raise symbolically as tax. Taxation is a recurrent process. It provides a stream of symbolic labour to be spent on real labour. Seigneurage is a one off process needed to start the tax process going. In the year the coins are minted, the Crown can purchase more than it taxes. This acts as a constant temptation for states whose tax raising power is weak, for minting coin is politically easier than raising taxes.

The issue of coin has to be a continuous process anyway.

- There is always a certain loss of coin due to accident or ware and tear. The subject's accidental loss is the Crown's gain. If a coin falls in a river, a record that services have been performed for the Crown goes with it. If taxes are to be met, the Crown must issue a new coin, and the subjects must perform new services to get it. This means that a certain level of seigneurage is built into the system. This seigneurage derives from the *lost information* caused by the imperfections of coin as a technology of record.
- Some additional minting is required to keep pace with the growth in the value of commodity circulation. As more people are drawn into commodity production - because of population growth, the expansion of the state, or because previously natural economic processes become commodified - then more coin is required to sustain this trade.
- Hoarding or saving withdraws coin from circulation. Of course many hoards are eventually lost either for ever, or to be found by archaeologists centuries later. But leaving those aside for the moment, the effect of hoarding is very similar to that of loss from the standpoint of the state. If a hoarder puts away 100 coins a year into a hoard, then, unless these are compensated for by other hoarders dissipating their hoards, the Crown can issue an extra 100 coins per year. Any net hoarding by the population allows a corresponding rise in the annual issue of coin.

  The consequence of a higher issue, is more seigneurage - real appropriation of labour and services by the Crown. But the money form hides this, both from holders of money, and from some economists who should know better. Open an elementary economics textbook

> and you read that money serves as a store of value. Hoarders believed this, but it is fundamentally an illusion. A miser with 1000 pennies under the floor, had not stored up value: he buried the ghost[8] of value departed. The King issued pennies in return for real value - work. Like as not, they were born as soldier's pay. Then the work of soldiering, like winter's snow, vanished leaving no material residue. Aside from the paltry scrap value of Royal cannon, no vendible commodity survived,
>
> What the miser stored was information, a number that assigned to him a tiny fraction of the power royal. Should he spend his hoard he would command the labour of others. But should all hoarders try to do this at once, they found themselves competing with the normal purchasers of labour and commodities. Prices would go up. The social power represented by each coin would fall.
>
> In time of famine hoards were spent. They helped ensure the survival of the hoarders, but only by grabbing them a larger share of a diminished crop. They redistribute starvation, but on a social scale do not represent an accumulation of value. A social provision against famine could only occur if there were a real accumulation of value in the form of corn in granaries.

If issue of coin goes beyond these limits, then the coin will tend to be devalued. A persistent excess of emissions over the various forms of withdrawal can lead to a devaluation of the coin.

10.1.4 Commodity-money space

We now extend our representation to include both commodities and money. Again we can use a table to summarise the position of our system in the new state space. Our new table will have a column for money followed by a column for each type of commodity. Table 10.1 shows the holdings of coin and commodities by the agents in our little society. Lets assume that the first thing to happen is that Femi buys 3 kola for 3 coins from Alande. This can be decomposed into two sub-operations

Pay(Femi, Alande, 3) which moves the money and

---

[8] See the discussion of ghosts in digression 10.2.

Table 10.1: Table of money and commodity holdings by agents

Agent	Coin	Cassava	Kola
State	2	0	0
Alande	0	6	6
Tunde	0	2	5
Femi	7	0	0

Transfer(Alande, Femi, Kola, 6) which moves the goods from Alande to Femi. Transfers conserve the commodity being transfered. After both of these operations we have:

Agent	Coin	Cassava	Kola
State	2	0	0
Alande	3	6	0
Tunde	0	2	5
Femi	4	0	6

In commodity exchanges these operations occur in matching pairs. In taxation the payment is unilateral with no corresponding transfer of goods. Note that in our example above we have abstracted from the prices of commodities. We will look at this in more detail in section 10.2.

10.2 EXCHANGE IS VALUE CONSERVING

We have asserted that the operations of payment and commodity transfer are conservative, in the sense that the amount of money and commodities is unchanged after them. We will now look at what it means to say that commodity exchange, that is to say linked pairs of payment and commodity transfer are value conserving as well as conserving commodities and money.

In science it is often worth questioning, problematising, the obvious.

It seems self evident that if Chantelle has a car, three chairs and a table we can add these up to obtain her total worth.

Chantelle	Car	£900
	Table	£50
	Chairs 3 at £12	£36
	total	£986

If Briony meanwhile owns a table, 4 chairs and a washing machine we can similarly add these up:

Briony	Table	£50
	Chairs 4 at £12	£48
	Wahing machine	£120
	total	£218

So Chantelle is £768 richer than Briony. All this is perfectly obvious to anyone in a comercial society like our own. But why do these arithmetic operations work, and what are we doing when we thus compare two peoples wealth?

Looked at in an abstract mathematical fashion, Chantelle and Briony have what economists term vectors of assets. For Chantelle this vector is $[1,1,3,0]$ denoting 1 car, 1 table,3 chairs, 0 washing machines. For Briony it is $[0,1,4,1]$. These vectors define positions in wealth space. When we compare their relative wealth we are deciding what 'distance' separates them in terms of wealth.

This problem of measuring 'distance' comes about in many domains; for example:

(1) We might want to measure the distance that you would have to walk between two street corners in Manhattan.

(2) We might want to know the distance 'as the crow flies' between two hilltops given their map coordinates.

(3) Given 3 variants of a conserved gene from chimpanzees, gorillas and humans we might wish to determine which two were closest.

Each of these uses distance in a different sense and for each there are appropriate mathematical techniques to work out the distance. An important property of all distances is that they are positive numbers, so the procedures used to measure distance must ensure we do not get negative results. A mathematical method of measuring a distance is refered to as a metric.

Manhattan distances On the regular street grid of New York, the walking distance between two street corners is just the sum of the distances along the two axes with which the streets are aligned. Whether you chose a simple route, or try to zig-zag, you end up going the same distance.

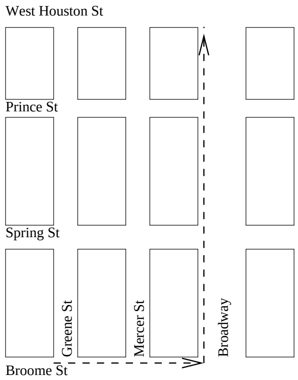


Figure 10.3: A walk in Manhattan.

If the starting and finishing points are $[a,b],[x,y]$ the walking distance is $|a-x|+|b-y|$. The use of absolute values ensures that the result is positive because $|x-a|=|a-x|$.

As the bird flies, or Euclid's distance Over street grids offering us no shortcuts, birds fly freely. Where we have to walk 3 miles East and 4 North, pigeons fly only 5 diagonal miles.

We calculate diagonal distance as the square root of the sum of the squared distances along the axes: $5=\sqrt{3^2+4^2}=\sqrt{9+16}$. For our points $[a,b],[x,y]$ the formula is $\sqrt{(a-x)^2+(b-y)^2}$, a positive result ensured, this time, by an initial squaring.

Hamming distance DNA can be computationally represented as strings drawn from the alphabet A, T, C, G. Here the letters correspond to the 4 bases that encode the information in a DNA molecule. Proteins can be represetned as similar strings drawn from a larger alphabet representing the amino acids in their sequence. A simple measure of the distance betweent two DNA or amino acid sequences is to simply count the places along the sequence where the letters disagree. For example the following amino acid sequences differ in 26 places.

```
MKPGRLASIALAIIFLPMAVPAHAATITITMTNLVISPTEVSAKVGDTI
MKAGAKIRLSWLAALALMAAPAAAATIEVTIDKLVPSPATVEAKVGTDI
```

```
  * *************  *  *    ** ***  *  ** *
```

The number of places where the sequence differs gives us a measure called the Hamming distance. Hamming invented it when trying to rank the seriousness of errors in digital codes.

Hamming distances are a useful measure of distance between two related DNA sequences because they measure the number of point mutations required to change one code into another.

Commodity value distance Suppose we have the two vectors representing Chantelle and Briony's assets: $\mathbf{c} = [1,1,3,0]$, $\mathbf{b} = [0,1,4,1]$. What mathematical formula can we use to obtain the distance d between their net worths?

A suitable formula is:

$$d = |\sum_i \mathbf{p}_i(\mathbf{c}_i - \mathbf{b}_i)| \tag{10.1}$$

where p is the price vector; $[900,50,12,120]$ in the previous example.

Surprisingly, it turns out that this metric is similar to one that occurs in physics when dealing with the conservation of energy. Suppose that instead of **b, c** being vectors of commodities, the vectors represent the height and kinetic energy of two flying balls. Then 10.1 would give us the difference between their energies.

Is this significant?

Perhaps it is. Marx said that in Capital he was trying to elucidate the laws of motion of capitalism. He was implicitly comparing the study of capitalism to physics. He devoted considerable space to analysing the logic of commodity exchange. In this context the fact that net worth has the same metric as the conservation of energy may well be relevant.

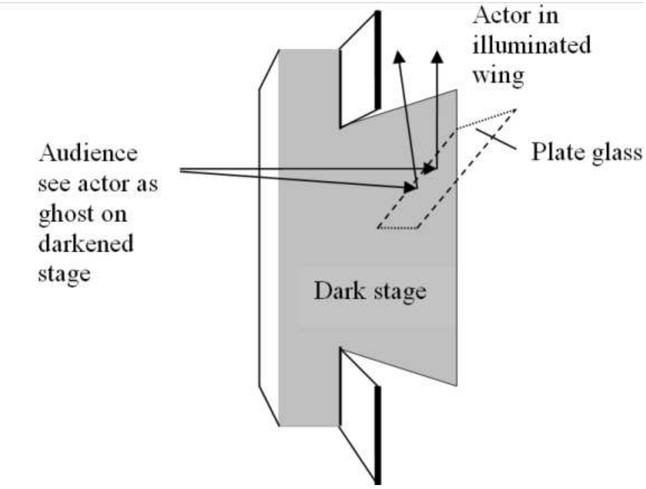
The law of energy conservation constrains the paths that flying balls follow. If we threw a ball on the moon, where there is no atmospheric resistance, then at each instant the ball will have combination of height and velocity that causes its total energy to be unchanged (Figure 10.4).

If value were just a matter of providing an ordering or ranking of combinations of goods, then a Euclidean, or indeed any other, metric would pass muster. It is some additional property of the system of commodity production that imposes this specific metric characteristic of a system governed

Digression 10.2 Money and the illusion of Pepper's ghost

In Victorian times there was a popular stage illusion known as Pepper's ghost. On a gloomy stage set a ghostly figure would appear floating in front of the fittings; its ethereal yet living qualities revealed by the fact that it was at once animated and transparent. The ghost was in fact the reflected image of an actor in the wings, reflected off a sheet of plate glass placed at 45° to the audience.

The ghostly and illusory properties of money and credit come because our views of them are projections of a partially hidden stage. One on which every entity has its mirror image, to every credit a hidden debit, to every visible coin a hidden tax. It is perhaps fitting that an age whose working lives were ruled by money and credit like none before, should have developed an obsession with ghosts, mediums and spirits.



How the Victorian stage illusion of a ghost was performed.

by a conservation law. This fits in rather nicely with the labour theory of value, where social labour would be the embodied substance conserved during exchange relations, which in turn provides us with some justification for casting the law of value in the form of a classical conservation law.

So far, however, this is merely a formal argument: the form of the phenomena is *consistent* with a conservation relation. To justify our formulation fully we depend on the arguments presented in chapter 5.

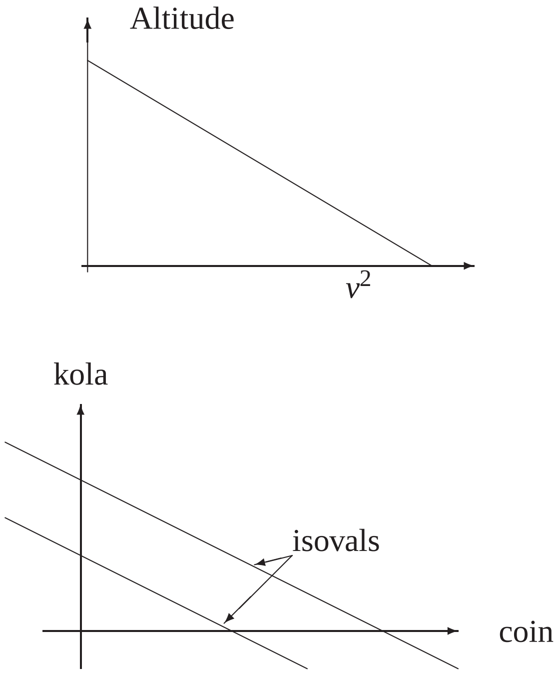


Figure 10.4: Points of equal net worth (isovals) in the space of commodities and money have the same form as points passed through in the phase space of altitude and velocity squared by a falling body.

Spatial metrics are so much part of our mode of thought that to imagine a different metric is conceptually difficult. Most of us have difficulty imagining the curved space–time described by relativity theory, Euclidean metrics being so ingrained in our minds. Conversely, when looking at commodities, a non-Euclidean metric is so ingrained that we have difficulty imagining a Euclidean commodity space.

But it is worth the effort of trying to imagine a Euclidean commodity space, what we referred to earlier as commodity vector space. By bringing to light the implicit contradictions of this idea, we get a better idea of the underlying reasons why value takes the particular form that it does.

Is a Euclidean metric for commodity space internally consistent? In commodity bundle space of order 2 the Euclidean isovals take the form of circles centred on the origin. In higher-order spaces, they take the form of spheres or hyper-spheres. (We assume in all cases that some linear scaling of the axes is permitted to convert them into a common set of units.) Let us suppose that the economic meaning of these isovals is that given any pair of points $\mathbf{p}$, $\mathbf{q}$ on an isoval, the bundle of commodities represented by $\mathbf{p}$ will be exchangeable as an equivalent with the bundle represented by $\mathbf{q}$.

If the state of an economic agent is described by her position in this commodity bundle space, then the set of permissible moves that can be made via equivalent exchanges is characterised by unitary operators on commodity vector space. The set of equivalent exchanges of $\mathbf{p}$ is $\{|\mathbf{p}|\,\mathbf{u}$ such that $|\mathbf{u}| = 1\}$, i.e. the radius-preserving rotations of $\mathbf{p}$. Mathematically, this is certainly a consistent system.[9]

But economically, such a system would break down. It says that I can exchange one, appropriately defined, unit of kola for one unit of coin, or for any equivalent combination such as ($\frac{1}{\sqrt{2}}$ coin, $\frac{1}{\sqrt{2}}$ kola). But then what is to stop me carrying out the following procedure?

(1) Exchange my initial 1 unit of kola for $\frac{1}{\sqrt{2}}$ coin plus $\frac{1}{\sqrt{2}}$ kola.

(2) Now sell my $\frac{1}{\sqrt{2}}$ coin for kola, giving me $\frac{1}{\sqrt{2}}$kola.

(3) Add my two bundles of kola together, to give a total of $\frac{2}{\sqrt{2}} = \sqrt{2}$ of kola in total.

I end up with more kola than I had at the start, so this cannot be a set of equivalent exchanges. The second step is illegal within the context of the Euclidean metric, since it involves operating upon one of the coordinates independently. But in the real world, commodities are physically separable, allowing one component of a commodity bundle to be exchanged without reference to others. It is this physical separability of the commodities that makes the observed metric the only consistent one.

The existence of a commodity-producing society, in which the individual components of the wealth held by economic agents can be independently traded, selects out of the possible value metrics one consistent with the law of value. In a society in which commodity bundles could not be separated into distinct components, and exchange obeyed a Euclidean metric, the labour theory of value could not hold—but that is not the world we live in.

---

[9] A very similar model is used in one of the standard formulations of quantum theory to describe possible state transformations (von Neumann, 1955).

10.3 LOGICAL PROPERTIES OF FINANCIAL TRANSACTIONS

10.3.1 Modelling Debt

We have already said that accounts based monetary systems are capable of recording both positive and negative amounts of money. This is required to represent debt.

Suppose that starting from holdings

Agent	money	kola	total
0	1d	0d	1d
1	0d	4d	4d
totals	1d	4d	5d

agent zero buys 2d of kola[10] from agent one. Since agent zero only has 1d in money to pay for it, the transaction leaves the following holdings:

Agent	money	kola	total
0	-1d	2d	1d
1	2d	2d	4d
totals	1d	4d	5d

We see that

- the totals for both money and kola are conserved,
- the total assets of each person do not change.

Sales on credit are still a conservative operation However this is only true if we abstract from coinage. If money were just coin then we have a contradiction. There was initially only one penny in circulation, but after the sale agent one has two pennies. Where has the extra penny come from?

The state has not issued a coin, so who created it?

The new penny is balanced by a new negative penny held by agent 0, the positive and negative new pennies constitute a debt between agent 0 and agent 1. This debt can not be supported by coinage, since coins can only represent positive numbers, so this implies some ancillary system of record

---

[10] 2d means 2 pence worth. The small denomination coin circulating in colonial Nigeria were pennies. The suffix ”d” used to denote pennies in the British imperial monetary systems was a relic from the Roman imperial monetary system standing for Denarius, the basic silver coin of Rome. The Denarius transmuted into the Penny in the early middle ages.

keeping to encode debt. We have already looked at an analogous situation with the tables recording tax obligations and coin holding in section 10.1.3. Here we represented the relationship between the state and individuals as a table with a pair of columns, one showing the coin that people held and the other their tax debts. These tax debts are assumed to have been recorded in some government tax ledgers.

One column of our tables showed physical objects - coins in this case, the other shows numbers that are recorded on paper. The numbers in the tax ledgers refer to coins, they are denominated in money. They describe the number of coins that subjects own the Crown in tax. At one level it seems that by being denominated in coin the ledgers refer to physical assets that must be handed over to the Crown. But we have seen that coins only get their value by virtue of the tax debts. That is because the circulation of the currency was imposed by a fundamentally coercive - non conservative operation, an enforced obligation to pay. The creation of private debt (disregarding interest for now) is a fundamentally equitable or conservative operation.

Non conservative operations associated with debts are:

a The formation of tax debts to the state.

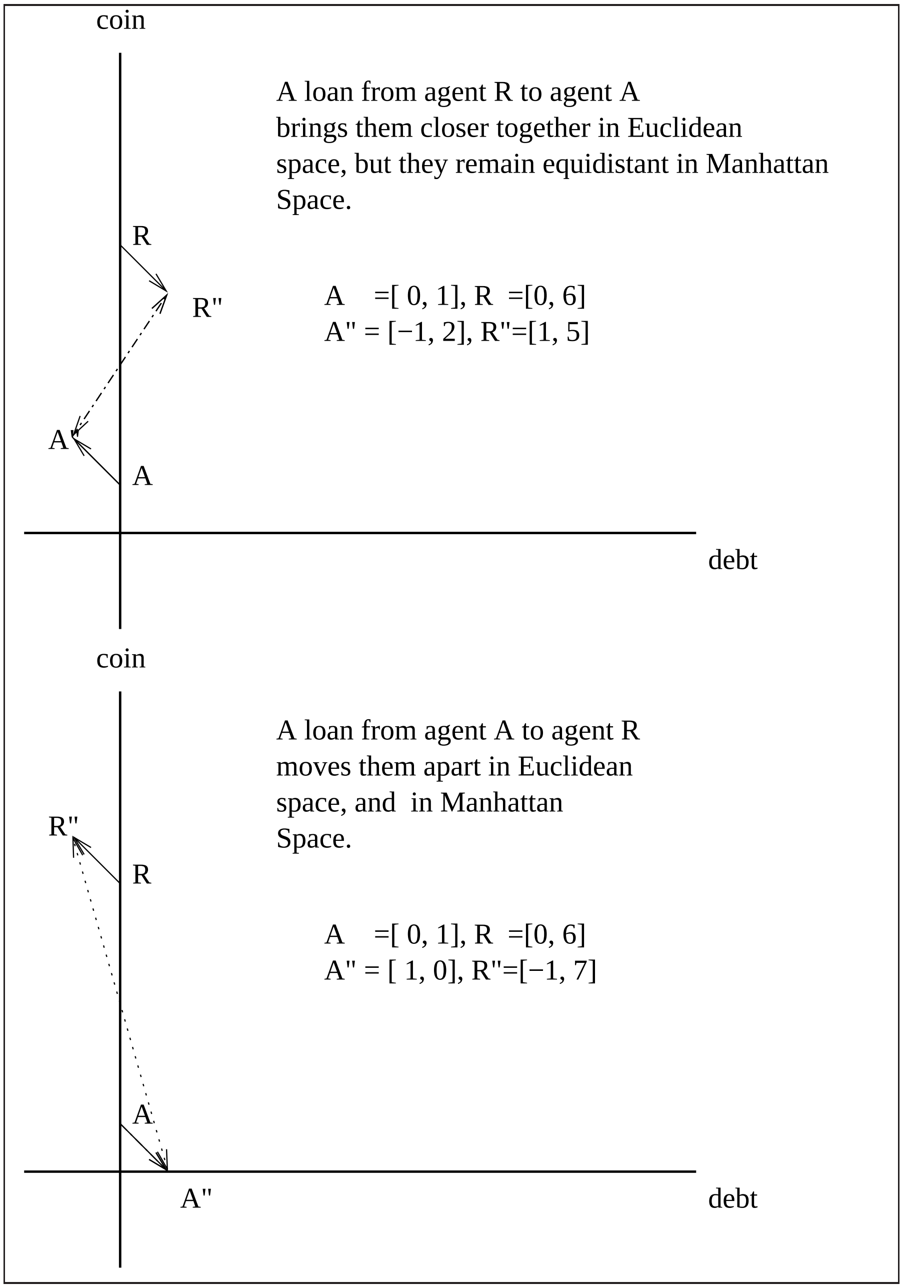
b The levying of interest on existing debts which increases the indebtedness of the original debtor.

Because of these cases it turns out to matter a lot that debt operations do not follow the same symmetry and conservation laws as commodity exchanges. Debt formation has its own symmetries, but these are not inherently conservative ones.

10.3.2 Relative movements caused by loans

Consider figure 10.5. This shows on a graph what happens when two agents Ajit (A), and Rakesh (R) engage in mutual loans. The vertical axis measures coins held and the horizontal axis measures their mutual indebtedness. In each half of the diagram Ajit starts out with 1 coin and Rakesh starts with 6 coins. At the beginning, as they have no mutual debts, both agents lie on the vertical axis.

Consider the situation where Rakesh lends one coin to Ajit, shown in the upper half of the figure. Both Ajit and Rakesh move along their isovals in opposite directions as the loan occurs. Since they remain on their isovals,



A loan from agent R to agent A brings them closer together in Euclidean space, but they remain equidistant in Manhattan Space.

A =[0, 1], R =[0, 6]
A" = [−1, 2], R"=[1, 5]

A loan from agent A to agent R moves them apart in Euclidean space, and in Manhattan Space.

A =[0, 1], R =[0, 6]
A" = [1, 0], R"=[−1, 7]

Figure 10.5: Effect of loans in moving the relative position of agents

it is clear that there is no alteration in each of their net worths. At the beginning Rakesh has a net worth of 5 coins richer than Ajit, and is still 5 coins richer after the loan has been made: on the assumption that Ajit will not default.

Although their net worths do not change, it looks as if this operation brings the two agents closer together. Indeed, in Euclidean space it does. They are initially 5 coin apart, but after the loan applying the Euclidean distance formula

$$\delta(A\text{''}, R\text{''}) = \sqrt{(1-(-1))^2 + (5-2)^2} = \sqrt{13} \approx 3.6055$$

they are closer. If on the other hand we look at the differences in Ajit and Rakesh's net worth, they are still 5 apart, and more surprisingly their Manhattan distance stays the same:

$$\delta_m(R\text{''}, A\text{''}) = |1-(-1)| + |5-2| = 5$$

Now look at the lower part of figure 10.5 which depicts the situation of Ajit lending 1 coin to Rakesh. Again the two agents move along their budget lines in opposite directions, but in this case the loan seems to move them further apart. This is obviously the case in Euclidean space, but the interesting thing is that they are now further apart in Manhattan space.

$$\delta_m(R\text{''}, A\text{''}) = |-1-1| + |7-0| = 9$$

The rule for small loans made by moneylenders who start out with much more money than the person they are lending to is that the loan does not alter the Manhattan separation of the agents. This means that the two agents are the same distance apart after the loan as before it.

If we consider the obverse relationship, where an agent with a small amount of cash lends some to an agent with much more cash, we find that the agents move apart in Manhattan space. The situation of an agent with a small amount of money making a loan to an agent that is much richer might seem improbable at first sight, but it is just what happens when an individual or a company makes a deposit with their banker. But does this movement apart in Manhattan space have any practical significance?

Surely in the real world we deal with differences net worth, not Manhattan distances between people. Since differences in net worth are unchanged, why worry about Manhattan distances?

Because the Manhattan distance points to something of huge social significance. Why should a poor person lend to a rich person in the first case?

Surely this would be a rare exception?

So it was in pre capitalist society, but with capitalism it becomes the rule. Name the rich person “Banker”, and it becomes clear. Whenever we deposit money in the bank we lend to someone much richer than us. The growing Manhattan distance between us and the bank this produces, measures the increasing social power of the banker.

Consider the theoretical entity that is termed the ’money supply’. This has various definitions, but is typically taken to be the sum of notes and coin in circulation along with the total of all current account deposits with the banks. At one level this seems obvious and unexceptional since a deposit in my bank allows me to buy commodities using a cheque just as I could with coins or paper money. At this empirical level bank deposits can obviously function as a means of purchase, and seemingly should be counted in the money supply.

However a moments thought indicates that when we talk of a money supply we are using a metaphor. We are making an analogy between a supply of some commodity - say petroleum, and money. But the two things are very different. Petroleum is something physical, it has mass. A supply of oil has a characteristic dimension, which we use to measure it. The world supply of oil would be expressed as x million barrels per day. In terms of dimensional analysis it is expressed in units of mass per unit time.

The money supply as conventionally measured is very different: it is measured in$ or £. We have argued above that such monetary units indirectly measure the amount of work that a social agent has done for the state. The transfer of monetary tokens allows agents who have done work for the state to transfer symbols representing this work to other people who use them to meet their tax obligations. It is thus evident that the money supply differs from a conventional supply in a number of respects:

(1) It does not measure the rate at which something is produced over time. Its measure has no time dimension, thus it is not a supply in the normal sense of a flow of things.

(2) Money is not a substance, it has no mass, instead it is a technology of record. As such it is information, or more properly, what we call money is a projection, in the strict relational sense, of an information structure. We have seen that holdings of coin by the public are one

column of a relation between subjects and the state. The other hidden column are individual tax debts. This column is hidden it takes the form of covert records held by the exchequer.

The illusion that it is a substance, which is implied by the term *money supply*, arose from a particular stage in its technological evolution. When the state of the monetary information structure is partly encoded by people's holdings of coins, these coins, which encode but one column of the monetary relation, were seen as money itself. Money was misidentified as a self sufficient substance. We should understand instead, that what exists is a relation, initially between the state and its subjects, later extending to credit relations between its subjects, whose partial projection appears as coin.

Digression 10.3 Relations

In this chapter we use the word relation in the sense in which it is used in formal logic and the *relational algebra*, the branch of mathematics used in computer databases. Much of what is modelled in computer databases are relationships between people and the social institutions. But the relational algebra can be used to model other types of information.

The concept of a relation comes from the idea of a predicate: a statement that can be true or false. For instance x is a dog is a predicate which can be true or false depending on what we put in the position x. Thus ”Fido is a dog” would be true but ”Mount Everest is a dog” would be false. That was a unary predicate. Predicates can have multiple arguments. The property of one number being less than another : x is less than y. We would normally express this as $x < y$ where the symbol $<$ for less-than is termed a relational operator.

In relational algebra we extend this idea to say that the relation less-than is the set of all pairs of numbers $[x, y]$ such that $x < y$. We can conceptualise this as an infinite table

1	2
2	3
1	3
17	203
-9	-1
..	.. etc

For less mathematical predicates we have relations as a finite table. The relation ”x is a known satellite of y” where y is an inner planet of our solar system, would give us a finite table:

Moon	Earth
Phobos	Mars
Deimos	Mars

Tables or relations can have more than two columns. We saw this with the table:

$T =$

Agent	Coin	Tax
State	2	6
Alande	0	-2
Tunde	0	-2
Femi	7	-2

Here the logical relation encoded by the table T is: x has y coins and is due to receive z in tax. When we *project* a relation we drop one or more columns from the table. Thus

T *Project* [Agent,Coin]=

Agent	Coin
State	2
Alande	0
Tunde	0
Femi	7

This is what we mean when we say that money as conventionally understood is a projection of an underlying relation.

The notion of projection in this context comes from the way that a camera-obscura projects or throws an image of three dimensional objects onto a flat surface. The key here is dimension reduction. A space of high dimension is represented in a lower dimensional one. Such projections involve a loss of information and can give rise to illusions, one only has to think of engravings by Escher such as ‘Belvedere’ or ‘Waterfall’, to realise how ambiguous a two dimensional projection can be.

The relation T above is composed of rows each of which has three values : a person, their cash and their tax position. These triples can be thought of as defining points in a three dimensional space. When we consider only the person and their cash we are projecting from three to two dimensions.

CHAPTER 11

CREDIT AND CAPITAL
Cockshott

In this chapter we discuss the formation of credit money and show that it necessarily arises from the basic law of motion of a capitalist society.

11.1 BANK CREDIT

In pre-capitalist economies lending was almost exclusively from rich to poor, and the rich lent out their own money[1]. In capitalist economies rich banks still lend to people who are much poorer, but the banks in turn borrow money from their depositors. The process of depositing money with banks is the crucial step in the creation of net credit.

There are three crucial innovations introduced by capitalist banking:

(1) The acceptance of deposits.

(2) The establishment of a system of mutual debt clearing.

(3) The issue of loans denominated in the liabilities of the bank.

These are tied to the creation of an independent system of records and ledgers recording the debit/credit position of bank customers. Bank accounts either as paper ledgers or as computer databases are an algorithmic system of record that allow debt relations to be held in a relatively centralised fashion. In section **??** we introduced the D matrix. This is a conceptual abstraction that models the mutual debts of agents in the economy. It is a square matrix, so that if there are *n* agents, people or firms, in the

---

[1] See Itoh and Lapavitsas (1999), chapter 3.

Table 11.1: The D matrix stored as a ternary relation.

debtor	creditor	amount
4	1	5
6	1	2
6	3	7
5	6	9

economy, then the D matrix will contain n^2 numbers. This is fine as a conceptual abstraction, but in a real society such an abstract entity can only exist if there is a technology capable of supporting it.

If there are 100 million people in an economy, the D matrix would contain 10,000 million million numbers. This is an impractically large number. Suppose the D matrix is stored in a distributed fashion, with everyone having their own double entry ledger system in which they wrote down their debts with everyone else. Then in an economy of 100 million people they would each need an account book with 100 million pages. Using standard accounting ledgers each person's account books would take up about 10 miles of shelf space. These sort of figures seem to suggest that the whole idea of a credit economy is impossible. But most of the people in an economy never meet one another, never trade, and never build up direct mutual debts. This means that the D matrix is *sparse*. A sparse matrix is one whose elements are mostly zero.

Sparse matrices can be recorded compactly, since the zeros never have to be written down. For instance the D matrix

$$\begin{array}{cccccc} 0 & 0 & 0 & 5 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 7 \\ -5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -9 \\ -2 & 0 & -7 & 0 & 9 & 0 \end{array}$$

is sparse and can be more compactly represented by a relation as shown in figure 11.1, that records each debt only once and lists the row and column numbers in the D matrix of the creditors and debtors. We have replaced a matrix containing 36 numbers with a relation containing 12 numbers.

Table 11.2: How a bank need only store credit information as a binary relation

customer	amount
1	2
3	7
5	-9

In this example agent 6 is acting as proto-banker. They borrowed from agents 1 and 3 and lent the proceeds to agent 5.

But the relation we have drawn is a holistic one, it still tabulates the totality of debt relations between all of the agents. From the standpoint of agent 6 it would be possible to keep a simpler private relation recording their account balances with their customers as shown in figure 11.2. Here the relation is reduced to two columns since the bank knows that it is one party to all of debts. If we assume that a large part of debt relations in a capitalist society take the from of loans to or from banks or other deposit takers, then these can be modelled by two column relations of the sort above. Indeed this is more or less how they now exist in the relational databases of the banks.

Other trading organisations who have credit relations with a large number of customers can use similar techniques to record their debt relations with their trading partners. Since the greater number of these trading partners will be private individuals who do not have the resources to maintain elaborate systems of account, it follows that the stored debt relations will largely be held by firms. Overall we can assume that the bulk of all the debt relations in society can be modelled by such binary (two column) relations.

Given this columnar representation, then amount of storage space required to record the information grows in proportion to the number of non-zero debts that actually exist, rather than in proportion to the number of potential debts which would be the case with the full D matrix. Whereas the size of the D matrix will grow in proportion to the square of the number of people in the economy, the total number of non-zero debts will grow much more slowly. We can express it as the product of two terms nd where n is the number of people, and d the average number of other agents with whom a person has debts. We can expect that as an economy grows, d will rise, but the rate of growth will be relatively slow: both with respect to time and with respect to the growth of the population.

The development of the banking system led to the growth of a specialised branch of the division of labour associated with the maintenance, storage, and updating of credit relations. These were typically recorded on paper in an indelible un-erasable fashion. Such records leave an audit trail of previous balances in the ledger books. However this is not an essential feature of private credit relations. They can also be recorded in the same way as subjects credit with the state - by the physical holding of tokens. At various times private individuals, firms and municipalities have issued coin like tokens. Typically these would be given as change for higher denomination royal money in small purchases. A grocer might issue his own farthings as change for state pennies. Other tradesmen in the neighbourhood would accept them, accumulate them and periodically have them redeemed in royal money by the issuer[2]. This practice continued on a larger scale from the 18th century with the issue of paper banknotes by private banks.

In return for a deposit of coin, they would issue paper notes, which, like tradesmen's tokens, were redeemable in royal money by anyone who presented them to the bank. The practice died out in England during the 19th century, but Scottish and Northern Irish banks continue to issue their own private banknotes in this way. Such tokens are an abacic rather than algorithmic system of record. They record a binary relation of indebtedness between the bank and other economic agents who physically hold the notes. Changes in this relation are brought about by physically handing over banknotes to another agent - a process analogous to physically moving the beads on an abacus to change the number it records.

In English we retain the word banknote to refer to paper money issued by the state bank though nowadays these are better seen as an extension of the coinage system. This reflects the history of paper money. The successful issue of paper money in the West was pioneered in Britain where capitalist economy was most advanced. The Bank of England started to issue paper money in 1694 followed by the Bank of Scotland in 1695. Whereas the latter existed primarily to finance private trade the Bank of England was owned by private shareholders but its main function was to make loans to the Crown. The Crown could pay for purchases using notes that the Bank

---

[2] See Berry (1988)

Digression 11.1 Paper money in socialist economies.

Our discussion has been about the institution of paper money under capitalism. 20th century socialist economies also used paper money. Why did this exist and what caused it to circulate?

Socialist states did not obtain their revenue from general taxation as capitalist ones did. Instead, their revenues came from the profits of state owned industry. It was the need of citizens to pay taxes that had forced Pounds, Dollars, Yen and D-Mark to circulate in the West. Since this need was absent in the USSR, why did the Rouble circulate?

The invention of money allowed states to separate the real appropriation of a surplus product from its symbolic appropriation as tax. Socialist states were the direct real appropriators of the social product, and as such had no need for money taxes. Money wages acted, instead, as a method for distributing the social product. Roubles circulated because state shops accepted them for purchase of consumer goods.

The Rouble was not a universal instrument of purchase like the Dollar. Rouble accounts did not entitle state enterprises to purchase arbitrary means of production except where these transfers had already been authorised by the Plan.

The Rouble could in principle have been replaced by notes indicating that the bearer had perfomed a given number of hours labour for society. That this did not occur was probably due to the continued existence of wage differentials, albeit small when compared to Western income differentials.

issued and would in return accept notes on the Bank issued as payment for taxes. This went alongside the process by which the Bank would issue notes to customers in return for coin deposits. Over the course of time the Bank of England became a state owned bank and its issue of notes became functionally indistinguishable from the issue of coin by the mint.

State paper money proper was invented much earlier in China. About one thousand years ago the Song dynasty had established an effective system of paper money.

> In 1161, the Southern Song government essentially replaced its bronze coin standard with a new paper money system known as huizi. Regional huizi currencies also proliferated in the Southern Song, and even in petty transactions the xiaoping coin was replaced by a two-cash coin known as zheher. *von Glahn (2004)*

In this context is worth noting that the monetary theories of European writers like Ricardo, Marx and Menger are parochial. They take as given that money must be made of a precious metal. In doing this they ignore the history of money in China. There money had frequently been made entirely of base metals such as bronze and iron, and, by the Song period, even this had been in large measure replaced by printed paper currency. The keys to this were:

(1) Availability of paper - at a time when Europe was still using parchment derived from animal skin.

(2) Knowledge of printing

(3) A large state with a professional salaried civil service made it quite feasible for the government to ensure that its money circulated by virtue of accepting it for tax payments.

According to von Glahn chartalism, the principle that the value of money is determined by the monetary authority irrespective of the use value of the substance employed as money, was a fundamental tenet of classical Chinese monetary analysis[3].

This would make it seem that the predominance of metallist doctrines in Europe is a reflection of the long period of European barbarism and disunity between the fall of the Carolingian empire and the establishment of the EU.

During this warring states period in Europe, states were small, and for much of the time lacked professional salaried civil services. Their tax collecting apparatus was poor and they were not able to so effectively enforce the circulation of national coinage unless it was backed by gold that could be used in internal European trade between the petty states. This local and temporary historical phenomena was universalised in metallist doctrines.

The first European bank-notes though were quite different from Chinese paper money. Instead of just meeting the needs of the state they arose also to meet the needs of capital. To understand this one needs to understand the signature of capital.

11.1.1 The signature of Capital

The notion of Capital having a signature is derived from Marx. He characterised the process of buying and selling commodities as having the signature $C \rightarrow M \rightarrow C$ a notation that he used to indicate that an agent starts off

---

[3] See von Glahn (1996), pp. 23-47.

with a commodity (C), sells it for money (M) and then uses the money to purchase another commodity (C).

> Let us now accompany the owner of some commodity say, our old friend the weaver of linen to the scene of action, the market. His 20 yards of linen has a definite price, £2. He exchanges it for the £2, and then, like a man of the good old stamp that he is, he parts with the £2 for a family Bible of the same price. The linen, which in his eyes is a mere commodity, a depository of value, he alienates in exchange for gold, which is the linen's value-form, and this form he again parts with for another commodity, the Bible, which is destined to enter his house as an object of utility and of edification to its inmates. The exchange becomes an accomplished fact by two metamorphoses of opposite yet supplementary character the conversion of the commodity into money, and the re-conversion of the money into a commodity. The two phases of this metamorphosis are both of them distinct transactions of the weaver selling, or the exchange of the commodity for money; buying, or the exchange of the money for a commodity; and, the unity of the two acts, selling in order to buy.
>
> The result of the whole transaction, as regards the weaver, is this, that instead of being in possession of the linen, he now has the Bible; instead of his original commodity, he now possesses another of the same value but of different utility. In like manner he procures his other means of subsistence and means of production. From his point of view, the whole process effectuates nothing more than the exchange of the product of his labour for the product of some one else's, nothing more than an exchange of products.
>
> The exchange of commodities is therefore accompanied by the following changes in their form.
>
> ```
>       Commodity - Money - Commodity.
>               C --- M --- C.
> ```
>
> The result of the whole process is, so far as concerns the objects themselves, C - C, the exchange of one commodity for another, the circulation of materialised social labour. When this result is attained, the process is at an end.
>
> *(Marx (1954), Chapter 3, Section 2A)*

In the circuit $C \rightarrow M \rightarrow C$ the value that weaver starts out with £2 of linen and ends up with £2 in the form of a Bible. Marx goes on to contrast this with the signature of capital $M \rightarrow C \rightarrow M'$, where M' represents an augmented sum of money $M < M'$.

> The simplest form of the circulation of commodities is $C - M - C$, the transformation of commodities into money, and the change of the money back again into commodities; or selling in order to buy. But alongside of this form we find another specifically different form: $M - C - M$, the transformation of money into commodities, and the change of commodities back again into money; or buying in order to sell. Money that circulates in the latter manner is thereby transformed into, becomes capital, and is already potentially capital.
>
> ....
>
> Now it is evident that the circuit $M - C - M$ would be absurd and without meaning if the intention were to exchange by this means two equal sums of money, £100 for £100. The miser's plan would be far simpler and surer; he sticks to his £100 instead of exposing it to the dangers of circulation. And yet, whether the merchant who has paid £100 for his cotton sells it for £110, or lets it go for £100, or even £50, his money has, at all events, gone through a characteristic and original movement, quite different in kind from that which it goes through in the hands of the peasant who sells corn, and with the money thus set free buys clothes.
>
> ....
>
> The exact form of this process is therefore $M - C - M'$, where $M' = M + \Delta M$ = the original sum advanced, plus an increment. This increment or excess over the original value I call "surplus-value". The value originally advanced, therefore, not only remains intact while in circulation, but adds to itself a surplus-value or expands itself. It is this movement that converts it into capital. *(Marx (1954), Chapter 4)*

The merchant having converted $M \rightarrow M'$ will want to build on his success, turning his M' back into commodities to sell again. The signature of capital thus implies a process of exponential growth

$$M \rightarrow C \rightarrow M' \rightarrow C' \rightarrow M'' \rightarrow C'' \rightarrow M'''$$

If the whole class of merchants are doing this it implies that over time there must be an exponential growth in the sum of money in their hands.

Table 11.3: Growth of the world gold stock 1840 to 2000

period		stock million troy oz	Annual growth rate %
1840	1850	617.9	0.27
1851	1875	771.9	0.89
1876	1900	953.9	0.85
1901	1925	1430.9	1.64
1926	1950	2130.9	1.61
1951	1975	3115.9	1.53
1976	2000	4569.9	1.54

Note that only some of this stock would have been available for use as coin, decorative and other uses absorbing the rest. Calculated using data on annual production published by Gold Fields and Mineral Services Ltd.

The thrust of this chapter has been to argue that money is a technology of record, that it is essentially information about social power. As an information structure there is no inherent obstacle to its exponential growth. If you use a place number system, either binary, decimal or the old Babylonian base 60 system, then the size of the number you can write down grows exponentially with the number of digits. But if you use a token number system like gold coins to encode social power, then exponential growth is a problem. It implies an exponential rise in the mass of gold, an altogether more difficult matter. As shown in Table 11.3 the annual rate of gold stock growth has been around 1.5% per annum for the last 100 years, about half that for the 19th century and was considerably lower prior to the discovery of the Californian fields in the 1840s. Were gold coin the only form of money in which capital could accumulate, then the circuit

$$M \to C \to M' \to C' \to M''$$

would have been limited to very low rates of return on capital.

The signature of capital was incompatible with gold money. It demanded new monetary technologies, the first of which was the paper banknote. It was no harder to print a £50 banknote than a £5 banknote.

At the end of the week workers have to be paid, the firms' products have to be sold, and stocks of raw materials replenished. This is composed of a vast number of atomic transactions each of which is, taken by itself, value conserving. In reality of course they do not all take place at the end of the week. That is just a pedagogic simplification. However, we can partition the set of atomic transactions into three groups:

(1) Transactions between capitalists.
(2) Payments of wages by capitalists.
(3) Purchases of consumer goods by workers.

An Iron Master taking delivery of coal would typically write a bill of exchange, a private certificate of debt, promising to pay within 30 or 90 days.

Payment of wages would generally have to be done in cash. Capitalists have tried at times to issue tokens as wages which would be redeemable only at company stores, but legislation by the state, eager to maintain its monopoly of coinage tended to put a stop to this. Payment in cash represents a transfer from the safes of capitalists to the pockets of their employees, with a corresponding cancellation of wage debts.At the end of the week, the wage debt has been cleared to zero, and there has been an equal and compensating movement of cash .

Workers then spend their wages on consumer goods. For the sake of simplicity we assume that there is no net saving by workers so that in the course of the week all of the money they have been paid is spent. This implies that immediately after payday, the money holdings of the workers were equal to one week's wages. If these wages were paid in coin this would have set a lower limit to the quantity of coin required for the economy of function.

When workers spend their wages on consumer goods they transfer money only to those firms who sell consumer goods: shopkeepers, inn-keepers etc. We can expect these firms to not only make up the money they had spent paying wages, but to retain a considerable surplus. Wages would add up to only a fraction of the value of the consumer goods. The final sellers of consumer goods will thus end up with more money than they paid out in wages. From this extra cash, they can afford to redeem the bills of exchange that they issued to their suppliers.

If we assume no bank credit, then suppliers of manufactured consumer goods would be entirely dependent for cash on money arriving when the bills of exchange, in which they had been paid, were eventually redeemed by shopkeepers and merchants. The payment situation facing raw materials firms was even more indirect: they could not be paid unless the consumer goods manufacturers like weavers, potters, and millers had sufficient cash to redeem bills of exchange issued for yarn, coal, grain etc.

The process of trade between capitalists leads to the build-up of inter firm debt.We suggest that the total volume of inter firm debt that could be stably supported would have been some multiple of the coinage available after allowing for that required to pay wages. If one takes the aggregate of all firms the ideal signature of this process can be represented as:

$$M \rightarrow [C \Rightarrow (C + \Delta C)] \rightarrow M + \Delta M$$

where $[C \Rightarrow (C + \Delta C)]$ represents the production process that generates a physical surplus of commodities after the consumption needs of the present working population has been met. If there is no new issue of coin by the state then the ΔM can not be real money, instead it must be in the form of bills of exchange and other inter-firm credit.

For the capitalist class considered as an abstract whole this should not be a problem since the ΔM is secured against the accumulated commodity surplus ΔC. There is a net accumulation of value as commodities, and accounting practice allows both the debts owed to a firm, and stocks of commodities to be included in the value of its notional capital. As the process of accumulation proceeds in this way the ratio of commercial debt to real money will rise. Suppose the period for which commercial credit is extended remains fixed - say at 90 days, then there a growing number of debts will be falling due each day. If these need to be paid off in money, then a growing number of firms will have difficulty meeting their debts in cash.

11.1.2 How much money is required

If we assume an economy in a steady state this causes no problems, but once you throw in the need for a capitalist economy to allow an exponential growth in capital values, problems arise.

How can we model this?

Consider an individual firm, what is the probability that it will not be able to meet its debts?

Let us first normalise the assets liabilities and cash of firms with respect to their turnover. We then assume that normalised to turnover, a firms expected gross assets and gross liabilities in terms of commercial credit will

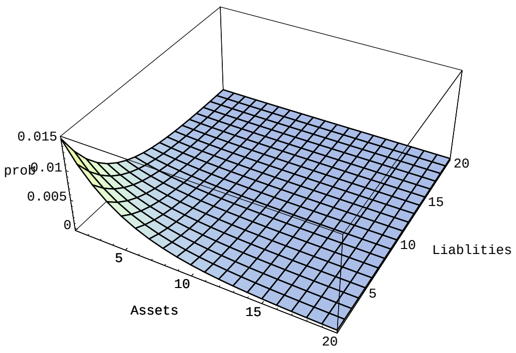


Figure 11.1: Conditional probability surface for assets and liabilities normalised to turnover.

follow a negative exponential distribution with a probablity density of the form:

$$P(d) = Ke^{-\frac{d}{t}} \tag{11.1}$$

where d is debt, K is a normalisation constant, and t is the firms turnover[4]. The same distribution is assumed to apply to the debts owed to the firm. There is thus a two dimensional probabilty surface relating assets to liabilities as shown in Figure 11.1. We can use this probability surface to estimate the probable distribution of firms along the net-creditor/net-debtor axis as shown in Figure 11.2. Note that the PDF peaks where firm has zero net commercial debts, but that this probablity is actually very small. It is more likely that they will have either a net credit or debit balance in their dealings with other capitals. The probability distribution is symetric, since to every commercial debtor there corresponds a commercial debtor. It falls of steeply on either side. We now have to consider two things:

a. How much of the debt will be falling due each month.

b. How likely is it that a firm will have insufficient cash to meet their debts at the end of the month.

---

[4] This is of course the Gibbs Boltzmann distribution discussed in Chapter 9 and in A. Dragulescu and V. M. Yakovenko (2000). The actual distribution may be either this or a power law but the argument that follows is robust in either case.

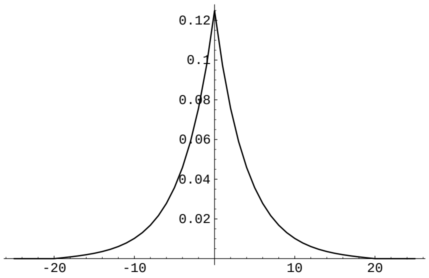


Figure 11.2: Plot of the probability distribution expected for firms along the net asset/net creditor axis with respect to commercial debts outstanding. Note that the plot is zero centered because of the symetry of the commercial debt relationship.

It is obvious that the longer the period p for which commercial credit is extended, the smaller will be the amount of debt falling due each month. If it were the custom to extend 90 days credit, then a $\frac{1}{3}$ of the debt will fall due each month, as opposed to all of it for a 30 day commercial credit rule.

In order to work out how likely it is that a firm will have too little money to pay its debts we need to have some model of the distribution of money balances in firms. We are provided with this from the data shown in Figure 9.10, which found that in the SA model, the probability of a firm having money holding x was given by the Paraeto distribution:

$$P(x) \propto (\frac{x}{m})^{-(\alpha+1)} \tag{11.2}$$

where $\alpha = 1.4$, where m is the average money holding of a firm. Using this we can plot the probability surface $C(l,\mu)$ relating liabilities l cash μ (Figure 11.3).

Firms will be unable to meet their bills if :

a. They have net liabilities l.
b. The net repayments on these debts, $r = \frac{l}{p}$ where p is the period on commercial loans, is greater than the current cash balance.

Thus the probability of a firm defaulting will be given by

$$P_{default} = \int_{-\infty}^{0} \int_{0}^{\infty} C(l,\mu).(\frac{l}{p} + \mu < 0)\, d\mu\, dl \tag{11.3}$$

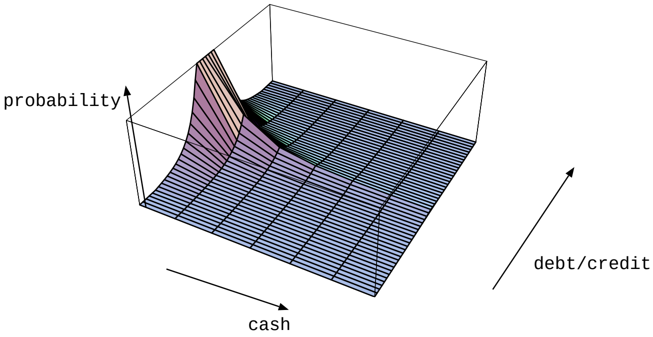


Figure 11.3: Plot of $C(l,m)$ the probability distribution expected for firms along the net asset/net creditor axis with respect to cash holdings.

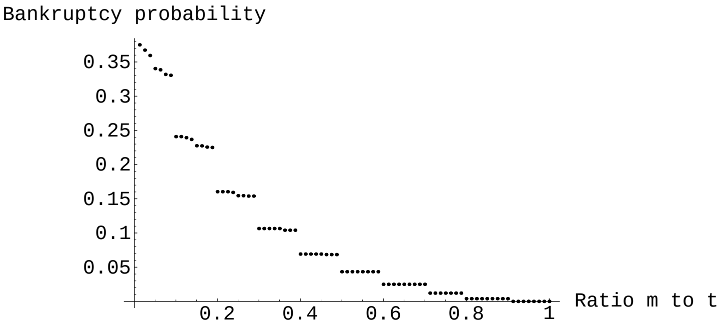


Figure 11.4: Plot of the fraction of firms going bankrupt in a time period as the ratio of the mean money holding m in equation 11.2 to the mean turnover t in equation 11.1, varies. The evaluation was done on the basis of a 3 month duration of commercial loans. The steps in the curve are the result of 'binning' as the functions were evaluated on a discrete grid.

However the shape of the surface $C(l,\mu)$ will depend on the amount of money in circulation. As the probable amount of money held by each firm rises, the default probabilty will fall. In figure 11.4, we see how the probability of bankruptcy declines as the ratio of cash to turnover rises.

It is important to note that what is being considered here is purely stochastic bankruptcy due to cash-flow fluctuations. It is quite aside from bankruptcies that may occur due to economic inefficiencies or long term rises in costs. It is the bankruptcy that can hit perfectly viable firms due to random fluctuations in in indebtedness.

11.1.3 Necessity of paper money

Let us outline the argument so far:

1. We know that the signature of capital implies an increase in the values of commodities being traded over time.
2. We know from the historical record of gold production, that the rate of increase of gold stocks was relatively low, certainly much lower than one would expect capital stocks to grow.
3. This implies that the ratio of cash to commodity turnover would tend to fall in an economy using gold for its coinage.
4. The consequence would be that an increasing fraction of capitals would have insufficient cash holdings to meet the liabilities falling due each month, and would thus become bankrupt.

This mechanism provided a basic engine of commercial crises under the gold standard, and, in the absence of other innovations to replace gold money, it would have limited the rate of growth of capital to the rate of growth of the gold stock. This applied to the proportionate rather than the absolute growth of the gold and capital stocks. From Table 11.3 we know that the world gold stock rose by about 154 million oz in the 25 years from 1851 to 1875. At that time a the price of gold in terms of UK currency was £2.87 per oz, being determined by the weight of metal in a gold sovereign coin. The gold mined during that period was thus worth £441million, an annual increase of about £17million. Did this mean that the total world capital accumulation in those years could only have been £17million per annum?

No. It means that the annual growth in monthly turnover that could be supported by a gold currency would have been limited to around £$(p \times$

17)million, where p was a constant determined by the period of commercial loans[5]. We would expect from the labour theory of value that the value of turnover would increase along with the population producing commodities. As more people became engaged in commodity production, both because because population had increased, and as natural peasant economies were replaced by production for the market, there would have been a proportionate rise in the turnover of commodities on the capitalist world market. Whilst the absolute increase in turnover supported by gold would be some multiple of the actual gold production, the *percentage* increase in turnover would still have been limited by the low percentage increase in gold stocks. It would have been, that is, had the banking system not been able to create alternative methods by which commercial debts were paid.

We have already shown how the process of banks accepting deposits in coin and making loans in coin will create net credit - increase the Manhattan distance between agents. The issue of banknotes accelerates the this process enormously. Suppose Mr Lang made a deposit of £100 in coin with the Bank of Scotland in 1696. If the bank were to then make a loan of £90 in coin to Mr Strang, there would be a creation of net credit as we showed above. But what the bank actually did was much more dramatic. It used the £100 as a reserve against which made loans of several hundred pounds in its own banknotes. The notes were redeemable on demand against royal coin, but because of the greater convenience of paper money, the bank could count on only a small fraction of the notes actually having to be redeemed on any banking day. This was termed fractional reserve lending.

The deposits and withdrawals by customers are the result of multiple independent cricumstances. As such they have a noisy character analogous to the ”shot noise” discussed in section 3.3.3. Recall that shot noise set a limit to the information capture accuracy of any camera or photo-sensor due to the discrete photon nature of light. Recall too that shot noise was proportional to the square root of the mean number of photons arriving at each sensor during the exposure period of the camera. As a result shot noise falls as proportion of the total signal the more photons we can capture. A similar principle applies to banking. The more customers that a bank has, the smaller will be the proportionate variation in the withdrawals from day to day.

---

[5] The *stocks* of capital taking the form of plant and machinery could have grown more than this, depending, among other things, on the depreciation period of capital equipment.

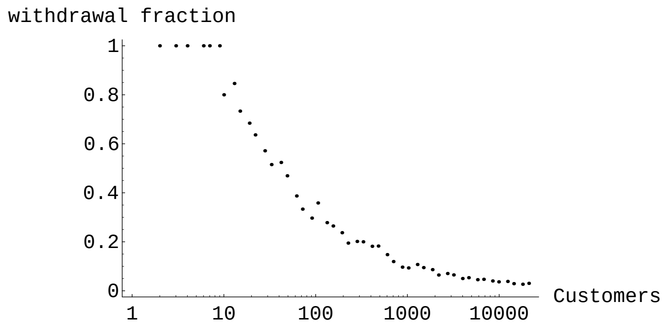


Figure 11.5: Plot of the largest fraction of a bank's deposits withdrawn in a single day over a 20 year period as the number of customers rises. The plot comes from a simulation in which it was assumed that a customer might withdraw any sum up to their maximum deposit, and that customers were as likely to make deposits as withdrawals. The slightly irregular nature of the trace reflects the underlying stochastic properties of such a simulation.

Look at Figure 11.5. The horizontal axis shows the number of customers, and the vertical axis the highest proportion of the bank's deposits withdrawn in any week over a 20 year period. As the number of customers rises, the variation in the amount withdrawn in any week falls, and so too does the maximum withdrawal that can be expected. A very small bank would have to keep all its deposits in the safe as an insurance against having to pay them out, but a bank with 20,000 customers might never see more than 3.4% of its cash deposits withdrawn in any week. A bank with that number of customers could safely issue as loans 20 times as much in paper banknotes as the coin that it held in its vaults, safe in the knowledge that the probability of it ever having to pay out that much in one day was vanishingly small.

This creation of new paper money by the banks was the hidden secret behind the signature of capital $M \to C \to M' \to C' \to M''$. Unlike a creation of token money by the state, or the issue of gold coin, there was no unproductive withdrawal of value in the form of seigneurage or the mining costs. Loans in bank money were made to capitalists (A) who were expanding their business. With these loans group A either directly purchased, or through wage payments financed the sale of, the surplus produced in the workshops of other established capitalists (B). The surplus product sold by group (B) was thus on the one hand converted into money, (phase $C \rightarrow M + \Delta m$), and at the same time became new capital for group A (phase $\Delta m \rightarrow \Delta c$). This money creation was distinct in its effects from the Chinese state's creation of paper money. In the latter, the money first appears as a purchase of commodities by the state. Whilst this would still create money profits for the merchants who sold to the state, the surplus so purchased was consumed by the state rather than becoming capital.

We are here considering bank money in its initial historical form, prior to the widespread use of chequing accounts or even more modern forms like debit cards. Such more modern forms have their own technical preconditions, which we shall discuss later. But the Urgeld form allows us to understand much of what followed.

The first thing to notice is that bank money is derivative on state money. Bank notes or chequing accounts are denominated in state money. They ultimately gain currency through being redeemable in state money. The banks lack an independent coercive power analogous to the power to tax.

The second thing to notice is the crucial importance of size. Unless a bank reaches a certain threshold number of customers, it will be unable to operate a system of fractional reserve lending. In order to pass this threshold the Bank of England, the Bank of France and the Bank of Scotland were initially given Royal monopolies.

Figure 11.5 makes it clear that a big bank had much more money issuing ability than a small one. The larger the number of customers a bank had, the small the proportionate variation in it's weekly withdrawals, and hence the smaller the proportionate reserve it was forced to maintain. This would be reflected in larger bank's being more profitable, since a higher proportion of their assets would be in the form of loans on which they earned interest, rather than cash which earned them nothing.

The curve in Figure 11.5 comes from a simulation in which the probability of each customer making a deposit or withdrawal is independent. The customers are assumed not to collude in making withdrawals. This would normally be true, but were the bank's credit to be impugned, then customers can cease to act independently. A panic can set in and lead to a collective run on the bank. Because a small bank faced greater proportionate variations in its weekly withdrawals, its credit would be less secure than that of larger banks, and bank failures would be more frequent in small banks.

The combined effects of these two processes lead to a process of gradual centralisation of banks, with larger and more profitable ones taking over the smaller. A crucial factor in creating a large customer base was the creation of a branch network. Initially banks would have a single office in the capital city - just as the Bank of England still does. Only later did they develop a branch network. So long as the private issue of banknotes was their main form of money creation, notes issued in the capital could circulate among merchants in outlying towns, and distance would slow down demand for these notes to be redeemed. Suppose the Bank of Prudence had only one branch in the capital, as had the Bank of Temperance. Each issued its own banknotes. The only way in which these notes were redeemed was when they were presented at their head office. These notes would circulate mainly in the capital, but some would go the rounds in the main provincial towns. If the Bank of Prudence opened branches in Gloucester, Chester, Bolton, Halifax etc, it gained several advantages:

- By opening a multitude of branches it obtained more depositors and thus increased its loanable capital stock.
- The increase in the number of customers reduced the variation in its weekly withdrawals and allowed it to operate with smaller reserves.
- Some of its depositors would have made their deposits, not in cash, but with the notes of its rival the Bank of Temperance. The Bank of Prudence would return these to the head office of the former bank, to be redeemed. This would have increased the rate at which Bank of Temperance notes were being cashed and forced it to hold higher reserves.

There were thus strong competitive pressures forcing the establishment of regional, provincial or national networks of bank branches. The existence of a network of branches, when combined with the fast transport provided by the new railways, allowed the next phase in the evolution of bank money : cheques.

Payment by drafts on Merchant Banks was a much older practice, dating back to the middle ages. International merchant companies - primarily Italian, would arrange payments in say Florence against a bill issued, for example, by their Paris office. This payment system grew out of the practice of issuing commercial bills of exchange, and was used primarily to finance international trade. The 19th century establishment of dense branch bank networks allowed chequeing transactions to penetrate into general domestic commerce between capitalists[6]. The development of cheque money marks a general social transition between abacic and algorithmic calculation. With banknotes, the social wealth relation persisted in the physical placement of notes: whose pocket were they in? With cheque accounts it persisted in the ledgers of the banks. Their establishment meant that the banks had to build up an information processing machine, a machine of flesh, brain, paper and ink. The routine processing of this information became a major branch of the division of labour and generated a whole social stratum of bank clerks devoted to its operation.

The basic operations being performed by this social computer were payments between accounts. With cash a payment had been a simple movement of coin. With cheques it looks almost as simple, you hand a cheque to the shop to make your payment. But this physical handover only aquires meaning when supported by an interpretation mechanism provided by the banking system. A cheque is a record volant, a cartesian of the form

(bank$\otimes$payee$\otimes$sum$\otimes$payer)

written out in longhand as :

National Commercial Bank
Pay *William Sydney*,
the sum of $\boxed{\text{£ } 10}$,
James Ross

A cheque $(x \otimes a \otimes n \otimes b)$ is a procedure call on the banking system to perform the action

```
Atomic Procedure   cash(   $(x \otimes a \otimes n \otimes b)$   )
    if   Account[b]>n   then
             Account[a]    ←    Account[a] + n
             Account[b]    ←    Account[b] - n
```

By **Atomic** we mean that the operation is all or nothing. It is impermissible for one account to be updated but not the other.

Even when using modern computer technology performing such atomic updates is not trivial. One has to write code to ”lock out” any other programs which might be updating accounts whilst this update occurs. The problem become greater when you consider that the algorithm above sim-

---

[6] Bank accounts did not spread to the mass of the working class population until the late 20th century

plifies things by ignoring which banks are involved. The actual process of cashing a cheque involves handing it into your own bank. William Sydney would hand the cheque in to his branch of say the British Linen Bank, but the cheque is drawn on the National Commercial Bank.

We must thus fill in more detail in the algorithm:

```
Atomic Procedure cash( y , $(x \otimes a \otimes n \otimes b)$ )
    if Account[b]>n then
        y.Account[a] ← y.Account[a] + n
        x.Account[b] ← x.Account[b] - n
```

We see that a's account at bank y is credited and b's account at bank x is debited. Nowadays these accounts would be held on different computers and the update involves what is termed a distributed atomic transaction. It took a considerable period before reliable techniques to do this were developed for electronic computers see Bernstein Bernstein et al. (1980) and Irving Traiger et al. (1982). In ink and paper days, the "computers" that held the accounts were systems of clerks and ledgers, but there were still two distinct systems: that of the British Linen Bank, and that of the National Commercial Bank in our example. The cheque instructed the National Commercial to pay money to Mr Sydney. If Mr Sydney went to the offices of the National Commercial they could just hand over cash and debit Mr Ross's account. The National Commercial could not directly modify Mr Sydney's account since that was recorded in the ledgers of the British Linen.

If Mr Sydney handed the cheque into his branch of the British Linen bank, they could easily credit his account but would want to know that Mr Ross had the funds to meet it. Had Mr Ross an account at the same bank, the updating clerk, could turn to Mr Ross's page and verify that his account was sufficiently in credit. But Mr Ross's record is held by another company, so the procedure was to provisionally update the Mr Sydney's record, and send the cheque to the National Commercial to *clear*. The provisional update would be cancelled were the National Commercial to return the cheque as invalid.

Clearing referred to the process by which, at a central place, the different banks exchanged the cheques drawn against them and computed the net inter-bank monetary transfers that resulted. Babbage, as ever a keen observer of the technical details of economic interchange, described it:

173. Clearing house. In London this is avoided, by making all checks paid in to bankers pass through what is technically called The Clearing House. In a large room in Lombard Street, about thirty clerks from the several London bankers take their stations, in alphabetical order, at desks placed round the room; each. having a small open box by his side, and the name of the firm to which he belongs in large characters on the wall above his head. From time to time other clerks from every house enter the room, and, passing along, drop into the box the checks due by that firm to the house from which this distributor is sent. The clerk at the table enters the amount of the several checks in a book previously prepared, under the name of the bank to which they are respectively due.

Four o'clock in the afternoon is the latest hour to which the boxes are open to receive checks; and at a few minutes before that time, some signs of increased activity begin to appear in this previously quiet and business-like scene. Numerous clerks then arrive, anxious to distribute, up to the latest possible moment, the checks which have been paid into the houses of their employers.

At four o'clock all the boxes are removed, and each clerk adds up the amount of the checks put into his box and payable by his own to other houses. He also receives another book from his own house, containing the amounts of the checks which their distributing clerk has put into the box of every other banker. Having compared these, he writes out the balances due to or from his own house, opposite the name of each of the other banks; and having verified this statement by a comparison with the similar list made by the clerks of those houses, he sends to his own bank the general balance resulting from this sheet, the amount of which, if it is due from that to other houses, is sent back in bank-notes.

At five o'clock the Inspector takes his seat; when each clerk, who has upon the result of all the transactions a balance to pay to various other houses, pays it to the inspector, who gives a ticket for the amount. The clerks of those houses to whom money is due, then receive the several sums from the inspector, who takes from them a ticket for the amount. Thus the whole of these payments are made by a double system of balance, a very small amount of bank-notes passing from hand to hand, and scarcely any coin.

Babbage (1832) Section II, Chapter 13, paragraph 173

Technology changed. The mid 19th century brought the electronic transfer of funds by telegraph, but behind this modernity stood the armies of clerks calculating and updating records by hand. The Morse impulses still had to be translated into paper and ink. Till the 1960's banks still closed their doors at three to allow the manual tallying up of accounts. Modern electronic bank money required two critical inventions: the programmable electronic computer described in section **??** and, less obviously, the random access disk drive.

The first electronic computers used a bewildering variety of memory devices. Turing (1937) had famously proposed his tape marked with symbols, a sort of half way house between the punched paper tape that was used by contemporary telex machines and the squared paper used by mathematicians or school children. When he came to collaborate on the Colossus, Hodges (1983), it was paper tape that he used as the storage medium. Paper tape was relatively cheap, and it could hold data in a persistent fashion - it did not need constant energy input to remember things. But tape, whether magnetic or paper, is a sequential store. In order to read the 100th character on the tape one must first read the other 99. The first generation of computers was constrained to the use of sequential stores. For instance it was found that a television tube could be used as a sequential memory. Charge deposited by the cathode ray on the screen during one scan, persisted long enough to be detected on the next scan: Lavington (1978, 1980). Acoustic delay lines were another early alternative, but both of these were volatile stores. The information had to be constantly refreshed or re-written to the store if it was to be remembered. Such stores are obviously useless for storing bank records.

During the 1920s and 30s there had been a fairly extensive development of business automation based on punched cards. These had been invented to hold census data, with a record card being punched for each person recorded in the census. Decks of cards could then be fed through sorting machines that would select cards from the deck if a particular pattern of holes was punched in them. Other machines termed tabulators, would then use mechanical adders to calculate and print totals from the sorted cards. This technology had moved from census taking to stock control applications. For instance shoes were distributed to shops with a punched card in each shoebox. When the shoes were sold, the shop retained the the card and returned it at the end of the day to the supplier's warehouse. By feeding the returned cards through sorters and tabulators the warehouse could make sure that appropriate replacement shoes were dispatched the next day, and that the total bill for each shop was computed.

During the 1950's and 1960's companies like IBM and ICL that had originally been active in the manufacture of punched card machines moved into the computer market, extending the power of the tabulators with simple stored program computers. But these card and tape technologies were not still not well adapted to accounting operations. To understand why, let us first look at how a Turing machine could have been used to update bank accounts.

Turing's original description of his machine allowed for it to have only a single tape, but theorists have subsequently proposed Turing machines with two or more tapes. We will assume that the National Commercial Bank had a three tape machine. In this scenario the bank keeps its accounts on tapes and each evening it feeds in yesterday's tape to tape head 1 and places a blank tape on head 2. On tape head 3 it places a tape onto which all pay-in slips and outgoing cheques have been transcribed. At the end of its work the Turing machine has produced on tape 2 an updated ledger on which all the accounts originally listed on tape 1 have been debited or credited by appropriate cheques found on tape 3. We assume that both the transcribed cheques and the account records start with an 6 digit account code. The program for the machine is outlined in Digression 11.2. As the digression shows, a Turing machine would be quite capable of doing banking transactions, but it would be very slow. For each account on tape 1 it has to read the whole of tape 3 searching for cheques that relate to that account. Its running time will be proportional to NC where N is the number of accounts and C the number of cheques.

The computers of the 1950s and early 60s were only slightly more powerful than this. In addition to multiple tapes, they had a small number of words of auxilliary working store into which the records of a few accounts could have been read. Suppose a machine had sufficient working store to handle 100 account records. It could read in 100 accounts from tape 1 in a block and then search the whole of tape 3 for updates to these accounts before writing the updated block of accounts to tape 2. Such a machine would have a running time proportional to $\frac{NC}{100}$. This is obviously much faster than a pure Turing machine without an auxilliary working store could have processed the accounts. Despite the acceleration, the time order of the process

---

Digression 11.2 A Turing machine program to update bank ledgers.

(1) Output a start of record character to tape 2, and read in the start of record characters from tapes 1 and 3. If instead of a start of record character on tape 1 we got an end of file character, stop. If instead of a start of record on tape 3 there is an end of file character, goto state 8.

(2) Copy the entire record from tape 1 to tape 2, then rewind tape 2 to the start of the record.

(3) Read in the first digit from tapes 3 and 2. If they match goto state 4 otherwise goto state 7.

(4) Repeat step 3 another 5 times and then go to state 5.

(5) We have now accertained that the cheque is to debit the current account and we have written a copy of the account number to tape 2. Read in the amount from the cheque on tape 3 and the current balance from tape 2 and write the result of debiting the account to tape 2. Then go to state 6.

(6) Rewind tape 2 to the first digit of the current account number and goto state 1. We are now ready to process the next cheque on tape 3.

(7)
    (a) Rewind tape 2 back to the character immediately after the *previous* start of record marker.
    (b) Skip tape 3 forward to the character after the *next* start of record.
    (c) Goto state 3.

(8) We get here if we have scanned the whole of tape 3 for a cheque affecting the current account on tape 1.
    (a) We rewind tape 3 to the start.
    (b) We skip forward to the next record on tape 1.
    (c) We move tape 2 forward till we come to a blank space.
    (d) Goto state 1

$\mathbf{O}(NC)$ is still the same. The formula, pronounced “Order *NC*” means that the running time is still proportional to the product of the number of cheques and the number of accounts. Suppose that in the early 1960’s a small US bank had an IBM 1400 which could read 1000 records a second from its tapes, then the following table shows the time it would have taken to process different numbers of accounts.

Accounts	Cheques	Seconds	hr:min:sec
10,000	1,000	1,000	00: 16:40
50,000	5,000	25,000	06: 56:40
100,000	10,000	100,000	17; 46:40

The technical change which made electronic banking possible came when IBM introduced the world's first magnetic hard disk for data storage. RAMAC (or Random Access Method of Accounting and Control) offered unprecedented performance by permitting random access to any of the million characters distributed over both sides of 50 two-foot-diameter disks. Produced in San Jose, California, IBM's first hard disk stored about 2,000 bits of data per square inch and had a purchase price of about $10,000 per megabyte.

> Information is stored magnetically on a stack of 50 disks which rotate continuously at 1,200 RPM. Each metal disk is two feet in diameter and is coated on each side with a magnetic material. The face of a standard disk contains 100 tracks, in each of which 600 digits may be stored. In double capacity disk files there are 200 information tracks on each disk face. Thus, standard memory file capacity is six million digits and the double capacity Model 2 IBM 355 disk storage unit can store 12 million digits. Up to four random access memory units may be used in a RAMAC 650 system.
>
> In each memory file there are three electronically-controlled access arms containing magnetic heads. They read and write the information contained on the rotating disks. They act independently of each other, but each arm can be directed to any track in the file. As a result, there can be simultaneous seeking of three different records and information constantly is available for processing.
>
> Instructions are given to the three access arms from the IBM 650 console. A ”seek” instruction sends an arm to the desired data track. A ”read” instruction brings data from the disk, through an access arm, into the system's immediate access storage unit. A ”write” instruction results in the recording

> on the disk of information which has been in immediate access storage.
>
> *Original IBM sales literature for the RAMAC*

The development of random access disks gave rise to relational databases which provide the material embodiment of modern money. Once monetary relations were encoded in relational databases, it was a relatively simple step to develop Switch Cards and electronic payments. The wiring of money by telegraph had of course preceded this by a century, but when money was wired, the telegraph messages that arrived at the banks were processed by hand. Clerks had to read them and adjust ledgers. This restricted the wiring of money largely to inter-bank and inter-firm transfers.

Card payment systems originated with Dinner's Club and American Express, which were closed loop payment systems. A single firm acted to pay the merchants, bill customers and manage accounts. Initially the process was paper based with the cards being used to make records on carbon paper stubs. With Visa cards, multiple banks became involved as franchisors. The Visa payment system was thus open-loop, in that different firms were responsible for crediting merchants and billing customers - the process being an extension of the prior cheque clearing mechanisms. There was still a manual processing phase of data entry from the sales stubs, but the data was processed electronically from then on.

Switch or electronic payment cards allow automation of the entire process.

The process of going from coin, to banknotes to Switch cards and relational databases is one of increasing abstraction. With each step, the symbolic or formal character of money becomes more apparent. With coin, it still appears as if their metallic substance is crucial. The illusion persists at one remove with banknotes which for years appeared to be promises to pay coin. They thus could be represented as symbols for which the concrete referent was coin. With relational databases the illusions are stripped off, and the logical essentials of money as a technology of record are revealed.

CHAPTER 12

UNDERSTANDING PROFIT

Cottrell, Cockshott

> "the difficulty which has hitherto troubled the economists, namely to explain the falling rate of profit" (Marx, *Capital* III, p. 230)

The concept of a tendency for the rate of profit to fall was a common theme in classical political economy. Smith, Ricardo and Marx all held such a theory. However, their grounds for believing in this tendency were quite various. Smith thought in terms of accumulation leading to an increase in competition between capitals, hence driving down prices and profits. Ricardo dismissed this as a confusion—competition between capitalists influenced the distribution of profit, not its overall amount—and held a theory whose motor lay in the confrontation between rising population and diminishing returns in agriculture. Marx's theory was in a sense more akin to Smith's (at least insofar as it had nothing to do with diminishing returns) but it was quite distinct. Marx had a historical view in which economic institutions like capitalism were seen as being transient. The tendancy of the rate of profit to fall was, for him, symptomatic of this transience.

There are three elements in Marx's writings that are particularly relevant. The first we call the "main argument": this is the argument Marx set out at length (and which appears in *Capital* I, his notebooks of the 1860s, and *Capital* III) to the effect that the rate of profit must tend to fall due to an increase in the organic composition of capital(see Digression 12.1), an increase itself driven by the search for maximum profit on the part of capitalists. The second is a brief coda to the main argument (appearing towards the end of Chapter XV of Capital, III) but worthy of mention in its

Digression 12.1 The composition of capital

Marx divided capital into two components: v, the sum of money used to pay wages, and c, the sum of money used to purchase raw materials and machinery. He called v variable capital and c constant capital, since, in the labour theory of value, only labour creates new value, and only the employment of labour can produce surplus value s.

The ratio c/v he termed the organic composition of capital, as it reflected the ratio of inanimate capital to human labour.

own right in view of its relevance to later discussions of the TRPF. We call it the “micro-macro bridge”. The third element is what Marx calls “absolute overproduction of capital” which we deal with in our discussion of the demographic constraints on the rate of profit.

We begin by sketching Marx’s main argument, for reference. The argument may be expressed thus:

(1) It is an inherent, intrinsic feature of capitalism that capitalists are driven to seek maximum profits.

(2) While profits can be gained in many ways, the most fundamental means of augmenting profit in developed capitalism is via increases in the productivity of labour.

(3) The enhancement of the productivity of labour involves workers working with an increased ”mass” of machinery or means of labour, and working up a larger quantity of materials per unit time.

(4) Although the value of the means of labour, materials, etc (in Marx’s terminology, constant capital) will not generally increase in full proportion to the ”mass”, it will nonetheless increase, and faster than the variable capital. The organic composition of capital tends to rise.

(5) A rise in the organic composition of capital lowers the rate of profit, other things equal.

(6) Other things cannot be expected to remain equal. The same increase in the productivity of labour, driven by profit-seeking, that expresses itself in a rising organic composition, also expresses itself in a rise in the rate of exploitation, which by itself raises the rate of profit.

(7) Nonetheless, as a long-run tendency, the increase in organic composition must outweigh the increase in the rate of exploitation, in its effects on the rate of profit.

Most critics of Marx's main argument have not questioned points 1 or 2 above, and neither will we.

Point 3 is perhaps more questionable. Certainly there are many examples of technological change that conform to this pattern: the switch from fluvial transport to railways, or from hand looms to machine ones, but there are also counterexamples. Sometimes, as for instance in the move from metal casting to plastic moulding for many uses, the more advanced process accomplishes its results more cheaply while deploying a lesser "mass" of means of production. But let's accept point 3 as broadly correct, at least for the sake of argument. Similarly for point 4: this may not always be true, but it is at least plausible and we will not question it here.

Points 5 and 6 are certainly correct within Marx's conceptual framework. That leaves point 7, the primary locus of controversy. To expose the issue here, it may be useful to write down the relevant equations. Marx's rate of profit (here denoted by r) is the ratio of surplus value (s) to the sum of constant capital (c) and variable capital (v):

$$r = \frac{s}{c+v} \tag{12.1}$$

The organic composition of capital is the ratio c/v and the rate of exploitation is the ratio s/v. We can write

$$r = \frac{s/v}{(c/v)+1} \tag{12.2}$$

which makes it plain that an increase in c/v lowers the rate of profit and an increase in s/v raises it. If the pursuit of profit (via the pursuit of higher labour productivity) has the effect of raising both c/v and s/v, does that not leave the overall effect on the rate of profit indeterminate?

Neither s/v nor c/v has any obvious theoretical upper bound. Why does Marx talk in terms of a basic tendency for the rate of profit to fall due to rising c/v, and treat rising s/v as merely an ”offsetting factor”?

Why not a tendency for the rate of profit to rise due to rising s/v, with the increase in c/v treated as an offsetting factor? (Or no "basic tendency" at all, just an indeterminate outcome.)

Marx clearly had an ideological investment in the idea that the falling rate of profit was primary. This proposition licensed the conclusion that "the real barrier of capitalist production is capital itself" (Capital, III, p. 248).

The very process that constituted capitalism's historical "justification"—namely, its development of the productivity of social labour to an unprecedented level—was at the same time the source of a falling rate of profit, which places a roadblock in the way of further development.

If the only reason Marx had for asserting the primacy of the tendency of the rate of profit to fall was that it fit well with his ideological agenda, one could accuse him of intellectual dishonesty. That is too harsh. It seems clear that he had a strong theoretical hunch or intuition that the rise in c/v must outweigh the rise in s/v. A further manipulation of the rate of profit may help here:

$$r = \frac{s/(s+v)}{c/(s+v)+1} \tag{12.3}$$

The ratio $s/(s+v)$ is not exactly Marx's rate of surplus value, but it is a closely related magnitude with an upper bound of 1.0, namely the fraction of the total social working time during which workers perform surplus labour, or generate profits for their employers. Similarly, $c/(s+v)$ is not exactly Marx's organic composition, but it is a closely related quantity which seems to have no upper bound, namely the ratio of the value of constant capital to the total "living labour".

Looking at this variant of the rate of profit equation it becomes easier to share Marx's intuition. Suppose the ratio $s/(s+v)$ is driven to its maximum (wages are effectively zero; the workers "live on air", as Marx puts it). In that case any rise in the ratio of constant capital to current labour is bound to lower the rate of profit. It then seems plausible that as $s/(s+v)$ gets closer to 1.0 it will become increasingly difficult to find an offset on this account for an ongoing rise in $c/(s+v)$, or in other words a rising rate of exploitation can't keep capitalism out of trouble for ever.

Our own argument in Chapter 11.1 can be summarized as follows. The essential signature of capital as a form of information is the process $M \rightarrow C \rightarrow M'$ in which money expands exponentially. We have argued that this was inhibited so long as the technology of record supporting money was the use of gold or silver coin. This constraint was removed through the development of the banking system in which the technology of record was first replaced by paper and ink and then by computer disks. But a complex social phenomenon like profit has multiple levels of causality. The ability of the monetary technology to support it is only one of these. One can look at the causes of profit in several ways:

(1) One can look at it from the standpoint of the social architecture of capitalism as we do in Chapter 9. In this case the occurrence of profit is seen as being caused by the property relations according to which the product belongs to capitalists whose workers have no property claim on it. The use of computer simulation indicates that these assumptions alone will suffice to generate realistic functional forms for the structure of incomes in society.

(2) One can look at it from the standpoint of monetary technology as we did in Chapter 11.1.

(3) One can look at it from the standpoint of production technology and the extent to which these constrain profits. This is the approach that was taken by economists like Sraffa (1960), Okishio (1961) or more recently by Roemer (1982).

(4) One can look at it as Marx did Marx (1954), from the standpoint of time, looking at how the working day of a labourer could be divided into two parts. In the first part the labourers generated the value to pay their wages. In the second part they generated surplus value which accrues as profit to their employer. We will argue below that this approach passes over, via dimensional analysis, to an examination of the role of demography in profit rates.

(5) One can look at it from the standpoint of macro-economic patterns of expenditure as was pioneered by the economist Kalecki Kalecki (1954).

12.1 INPUT/OUTPUT CONSTRAINTS

In Section 7.3 we introduced the idea of an input–output table. We brought this in to explain how one could estimate labour values. In 1960 Sraffa showed how a sufficiently detailed input output table could be seen as determining the entire price and profit structure of an economy, provided that one crucial simplifying assumption is made.

He starts out by considering a simple self-reproducing economy producing wheat and iron. Each of these goods are used both for the sustenance of the workers and as inputs to the agricultural and industrial production processes. 280 quarters of wheat and 12 tons of iron are used to produce

400 quarters of wheat; whilst 120 quarters of wheat and 8 tons of iron go to produce 20 tons of iron. In schematic form we have

Input		Output
280 qr. wheat + 12 t. iron	→	400 qr. wheat
120 qr. wheat + 8 t. iron	→	20 t. iron

All outputs are consumed either as wages or as means of production. Sraffa goes on to argue that there is a unique set of prices which, when the wheat and iron are sold will ensure that each industry can buy enough inputs to continue producing at the current scale. In this case the net output of industry 1, which was 120 quarters of wheat must exchange for the net output of industry 2, namely 12 tons of iron. This establishes an exchange rate of 10 quarters of wheat to one ton of iron.

With two commodities there is only one exchange ratio. With n different commodities there would be $n-1$ exchange ratios. With n commodities we would have n input output equations.

$$A_a p_a + B_a P_b + .. + N_a p_n = A p_a$$
$$A_b p_a + B_b P_b + .. + N_b p_n = B p_b$$
....
$$A_n p_a + B_n P_b + .. + N_n p_n = N p_n$$

Where A means the total output of commodity A, B_a means the amount of commodity B used up in industry A, and p_a is the price of commodity A. He then assumes that one of the commodities is used as the standard of value—suppose this is gold. The price of gold in terms of itself is obviously unity, so we are left with $n-1$ unknowns. It might appear that we have more equations than unknowns and thus run the risk of having no solution, but it turns out that since the total quantities of each good occur twice—once as inputs and once as outputs, so any one of the equations can be inferred from the sum of the others. It follows that we have only $n-1$ independent equations, and thus the set of prices must be unique.

Next Sraffa went on to consider what will happen if there is a surplus product. The immediate effect of this is to make the equations independent of one another, since it is no longer the case that the input and output totals are equal. In general

$$\sum_{i=a}^{n} A_i \leq A \tag{12.4}$$

This would appear to make the prices indeterminate. To compensate, he introduces a new constraint, assuming that all industries earn the same rate of profit r. This allows him to obtain a new system of price equations:

$$\begin{array}{rcl}
(A_a p_a + B_a P_b + \cdots + N_a p_n)(1+r) & = & A p_a \\
(A_b p_a + B_b P_b + \cdots + N_b p_n)(1+r) & = & B p_b \\
\ldots & & \ldots \\
(A_n p_a + B_n P_b + \cdots + N_n p_n)(1+r) & = & N p_n
\end{array}$$

These equations simultaneously determine all prices and the rate of profit. It is relatively easy to show that if the productivity in any industry rises, so that either its output goes up for the same inputs or it uses less inputs for the same outputs, this will lead to rise in the rate of profit. Given the assumption of an equal rate of profit, any technically advantageous invention in a given sector will raise the rate of profit for the economy as a whole. But Sraffa shows that this only holds if every output is also used as an input by other industries. Luxury goods industries, whose output does not enter back into production are different. He divided industries into two sectors:

(1) The basic sector. This is made up of those industries whose output is a direct or indirect input to every other industry.
(2) The non-basic sector. This is made up either of luxury goods, or of goods that are only used as inputs in other non-basic industries.

An improvement in productivity in the basic sector will raise the rate of profit. An improvement in a non-basic sector will leave the rate of profit unchanged. If the manufacture of bombs becomes more efficient, for example, bombs get cheaper but here will be no knock-on effect to raise the general productivity of the economy.

Up to this point Sraffa has treated the goods consumed by workers as part of the necessary inputs to a production process. His iron industry used up wheat to feed its workers for instance. Once one recognizes that workers are paid money wages rather than getting paid in kind, there is an additional variable to deal with: the wage rate w. To handle this he extends his input output table to include labour inputs as another column:

$$\begin{array}{rcl}
(A_a p_a + B_a P_b + \cdots . + N_a p_n)(1+r) + L_a w & = & A p_a \\
(A_b p_a + B_b P_b + \cdots + N_b p_n)(1+r) + L_b w & = & B p_b \\
\ldots & & \ldots
\end{array}$$

$$(A_n p_a + B_n P_b + \cdots + N_n p_n)(1+r) + L_n w = N p_n$$

Table 12.1: A physical input/output table of economy with a surplus.

	iron	corn	labour	output	surplus
iron	440	1100	110	825	185
corn	100	500	50	2250	550
silk	100	100	20	1000	1000
totals	640	1700	180		

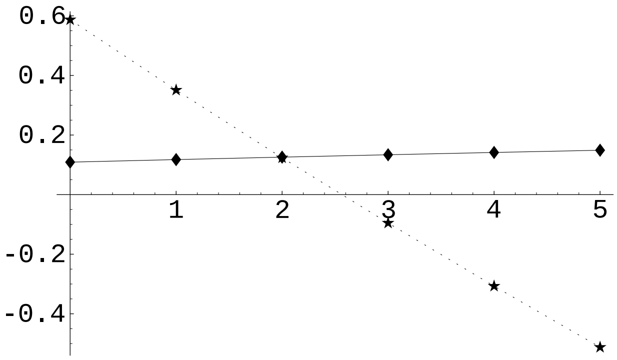


Figure 12.1: Plot of how the rate of profit (dotted line) falls as the wage rises, with wages being expressed in tons of iron. The solid line shows the change in the price of corn expressed in tons of iron as the rate of profit and wage rate change. The data for the graph is drawn from the model economy shown in Table 12.1.

This gives him n equations and $n+1$ variables, which implies that the system has one degree of freedom. If one fixes the wage rate, the system becomes determined.

Table 12.1 shows an example of of an economy with a surplus of the sort described by Sraffa. Sraffa was able to show that there was an inverse relationship between profits and wages. As wages rose profits would fall as shown in Figure 12.1. The rate of profit falls as a straight line or linear function of the wage rate. At the point where wages were high enough for workers to purchase the whole net product, then profits would be zero. This by itself is hardly surprising. More interesting is the result he obtained for

Table 12.2: Example of a Sraffian Standard System. Note that the iron/corn ratios in both input and output are $\frac{1}{3}$, and the the ratio of the total output to the total input is $\frac{3}{2}$. This gives an expansion rate R of 0.5 or profit rate of 50%.

	iron	corn	labour	output	R
iron	400	1000	100	750	0.5
corn	100	500	50	2250	
totals	500	1500	150		

the maximal rate of profit - the rate of profit that would obtain were wages to fall to zero.

To arrive at the maximal rate of profit he introduces the notion of the Standard System. The Standard System is made up of part of the output of each the industries in the basic sector. In it, the industries of the basic sector are scaled in such a way as to ensure that the ratios in which outputs are produced is the same as the ratios in which the inputs are used. He termed this the Standard Ratio. Sraffa wrote that every economy contains such a standard system, which can be discovered by:

(1) Discarding all non-basic industries.

(2) Scaling back those basic industries whose share of the output mix is excessive compared to their share of the input mix.

Table 12.2 shows the result of applying this rule to the economy introduced in Table 12.1. We have first discarded the silk industry as non basic. Then, observing that the ratio of iron to corn in the output was $\frac{825}{2250} = \frac{11}{30}$ but the ratio of iron to corn in the input of the basic sector was $\frac{540}{1600} = \frac{11}{32} < \frac{11}{30}$, we scale back the iron industry until the iron/corn ratios are equal in both the input and the output at $\frac{1}{3}$ giving the Standard System shown in Table 12.2.

In the Standard System we can express the maximum profit rate R in terms of the physical expansion rate of the economy. Recall that commodities occur in the input vector in the same proportions as they occur in the output vector. Thus whatever the relative prices of different commodities the ratio of the total money value of the collection of input commodities to the total value of the output will be $(1+R)$: the same as the physical expansion rate if all the surplus were re-invested. Note that the value of $R = 0.5$ obtained in Table 12.2 is not exactly the same maximal rate of profit as we observed in Figure 12.1. This is because the profit rate in that Figure was expressed in terms of iron as the numeraire. Since the relative prices of commodities changes with the rate of profit (see the corn price in Figure 12.1) no single commodity can act as a reliable numeraire for measuring prices or profit rates. Sraffa showed that the only reliable numeraire would be to use a bundle of basic commodities, mixed in the Standard Ratio, as the numeraire. This weighting ratio will precisely compensate for the fluctuations in relative commodity values that occur as the rate of profit changes. The maximal rate of profit when measured in this weighted bundle is then the physical expansion rate R.

Sraffa's analysis has a number of very interesting implications, but it also suffers from some weaknesses. An initial conclusion that other economists drew, was that Sraffa had shown that the labour theory of value espoused by the classical economists was redundant Steedman (1981). Sraffa had been able to derive all prices and the rate of profit in an economy from the technology matrix and the wage rate. Although Sraffa discussed the feasibility of deriving labour values from the technology matrix, his price theory did not rely on these. Sraffa's assumption of equal rates of profit in every industry amounted to assuming that labour inputs had no independent causal effect on values. But this is just an assumption. One should see Sraffian price theory as being conditional on this assumption. It amounts to saying, if we were to assume an equal rate of profit across the economy, what conclusions could we draw?

12.1.1 Non equalization of profit rates

But the equalization of profit rates across industries is just a simplifying hypothesis. It should not be assume that it is a realistic hypothesis. Farjoun and Machover Farjoun and Machover (1983) showed that if one drops this assumption, prices will tend to follow the predictions of the classical labour theory of value. Empirical studies have shown that whilst Sraffa's model provides very good predictions of actual prices, it is not significantly better than the classical labour theory of value in this respect Shaikh (1998). The fact that profit rates are far from equal across industries Cockshott and Cottrell (1998) is probably the reason for the predictive parity of the two

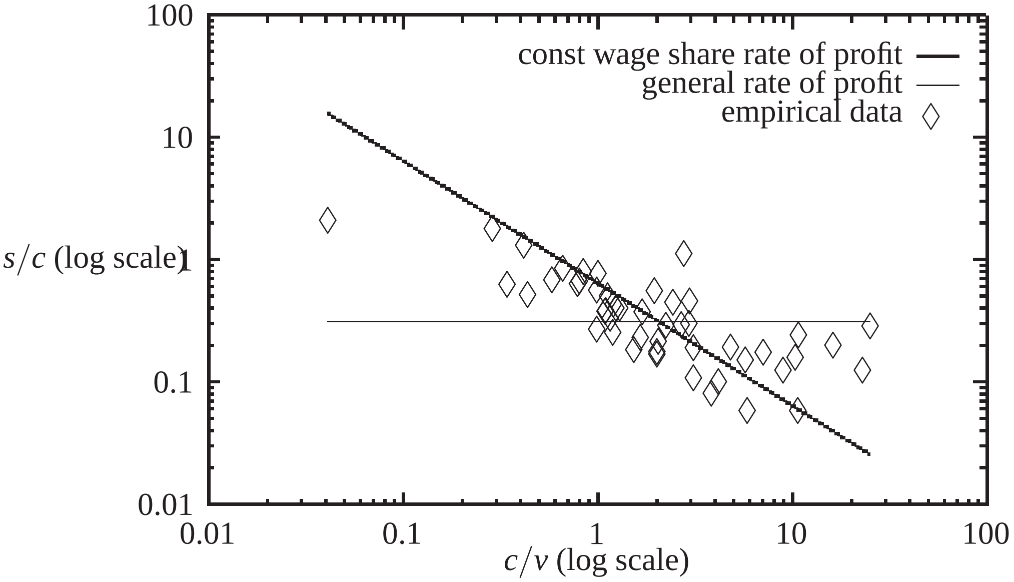


Figure 12.2: Graph of relation between profit rates and organic composition using buildings and structures as the estimate of constant capital, Cockshott and Cottrell (2003).

theories. Using capital stock data from the Bureau of Economic Affairs for the USA Cockshott and Cottrell Cockshott and Cottrell (2003) examined how the profit rate of US industries depended on their organic composition. In computing the organic compositions by industry it was necessary to aggregate some of the industrial categories in the I/O tables as the capital stock figures were not so broken down into such fine categories. But it was found that the results indicate, that any tendency toward profit rate equalization is very weak, and that the effects of the raw labour theory of value predominated. If one defined the constant capital stock for the US using figures for industry by industry stocks of buildings and structures, then organic composition was negatively correlated with profit rates for the US. This is illustrated in Table 12.3.

The consequences of this are indicated in Figure 12.2, which shows three sets of points:

(1) the observed rate of profit, measured as s/c,

Table 12.3: Relation between profit rates and organic composition of United States industries, using buildings and structures as the estimate of constant capital.

	$\frac{s}{c}$	$\frac{c}{v}$	$\frac{s}{v}$
Mean	0.310	9.36	1.178191
Standard Deviation	0.249	9.57	0.162321
Coefficient of variation	0.802	1.02	0.14
	s/c and c/v	s/c and v/c	-
	(weighted by c)	(weighted by c)	-
Correlation coefficient	-0.306	0.685	-

(2) the rate of profit that would be predicted on the basis of assuming a constant wage share, where it would be given by $\frac{vs'}{c}$ where s' is the mean rate of exploitation in the economy,

(3) the rate of profit that would be predicted by volume III of *Capital* or any other variant of transformed values (mean s/c).

It can be seen that the observed rates of profit fall close to the rates that would be predicted by the “Volume 1” theory. The exception is for a few industries with unusually high organic compositions > 10. But what are these industries?

At an organic compition of 23.15, one has the electricity supply utilities with a rate of profit half way between that predicted by the simple labour theory of value and that predicted by the price of production theory. Then at an organic composition of 16.4, one finds the crude petroleum and natural gas industry, with a rate of profit substantially in excess of that predicted by the labour theory of value, and approximating much more closely to that predicted by an equalization of the rate of profit. But an industry like this, would, on the basis of the Ricardian theory of differential rent, be expected to sell its product above its mean value, and hence report above average profits. In a similar position we find the oil refining industry with an organic composition of 9.4. Both oil production and oil refining have similar rates of profit, at 31% and 32%. Since the industry is vertically integrated, this would indicate that the oil monopolies chose to report their super profits as earned pro-rata to capital employed in primary and secondary production. In both cases, however, the super profit can be explained by differential rent.

Next one comes to the gas utilities with a rate of profit of 20% on an organic composition of 10.4. The labour theory of value would have predicted 7% and the production price theory 32%. But like the electricity utilities, these industries are regulated and the assumptions built into the regulatory system include that the utilities should earn an average rate of profit.

The conclusion that one has to draw from this is that the assumption of equal rates of profit by the Sraffian model are a serious over-simplification.

12.1.2 Technical change and the rate of profit

Sraffa's conclusions regarding the determinants of the average level of profit still stand. He shows that profit levels can be seen as deriving from two types of cause—overall technical productivity which sets the maximal profit rate R, and the struggle between labour and capital over the wage rate, which sets the actual average rate of profit r. In particular, he has shown that technology advances can only raise profit rates if they occur within the basic sector. Only those innovations whose product enters directly or indirectly into the production of every other commodity, can raise the general rate of profit. In his first examples Sraffa treated the real wage consumed by workers as part of the necessary inputs to the production process, since in the absence of such consumption they would not survive. He then says that in practice wages are made up of a necessary component required to ensure survival, and a surplus component over which capital and labour content. If one treats the necessary component of the real wage as part of the production inputs to all branches of production, then the definition of the basic sector becomes more general. It can now be defined as all those industries whose output is directly or indirectly necessary to reproduce the working population.

From this concept of the basic sector and Standard System it is easy to see why a technical change which increases the rate of profit in a single industry above the industry average will tend to raise the overall rate of profit in the economy as originally argued by Okishio (1961). Okishio made the assumption of a fixed real wage which is equivalent in Sraffa's terms to assuming a zero surplus component of the wage, and including the necessary the real wage be included as part of production inputs. Under these assumptions Okishio's rate of profit is equivalent to Sraffa's R. Suppose that a change occurs in a single industry. If the change is to be profitable to the individual firm it must involve either a reduction of at least one input for unchanged output, or must increase outputs with the same inputs. Either of these eventualities will reduce the ratio of inputs required to produce the output of the basic sector, and thus increase R.

12.1.3 Computers and the productivity paradox

Technology advances in non-basic sectors like banking, the manufacture of executive jets, or warship construction will not affect the average rate of profit. This may have relevance to the much discussed 'productivity paradox' of computer technology.[1] The paradox stems from the fact that economists have had the greatest of difficulty in detecting any significant increase in economic productivity stemming from the use of computers. The discussion of the productivity paradox has taken place by economists working within the framework of neo-classical economic theory. This framework, which Sraffa was criticizing in his work, assumes that output is determined by an exponential production function of the general form:

$$Y = aL^b.K_1^b.K_2^c \tag{12.5}$$

where Y is the money value of output, $a, b, c, \ldots$ are constants, K_1 is the stock of computer capital goods, K_2 the stock of non-computer capital goods, and L is the labour input. This model differs from Sraffa's in that it is non-linear and causality operates in the reverse direction. Sraffa says that production of 1 ton of iron *uses up* 0.4 ton of iron and 6 qrtr wheat. The neo-classical model says that if we *put in* quantities K_1, K_2, L of inputs, then we will produce Y of output. Perhaps most significantly, the Sraffian model measures all inputs and outputs in physical terms whereas the neo-classical model measures them in money terms. From the Sraffian point of view the measurement of capital in money is a serious flaw since the valuation of commodities depends upon the distribution of income between labour and capital (Foley (2003)). One can thus not hope to measure the productivity of aggregations of capital goods since the valuation of these aggregations is itself a function of the class distribution of income.

From a classical standpoint the notion of productivity measured in money terms was ill-defined. The only context in which one could define productivity was as the inverse of labour values, an increase in productivity was then equivalent to a fall in the labour required to make goods. Sraffa added to this concept the idea of the productivity of the basic sector measured in terms

---

[1] See Brynjolfsson (1993) for a review of this.

of its own inputs. One could in principle measure R for different years and see if it has gone up after the introduction of computer technology. Since there were many other technical changes at the same time, it would be hard to say whether such an increase in R might have stemmed from computer technology or from other innovations. Beyond this point though, the concept of the basic sector may provide another reason why productivity gains due to computers are so hard to discover.

Since computers are largely used in non-basic sectors, Sraffian theory predicts that they will leave R unchanged.

12.2 ROEMER AND EXPLOITATION

Classical Marxism conceives of exploitation as the extraction of surplus labour by an exploiting class from the exploited. By some mechanism – which varies from one mode of production to another – the exploiting class is able to compel the exploited class to perform more labour than is required for the maintenance of the latter. The fruits of this surplus labour are available to the exploiters, to support their consumption and/or to augment their wealth. Under capitalism, the extraction of surplus labour proceeds via the exchange of labour- power for wages: the worker receives a wage equal (on average) to the cost of (re-)production of labour-power, but once he has purchased the worker's labour-power, the capitalist is able to make the worker perform more labour than is needed to reproduce the wage.

The underlying precondition for this mode of exploitation is the capitalist pattern of ownership of the means of production. The compulsion of the workers to submit to exploitation via the wage system stems from their propertyless status: possessing no means of production, they are unable to secure their subsistence outside of the wage-contract. At the other pole, it is their exclusive possession of society's means of production which permits the capitalists to set the terms on which the workers get access to those means.

The assumption of classical Marxism is that there exists a strict correlation between the capitalist/worker distinction and the exploiter/exploited distinction: capitalists are exploiters and wage- workers are exploited. The petty bourgeoisie – agents who both own means of production and work on their own account – may be harder to classify in terms of surplus labour accounts, but the situation is clear with regard to the ’basic’ classes of capitalism.

Against this background, Roemer’s critique involves unpicking the relationship between unequal ownership of the means of production, on the one hand, and the extraction of surplus labour on the other. According to Roemer, the proper object of ethical critique on the part of socialists is inequality in the ownership of productive assets *rather than* exploitation in the form of extraction of surplus labour.

This argument is set out clearly in Roemer (1986) – all page references below are to this piece. The target: A theory of exploitation which conceives “goods as vessels of labour, and calculates labour accounts for people by comparing the ’live’ labour they expend in production with the ’dead’ labour they get back in the vessels” (p. 261). The conclusion: “[T]here is, in general, no reason to be interested in exploitation theory, that is, in tallying the surplus value accounts of labour performed versus labour commanded in goods purchased” (262).

Roemer identifies four possible uses or justifications for a theory of exploitation as extraction of surplus labour, before attempting to cut the ground from under each one. We shall concentrate on two of these uses: (a) exploitation of workers might provide an explanation of profits; and (b) exploitation may be seen as a measure and consequence of the underlying inequality in the ownership of the means of production.

12.2.1 Does exploitation explain profits?

Granting Morishima’s formal ’Fundamental Marxian Theorem’, according to which the exploitation of labour is a necessary condition for positive profits under capitalism, Roemer nonetheless claims that it is erroneous to infer from this theorem that the exploitation of labour serves to *explain* profits.

> For, as many writers have now observed, every commodity (not just labour-power) is exploited under capitalism. Oil, for example, can be chosen to be the value numeraire, and embodied oil values of all commodities can be calculated. One can prove that profits are positive if and only if oil is exploited, in the sense that the amount of oil embodied in producing one unit of oil is less than one unit of oil – so oil gives up into production more than it requires back. Thus the exploitation of labour is not the explanation for profits and accumulation any more than is the exploitation of oil or corn. The motivation

> for the privileged choice of labour as the exploitation numeraire must lie elsewhere... (265-6)

First response: The ’exploitability’ of oil is a consequence of technology (i.e. it is a technological datum that a barrel of oil can be extracted at a total oil-cost of less than one barrel). This is not so for labour: the ’exploitability’ of labour depends in part on the consumption bundle, which is socially determined. Here is a real economic difference, which gives a special role to the exploitability of labour in explaining the existence of profits. By raising the price of labour-power sufficiently, workers could, in principle, render themselves ’unexploitable’ – which observation points us towards the socio- economic factors that *prevent* the workers from doing so: these factors explain the possibility of profit.

Second, consider the whole list of commodities which capitalists *might* choose to produce. Some of the items on this list may turn out, given current technology, to be inherently ’unexploitable’ (e.g. currently it takes a larger energy input to produce a given energy ouput from a nuclear fusion reactor). This is not a problem: capitalists simply don’t try to produce them (there are no commercial fusion reactors for electricity generation). From this perspective, the ’exploitability’ of all of the commodities actually produced in capitalist economies is not an *explanation* of profits. Rather, the need for profitability explains why only ’exploitable commodities’ get produced. labour’s special role – in this context – consists in the fact that its use is *not* optional (short of a science-fiction world with robots of the type described by Asimov). Labour is not employed because it ‘happens to be exploitable’, but rather it is the exploitability of (non-optional) labour that explains the possibility of profit.

As noted above, classical Marxism involves the assumption that all capitalists and no workers are exploiters, while all workers and no capitalists are exploited. Roemer calls this the ’Class-Exploitation Correspondence Principle’ or CECP. Question: Can it be derived as a formal theorem from basic and self-evident axioms?

Roemer’s procedure is to employ the methods of modern neoclassical economics to reconstruct Marxism. He starts out from the postulate of rational self-interested agents who attempt to maximize their utility given certain constraints (including their initial ’endowments’ of various kinds of assets). On this view, social classes are not ’basic’ theoretical objects; rather the task is to show how rational agents with differing endowments will choose to enter certain class positions (e.g. to sell their labour-power, to work on their own account, to hire others). Marxist propositions that can be supported in this way are regarded as definitive, while those that cannot be so derived are rejected.

In some of his earlier writings Roemer provided formal proofs of the validity of the CECP – hence ’confirming’ the classical marxist identification of capitalists as exploiters and workers as exploited – under various assumptions regarding production functions and preferences. But in his 1986 article he criticizes his own past work as reliant on an overly restrictive account of agents’ preferences. He then sets up an example (p. 274 ff.) where the pattern of preferences leads to a wealth-elastic supply of labour (i.e. the poor don’t like to work, but the rich do), and in this context shows that the flow of surplus labour may end up going the ’wrong way’ (from rich to poor). The poor man, who is averse to labour and prefers to make a living by lending out his meager capital, ends up ’exploiting’ the workaholic rich man. Yet we should still want to say it is the poor man, with the much smaller initial endowment of productive assets, who is subject to injustice.

But if the exploitation of the rich by the poor is theoretically possible – and hence the Class-Exploitation Correspondence Principle breaks down – this means that the concept of exploitation in terms of flows of labour time should be abandoned.

The example which leads Roemer to this conclusion may be mathematically correct, but it makes no contact with social reality. Much more than just an extended notion of preferences is required to make relevant the putative exploitation of the rich by the poor. Under current circumstances, for a ’poor’ person to make a (meager) living as a lender to the rich he would need to have a capital of perhaps $200,000. But having that much money would place our pauper in something like the richest 10 per cent of the population!

Roemer himself seems to recognize that there may be a practical cost to following through on his theoretical admonition, when he admits that we still need some index of the unjust income flows which arise from an unjust distribution of stocks: ”In cases where exploitation does render the correct judgment on the injustice of flows, then perhaps the degree or rate of exploitation is useful in assessing the degree of injustice in the flow” (277). Furthermore, such cases are admitted to be preponderant: ”as an empirical statement, surplus value accounts mirror inequality in ownership of the means of production pretty well...”. (Roemer’s candor is to be applauded: not all writers who have constructed tricky ’counter-examples’ to the Labour Theory of Value are so forthcoming on the empirical status of their constructions.)

It would appear, then, that this critique does not really have much sting. Moreover, when it comes to formulating socialist objectives Roemer’s position has a serious weakness. As Cottrell and Cockshott (1992) argued, ’ending exploitation’ is a clearly-defined goal. On the other hand, ’achieving a just or equal distribution of the means of production’ (Roemer’s preferred expression) is much less clear. This obviously *cannot* mean, in the modern context, giving every worker his or her per capita share of the total stock of means of production. The notion of a ’just distribution of the means of production’ is very problematic. Socialists aim for the employment of the stock of means of production for the benefit of all working people and their dependents: it is not helpful to conceive of this as a ’distribution’ of stocks across agents; rather it is a pattern of democratically-controlled, socially-planned, allocation and use. The call for a ’fair’ distribution of the means of production may be applicable to the struggle of landless peasants against a landlord class – for the redistribution of land – but it is not applicable to the struggle of wage-workers against a capitalist class.

12.3 MACROECONMIC CONSTRAINTS

Sraffa approached profit from the standpoint of the capacity of the economy to produce a physical surplus. An alternative approach, based on the sort of analysis presented in Chapter 11.1 looks at the way the accounting identities imposed by commodity sales, constrain the overall level of profit. Such a model for the determination of profits was given by Kalecki (1954). He showed that in an abstract capitalist economy with only two classes, no government and subsistence wages, profits are jointly determined by the levels of capitalist consumption and investment. We give an algebraic demonstration of this in the following section, but the basic argument is of disarming simplicity.

In the absence of workers’ savings and taxation the capitalist class will exactly cover its wage bill by selling consumer goods to workers. Since the only other sales are of investment goods and capitalists’ consumer goods, it follows that profits, as the only other form of revenue, must be derived from these sales. Any money that capitalists spend on commodities other than labour power is a revenue for another capitalist. Since wages have been accounted for it follows that investment and capitalist consumption must instantaneously determine gross profits. Taken collectively capitalists' expenditures determine their own revenues. Whilst the existence of credit allows for wide fluctuations of investment and capitalist consumption independent of past profits, current profits are absolutely determined by investment and capitalist consumption. As collective owners of the means of production, capitalists finance their own appropriation of the surplus product.

They have the ability to appropriate the entire surplus product, but this ability belongs to capitalists as a class. The units of economic calculation are individual enterprises and capitalist households, and there is no reason to suppose that the quantity of commodities appropriated by them will coincide with the available surplus product. An investigation of the determinants of profit will therefore require some conception of the determinants of capitalists' expenditure, particularly investment expenditure. To do this you have to make assumptions about the forms of property in existence and the types of economic calculation accompanying them. In particular this involves assumptions about the existence of financial institutions and about the possibility of substitution between real and fictitious capital assets, and this in turn requires assumptions about the establishment of a rate of interest.

Marx argued that a general theory of interest rates was impossible Marx (1971). It is only because the conditions of production set limits to the rate of profit that it is possible to have a theory of the long run rate of profit that abstracts from specific property forms. Since there is no particular relationship between interest rates and conditions of production, any assumptions that we make about the determinants of interest must be related to the specific institutions that determine the interest rates. In our analysis we assume that interest rates are determined by the economic calculations of bank capital.

12.3.1 Simple Kaleckian model: Direct Personification of the Enterprise, no Rentiers, no Banks

Our starting point is the simple Kalecki model of realization, wherein

$$\text{Profits} = \text{Investment} + \text{Capitalist Consumption}$$

expressed symbolically:

$$P = I + C \tag{12.6}$$

Two points must be noted. It implicitly assumes only two classes, workers and capitalists. It is assumed that subsistence wage rates result in there being no saving by workers, capitalists are sole owners of their enterprises and there is no capital market. Secondly, the order of determination is not what might initially be thought. Common sense tells us that capitalists receive a certain sum of profits, part of which they consume and another part of which they invest. The level of profits seems to determine the level of capitalist consumption and investment. In fact the reverse is the case, investments and capitalist consumption are the determinants of profits, as is explained by Kalecki.'s article 'The Determinants of Profits' in Kalecki (1971). This means that profits are necessarily equal to the real appropriation of value as elements of constant capital or articles of consumption by the owners of capital. Real appropriation determines profits rather than vice-versa. Profit as a monetary expression of real commodity movements is ultimately determined by them.

If we were to include the whole national income in our calculations the following would apply:

$$P + W = I + C + W_c \tag{12.7}$$

where W denotes wages and W_c denotes workers' consumption.

If wages equal workers consumption, 12.7 reduces to 12.6. To obtain the net value product we must subtract consumption of constant capital, i.e., depreciation.

$$nvp = P + W - \Delta \tag{12.8}$$

where nvp is the net value product and Δ denotes depreciation.

If we assume that there are no unproductive workers then variable capital is identical to total wages and the rate of surplus value is given by:

$$s' = \frac{P - \Delta}{W} \tag{12.9}$$

where s' is the rate of surplus value, and the net rate of profit will be given by:

$$p' = \frac{P - \Delta}{K + W.t} \tag{12.10}$$

where

p' = net rate of profit
K = stock of constant capital
t = turnover time of variable capital

Since the turnover time of variable capital is very short, its effect is negligible so that by substituting 12.6 into 12.10 we obtain the net rate of profit as a function of consumption and investment by capitalists.

$$p' = \frac{I + C - \Delta}{K} \tag{12.11}$$

Given the skeletal assumptions we have made about property forms in this model we cannot go on to say anything credible about the determinants of investment. This requires a more elaborate model.

12.4 DEMOGRAPHIC CONSTRAINTS

The whole debate in Okishio and Roemer is cast within the context of a choice over techniques, and whether any rational choice of techniques by a capitalist will result in techniques being chosen that will lower the rate of profit. There are other possible ways of approaching the question though. Instead one could focus on what the rate of profit tells us. It tells us something about the potential rate of expansion of capital stocks. It sets an upper limit on the rate of expansion that can be achieved out of internal funding - the rate of capital growth that will be achieved if all profit is reinvested. From Sraffa's notion of the Standard System, the rate of profit tells us something about the rate of material expansion of the productive base of the economy.

The focus of the analysis should be on how this rate of expansion will change over time if capital actually is reinvested. If we answer this question we can then go on to look at the circumstances under which capital might be reinvested and also look at the consequences of capital not being reinvested.

If we approach the time evolution of the rate of profit from the standpoint of capital accumulation, then the issue becomes simpler. Initially we will assume that all measurements are performed either in labour hours, or what amounts to the same thing in a monetary unit of account whose labour hour equivalent does not change from year to year. Let us further assume that half of all profits are reinvested. Thus a 5% rate of profit will imply a 2.5% growth per annum of the capital stock. Let us also assume at first that the division of value added between wages and profit remains unchanged over time.

This means that total profit per year will be a constant multiple of total wages per year. Under these circumstances it is clear that the rate of profit will fall over time if rate of growth of wage income is less than the rate of profit, and the rate of profit will rise if the rate of growth of wage income is higher than the rate of profit. We then focus on the determinants of the rate of growth of wage income - measured in labour hours per annum. The dimensions give it away, since wage income in these terms corresponds to a number of people - the number of people whose direct and indirect labour supports the employed population. The rate of growth of wage income comes down to the rate of growth of the working population (given the assumption of a constant rate of surplus value). The appropriate focus for analysis of the falling rate of profit is not technological choice but historical demography. This is made even clearer if we take on board Marx's approach which treats wages and profits in terms of time. Profit can be measured as a flow of labour value, in which case its dimension its units are person hours/annum, which in dimensional terms is just persons since the hours/annum just give us a scalar. Thus the annual flow of profit when measured in labour terms corresponds to a certain number of people—the number of people whose direct and indirect output is materialized in the goods purchased out of profits. These people are the surplus working population, the population over and above those who would be required to maintain the employed population at its current standard of life.

The capital stock of a nation is, in these terms, a quantity expressed in millions of person years. The rate of profit is then:

$$r = \frac{\text{Millions workers whose product is bought by profits}}{\text{Millions of worker years represented by the capital stock}}$$

The evolution of r is here seen to depend on how rapidly the capital stock is built up compared to how rapidly the number of surplus workers grows. Once the argument is on this terrain one has to ask what determines the rate of growth of the working population. Two factors are clearly important, the natural rate of population growth and the fraction of the total population that is employed as wage labourers under capitalist relations of production.

Economies undergoing transition from peasant farming to capitalist industry typically have a rapid rate of growth of the working population from

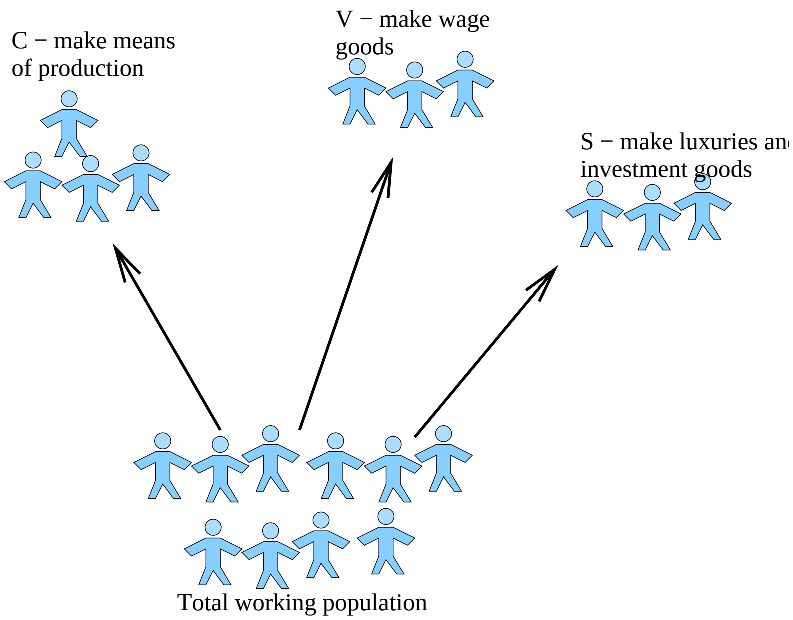


Figure 12.3: The working population can be divided into three groups: V, those whose product is workers' consumer goods; C, those whose product is replacement means of production; and S, whose products are either luxuries or net additional means of production.

.

both factors. The birthrate tends to be high and infant mortality falls during the transition from peasant agriculture to capitalist industrial economy. This gives a rapid rate of natural population increase. At the same time the fraction of the population employed as wage labourers rises to give a high compounded rate of growth of the employed population.

In a mature capitalist economy things are different. Although infant mortality continues to fall, this is offset by a falling birth rate, which in many advanced capitalist economies falls below replacement level. At the same time the share of capitalistically employed wage labour in the population tends either to reach a plateau or even to fall. The result is a relatively stagnant or declining capitalistically employed population.

If we assume that the rate of growth of the employed population is fixed then the effect is that the actual rate of profit tends towards some multiple

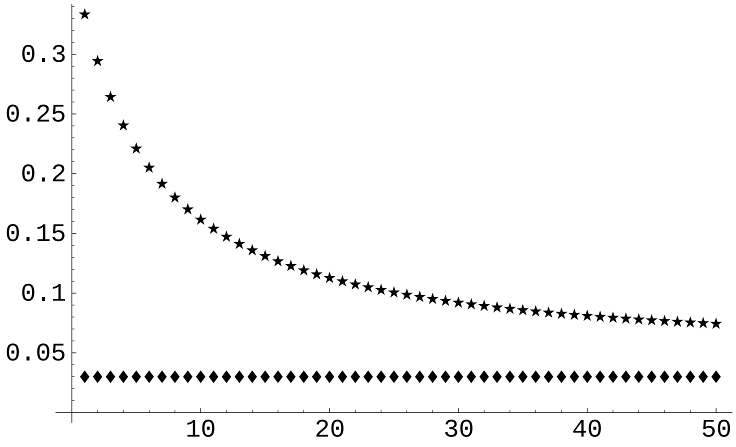


Figure 12.4: Evolution of the profit rate under constant population growth. The upper line is the profit rate, the lower line the rate of population growth. . Years are measured along the horizontal axis.

of the rate of growth of the employed population. In Figure 12.4 we start out with an initial rate of profit of 33% and have the population growing at 3% a year. Half of all profits are reinvested. The rate of profit on capital tends towards 6% as this is the only rate of profit at which the rate of growth of the capital stock will equal the rate of growth of the population. It is the latter that constrains the rate of growth of value production.

If we take a more realistic model as shown in Figure 12.5, where the rate of growth of the population declines with time, then the rate of profit chases the rate of population growth downwards. If the share of accumulation out of profit is α and the rate of population growth is g then the rate of profit will tend towards g/α.

In our examples up to now we have been assuming that real wages are rising over time. This is because we have assumed a constant rate of surplus value of 50%. Because technology and productivity can be assumed to be going up, a wage share of 2/3 of the national income will correspond to a rising real standard of living. Okishio's paper in 1961 assumed that real wages were constant. This corresponds to a gradually increasing profit share in national income. Figure 12.6 models this process. We here assume that technical productivity in the wage goods sector grows at 3% a year, and that

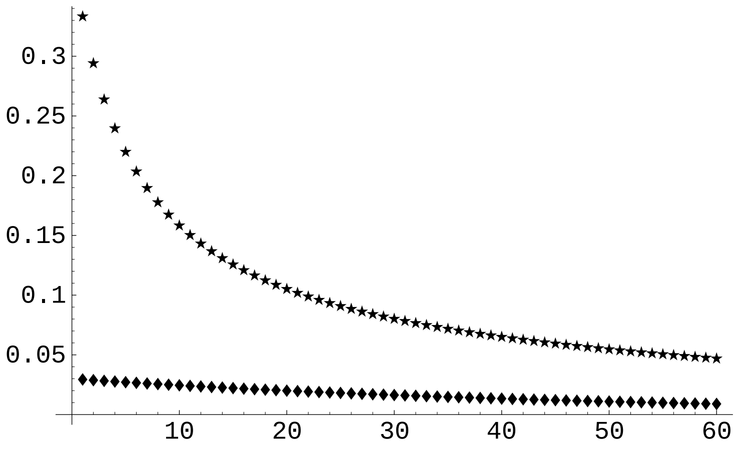


Figure 12.5: Evolution of the profit rate under conditions of declining population growth. The rate of profit declines further than in the case of Figure 12.4

real wages remain constant. Under these circumstances the wage share will fall at 3% a year. Observe that the rate of profit still falls.

Investment in new plant and equipment can be expected to improve production techniques and reduce the prices of capital goods. Under these circumstances the value of the stock of invested capital will depreciate. This will tend to slow the growth of the capital stock. At the same time, it will result in losses on the capital account to firms whose assets have depreciated. If we take these into consideration when calculating profits we find two opposite effects. The depreciation of capital stocks slows down the growth of stocks which tends to mitigate any decline in profit rates. Conversely, the losses on the capital account tend to directly reduce profits. These consequences follow from equation 12.11 summarizing the Kaleckian model of profit causality.

The simulation shown in Figure 12.7 shows what happens when we allow productivity to improve both in the wage goods and capital goods sectors. During years 1–14 and 55–70 there are no improvements in productivity, but over years 15 to 44 labour productivity throughout the economy goes up 5% a year. Real wages are assumed to be held constant.

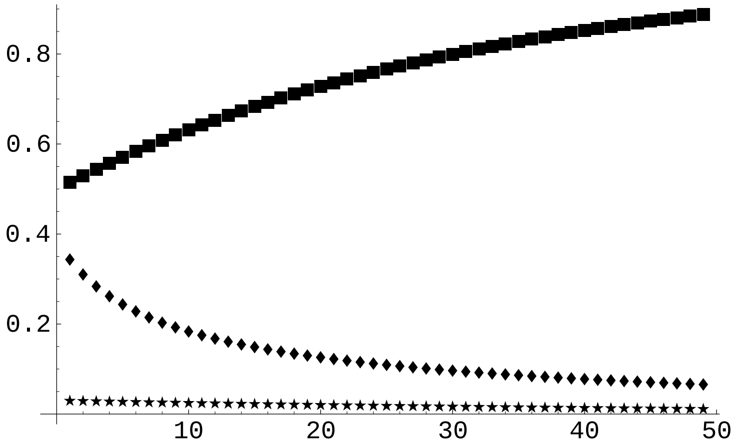


Figure 12.6: Evolution of the profit rate under conditions of declining population growth and constant real wages. The rate of profit still declines. Top curve, the share of profit in national income. Middle curve the rate of profit. Bottom curve, the rate of population growth.

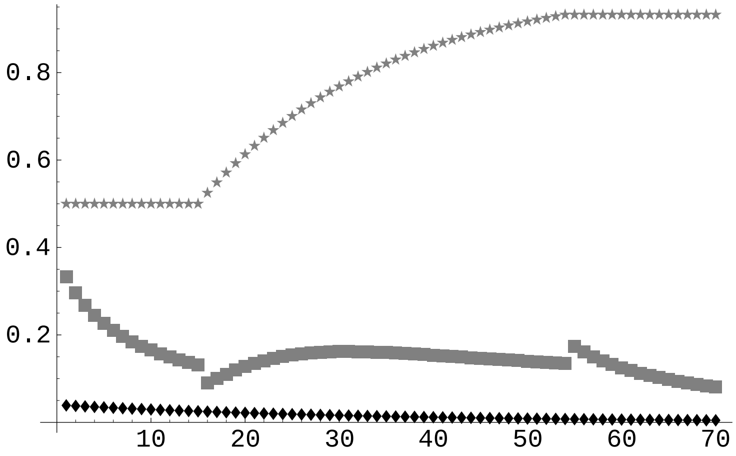


Figure 12.7: Evolution of the profit rate under conditions of declining population growth and technical improvement. From years 15 to 55 there is a 5% improvement in labour productivity. Other years see no improvement. Top curve, the share of profit in national income. Middle curve the rate of profit. Bottom curve, the rate of population growth.

In year 15 the rate of profit drops sharply from 13% to 9% because of the higher rate of depreciation brought about by technical change. But for the following 15 years the rate of profit rises again to a peak of just over 16%. This rise comes about as a result of capital stocks growing more slowly because of depreciation which temporarily lowers the organic composition of capital. The other factor bringing an initial rise in the rate of profit is the rise in the profit share brought about by falling real wages. The rate of profit subsequently settles into a declining trend because of the slowdown in population growth.

What one finds in such simulation studies is that the long-term rate of profit $\mathcal{R}$ tends to:

$$\mathcal{R} = \frac{g+m}{\alpha} \tag{12.12}$$

where g is the rate of growth of the working population and m is the rate of mechanical improvement to labour productity, and α is the share of profit being reinvested. This rate $\mathcal{R}$ is the only one at which the organic composition of capital is stable. If the rate of profit is higher, then the organic composition is driven up, if it is lower, then the organic composition falls. This has the interesting implication that were an economy to have a declining population, which given trends in birth rates is quite plausible for many capitalist nations, then the long term profit rate might be zero or negative. To retain a positive profit rate with a declining workforce, the rate of labour productivity must improve faster than the size of the workforce falls.

12.4.1 Monetary illusions

The argument in Section 12.4 is formulated on the assumption that all accounting of profit and loss is done in value terms. In practice, of course, profits are calculated in terms of money not labour values. The argument assumed that the value of monetary in labour terms did not change over time. If the product of a days labour sold for £1 in 1900, it still sold for £1 in 1910, 1920, 1930..., etc, which is obviously false. Over time the value of money has gone down in two senses:

(1) £1 bought less and less labour as the decades progressed.

(2) The price of many, but not all, commodities tended to rise. Bread in Britain is about 15 times as expensive as it was 40 years ago.

We can measure the purchasing power of money in goods that it buys, or in labour it purchases. In both senses the value of money has fallen. Such inflation makes commerce *appear* more profitable. You have only to hold assets a while then sell, and you make a profit. Millions of houseowners know this. Inflation transfers resources from lenders to borrowers and can also hide a falling real rate of profit. Banks do notice that the money they are being paid back is worth less than when they lent it. They know the difference between real and money profits and compensate by charging more interest.

The same applies to other businesses. Their accountants distinguish between nominal profits arising from inflation and real profits. But how should they make the adjustment? Should they measure the value of money in terms of commodities or in terms of labour?

Official inflation is measured using cost of living indices. These track the price of a shopping basket of typical consumer purchases. The index tells you what wage increases are required to maintain living standards. So the cost of living index is more useful to trades union negotiators than to capitalists. What interests the latter are the prices of labour and raw materials. Adam Smith said that money was the power to command the labour of others. The entrepreneur seeks this command over labour. He wants to 'grow the business', and this growth comes down either to having more employees, or what amounts to the same thing, indirectly employing more people via suppliers and subcontractors. Unless the business grows in these terms, his social position has not improved. From the capitalist standpoint, the value based accounting that we presented above is indeed the most appropriate. It is only when he deflates the monetary profit rate to get the value profit rate that the capitalist can measure the growth of his social power. This accumulation of social power is what capital accumulation is ultimately about. You don't get to be a Bill Gates by growing your wealth only at the rate that industrial productivity rises.

If a capitalist wants to go up in the world, he should watch his value rate of profit. If he does not want his absolute standard of living to decline, he should ensure that his money rate of profit is greater than price inflation.

The equation 12.12 showed how, for a representative capitalist, the growth of his power comes to be constrained by investment (α), population growth (g) and technical progress (m).

We can transform the attractor of the value rate of profit $\mathcal{R}$ into an attractor for the monetary rate of profit $\mathcal{R}_{\mathcal{M}} = \mathcal{R} + I_v$ by adding the rate of inflation in value terms I_v. Now, suppose that the cost of living index, measured against a representative bundle of commodities, is zero, what does this imply for the evolution of the monetary rate of profit?

I_v measures the annual price inflation of a commodity bundle containing 100 hours of labour. As labour productivity rises, this 100 hour commodity bundle will get physically bigger at the rate m: the rate of mechanical improvement to labour productity. If we assume zero rate of price change then obviously $I_v = m$. In the general case : $I_v = m + p'$ where p' is the conventional rate of price inflation as measured by, for instance a cost of living index. This enables us to deduce that the attractor for the monetary rate of profit will be:

$$\mathcal{R}_{\mathcal{M}} = m(1 + \frac{g}{\alpha}) + p' \tag{12.13}$$

Let us look at six possible scenarios to get a feel for the different determinations of the long term value and long term money rates of profit. The scenarious are labeled by historical periods that have the same general features. Remember that in what follows $\mathcal{R}, \mathcal{R}_{\mathcal{M}}$ are not the actual rates of profit, but the limits towards which the rates of profit evolve.

	$\mathcal{R}$	$\mathcal{R}_{\mathcal{M}}$	m	g	α	p'	
1	2.75%	10.26%	2.25%	0.5%	100%	8%	UK 1970
2	2.75%	2.26%	2.25%	0.5%	100%	0%	UK 1970 allowing for inflation
3	15%	1.10%	2%	1%	20%	-1%	UK 1870
4	6.67%	2.90%	3%	-1%	30%	0%	Europe 2020
5	2.86%	2.96%	3%	-1%	70%	0%	Europe 2020 high acc
6	15%	10.50%	10%	5%	100%	0%	China 2000

The first is period of high inflation and high accumulation and slow population growth - for example the UK at the end of the 1960s and early 1970s. The long run money rate of profit $\mathcal{R}_{\mathcal{M}}$ is high, but when this is corrected for inflation in the next row, we see that the long run money rate of profit is very close to m the improvement in mechanical productivity. $\mathcal{R}$, the attractor of the value rate of profit is somewhat higher because of the growth of the working population.

If we go back a century, we have a faster population growth, but a much lower rate of accumulation out of profits, and because of the deflationary effect of the gold standard, slightly declining money prices. The long term money rate of profit would be mainly determined by the rate of deflation and the rate of technical change - it is approximately the rate of technical change less the rate of price decline. The long term value rate of profit is considerably higher than the money rate because

- The value rate is not affected by price deflation.
- Under these circumstances of low accumulation the equilibrium organic composition of capital is also low which keeps up the profit rate.

For a comparison with history look at Figure 14.2.

Scenarios 4 and 5 look at a future European economy with a declining population. If we assume that the Euro is managed to maintain a zero rate of price inflation, and assume a 3% long run growth in labour productivity, then the attractor of the monetary profit rate will also be close to 3.̂ The attractor of the value rate of profit on the other hand will vary inversely with the rate of investment.

Scenario 6 looks at a rapidly emergent capitalist economy like China at the end of the 20th century. Here the population is still growing fast and the rate of investment is high. The importation of advanced technology allows a much more rapid growth of labour productivity than can be attained in a mature capitalist economy. The limit of the value rate of profit is again higher than that of the money rate because of the growth in population.

What do we learn from these examples?

(1) The long term attractor of the money rate of profit will be mainly determined by the rate of technical progress and the rate of inflation.

(2) The long run attractor of the value rate of profit is more complex; being determined by technical progress, population growth, and investment rates. For a positive long run rate of profit, the rate of technical progress must exceed any decline in the population.

Whilst the money rate of profit is the easier for a firm to calculate, it's value rate of profit gives a better social measure of how the firm is doing.

12.5 KALECKIAN CONSTRAINTS ON PROFITS

The demographic arguments in section 12.4 implicitly assume full employment which is, of course, unrealistic. We will now consider a new model in which we take into account how employment may fluctuate and explain accumulation taking into account the rate of interest. We shall:

(1) consider enterprises as abstract juridical personalities rather than individual capitalists;

(2) assume that the entire capitalist class have become rentiers, making capitalist consumption rentiers' consumption;

(3) assume the existence of credit money with banks providing the main financial intermediary, between rentiers and enterprises, but allowing for a marginal market in direct loans from rentiers to enterprises.

Such a set of property forms has never existed in pure form, but corresponds to the hypothetical presuppositions for the complete dominance of finance capital in the specific form of bank capital, but without the merger of bank with industrial capital Steindl (1952). Using this model we can explain some of the weaknesses of the Okishio approach.

In the previous model the sole forms of revenue were wages and profits. Interest now appears as an additional category, so that profit is divided into interest and profit of enterprise Marx (1971).

For profit of enterprise we thus obtain:

$$E = P - R - \Delta \tag{12.14}$$

where

E = profit of enterprise

R = total interest

Now the level of total interest payments is a function of the rate of interest (r) and the outstanding debts of enterprise (D) since we assume enterprises to be the only net debtors in the system. It follows that:

$$R = D.r \tag{12.15}$$

Similarly we may divide the total capital of the enterprises into two parts, one of which belongs to the enterprise, and the other matching the enterprise's outstanding debts to the banks.

$$K = D + H \tag{12.16}$$

where

H = enterprise capital

On the basis of the original model we can now give the determinants of profit of enterprise.

$$E = I + C - D.r - \Delta \tag{12.17}$$

If rentiers consumption is given by C, any income that they get in excess of this must be their savings, or accumulation of money capital. On the other hand, since enterprises are pure juridical personalities they have no personal consumption. Any excess of profit over interest payments constitutes their internal accumulation. This may either take the form of the acquisition of new elements of constant capital (again assuming variable capital to be insignificant in stock terms) or as money capital. Given our assumption that there is no market in loan capital other than through the banking system, this means that individual enterprises must divide its internal accumulation between purchases of elements of constant capital and the accumulation of bank deposits. Purchases of elements of constant capital are the active factor, accumulation of deposits the residual. If investment exceeds profits, then the enterprise has a negative accumulation of money capital. Either it runs down its deposits or it borrows from the bank. Conversely a positive accumulation of money capital may involve either an absolute rise in its deposits with the banks or a fall in its outstanding debt.

The two portions of total accumulation are defined as follows:

$$A_e = E \tag{12.18}$$

$$A_r = D.r - C \tag{12.19}$$

$$A_e + A_r = I - \Delta \tag{12.20}$$

where A_e denotes the accumulation of enterprise capital and A_r denotes the accumulation of rentier capital.

We will assume that the consumption of rentiers is determined both by their income and their money stocks. They are buffered from the immediate effects of fluctuations in their income by their holdings of money capital. In this they are different from workers whose consumption is directly related to their current wages. We can treat rentiers' consumption as being made of two components:

$$C = x.D + y.D.r \tag{12.21}$$

where x and y are constants of value less than 1.

Substituting into 12.17 and 12.18 we can see that

$$A_e = E = I - \Delta + D[x + r(y-1)] \tag{12.22}$$

This shows that accumulation of enterprise capital is a decreasing function of the rate of interest and an increasing function of the coefficients of rentiers' consumption. The higher is rentiers' accumulation the lower is enterprise's own accumulation. Here we see an instance of the contradiction between industrial and rentier interests. It is also clear that if the rate of interest falls sufficiently low (below $\frac{x}{1-y}$ in our example), then it will not cover rentiers' expenditure on consumption, and will lead to a negative level of accumulation by rentiers. Low interest rates accelerate the accumulation of enterprise capital whilst undermining the position of the rentiers. Okishio's argument was that capitalists will not carry out investments in new production technologies if the result of these would reduce the average profit rate in the whole economy. On the other hand we have argued in section 12.4 that net accumulation greater than the rate of growth of the working population will tend to reduce profit rates in the long run.

One conclusion from this might be that our scenarios shown in Figures 12.5 or 12.6 will never occur. There will never be any accumulation faster than the rate of population growth. It is possible that the microeconomics sets limits to the maximal rate of accumulation that macro-economy can exhibit. Another possibility is that some of Okishio's micro-assumptions are invalid and should be dropped. In the end, as with all scientific hypotheses, the criterion has to be their ability to predict what actually happens.

12.6 HISTORICAL TREND DATA

There is evidence that over periods of decades in individual economies the rate of accumulation has exceeded the rate of population growth and that as a consequence the organic composition of capital has risen Michaelson et al. (1995), Edvinsson (2003, 2005). As Table 12.4 shows, during the post-war boom a sustained period of rapid accumulation drove the organic composition up and the rate of profit down. The fall in the rate of profit, was greater than what could be accounted for just by the change in the organic composition. Other factors, such as rising real wages and a rise in the proportion of unproductive workers in the workforce over were also factors holding down profits.

Table 12.4: Development of organic composition c/v, rate of profit $s/(c+v)$ and share of profit being accumulated α in the UK 1948 to 1972. Sources described in Michaelson et al. (1995).

Year	c/v	$s/(c+v)$	α
1948	4.57	3.75	0.34
1950	4.58	3.61	0.68
1952	4.98	5.61	0.36
1954	4.96	4.95	0.47
1956	5.15	4.02	0.72
1958	5.68	3.72	0.83
1960	5.59	4.73	0.72
1962	5.98	3.53	0.94
1964	6.37	4.16	0.89
1966	6.57	2.50	1.41
1968	7.39	2.79	1.29
1970	7.85	1.25	2.62
1972	8.35	1.09	1.97

Table 12.5: Rising organic composition of Capital, Swedish data. Figures for Manufacturing and Mining. Source Edvinson 2003, table 7.5.

	1871–1900 average	1971-2000 average	%change
$\frac{c}{s+v}$	184%	305%	66
$\frac{s+v}{c}$	54%	33%	-40
$\frac{s}{s+v}$	34%	21%	-38
$\frac{s}{c}$	19%	7%	-61

Edvinsson shows data for Sweden indicating that over a prolonged period there had been a significant rise in the organic composition of capital and a fall in the rate of profit. Duménil (2002) show that there was a prolonged decline in profit rates in the USA in the postwar period analogous to that observed in the UK (see Figure 12.8).

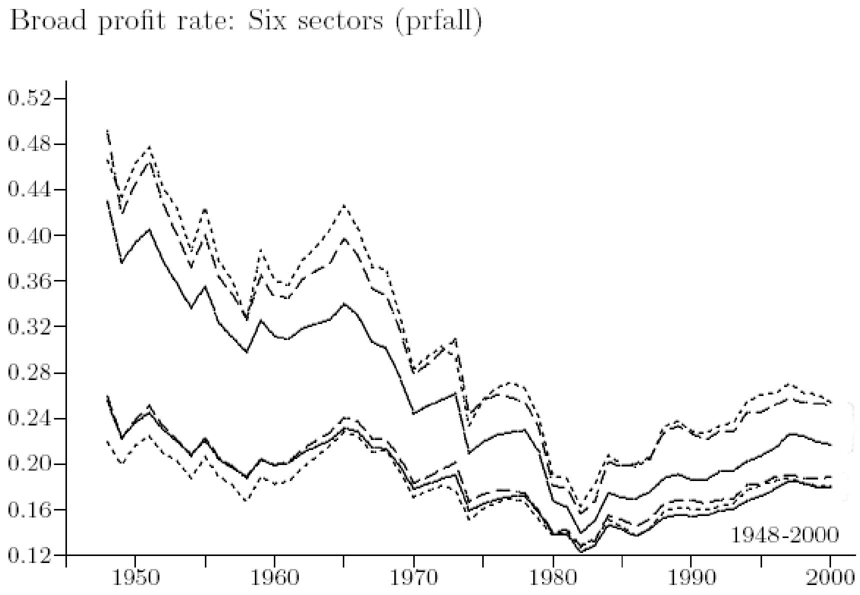


Figure 12.8: Evolution of the profit rate in the USA after Duménil and Lévy.

Since this should not occur on the basis of the microeconomic arguments put forward by Okishio, this predisposes us to believe that there must be some premises in his argument that are not an accurate reflection of the way capitalist economies actually work. A possible weakness in the Okishio theory comes from the assumption of an equalized rate of profit. This rate of profit is used as a benchmark against which possible improvements in productivity are measured. We have argued that this assumption is unrealistic. Actual profit rates show a wide dispersion,[2] wider than the dispersion in the rate of surplus value for instance. The general rate of profit is not a given or datum for an individual firm. A firm knows what its rate of profit last year was, and it knows what the interest rate is but the general rate of profit is of interest only to economic statisticians. The process by which equilibration of profit rates is supposed to come about was originally invoked by classical economists in the context of comparing things like wine maturation and forestry which had multiple year turnover times, with corn growing which had an annual turnover period. The argument was that capital would only be

---

[2] See Table 12.3.

invested in low turnover activities if it yielded the same return as in normal agriculture. Whilst this argument may have some plausibility when applied to activities agriculture and the production of vintage wine where the rate of technical change is low, and decades or centuries can be allowed for the establishment of relative prices it is less clear that it can be invoked where there is rapid technical change. In this case the time taken to establish equilibrium could be much longer than the lifespan of the technology. This is especially true in some industries with a very high capital/labour ratio, ones which are particularly relevant to the question at hand. Consider the Victorian railways. Here was an entirely new technology requiring huge capital investment. The lifetime of the capital in the form of bridges embankments and stations would be a century or more. The railway booms resulted in over capacity, which, by the early 20th century resulted in a process of line closures. But before the capital invested in railways could depreciate to a level at which the return of railway capital reached equilibrium levels, the whole technology was superseded by road transport. The sort of equilibrium that is required for the Okishio theorem can be so long in coming that the industry has died before it is relevant.

One can distinguish three possible rates that might act as investment benchmarks:

(1) the statistical average rate of profit;
(2) the average rate of return on equities; or
(3) the rate of interest available from the banks.

We have ruled out item 1, what about the rate of return on equities?

The rate of return on equities is much more accessible as data, since there are well developed stock markets that make this data available. This makes it a more plausible investment benchmark.

Suppose we have a static working population, and a rate of return on equities equal to the general rate of profit. By the Okishio theorem, any net investment in fixed capital which would raise the organic composition of capital would give the investing firm a lower return on capital than allowed by the equity market. Firms will thus tend to select only capital saving technical innovations. This implies that firms, taken as a whole will need no net infusion of capital, so that there should be zero net issue of new equities. If the rentier class as a whole decides to make no net saving, then the situation is stable. But this is unlikely. If the rentiers attempt to accumulate capital by investing in shares, the net effect will be to bid up the price of equities, given that no new equities are being issued.

The effect of this is to depress the rate of return on equities. This will affect the discount rate used in assessing investment projects. Previously unprofitable ones will seem profitable. New equities will be issued and the proceeds invested. Given that the population is static, this will raise the organic composition of capital and depress the real rate of profit.

The weakness in Okishio's argument stems from a fundamental failing of the entire price of production school of Marxian economists: they identify the formation of a general rate of return on equities with the formation of a uniform rate of profit on real invested capital.[3] The tendency towards the formation of a uniform rate of profit on equities will be much stronger than the tendency towards a uniform rate of profit on capital stocks. Stocks of capital goods held by companies originated as manufactured commodities with a price and a corresponding account book-value. T his book value may be written down due to depreciation, or written up due to inflation, but at heart their valuation remains grounded in commodity prices. In contrast the stock market valuation of a company represents, the discounted present value of its anticipated future earnings. If the company owns readily saleable capital assets, these can set a lower bound on its share price. Below this price takeovers by asset strippers become likely. Even allowing for this constraint, share prices have great flexibility and respond rapidly to changes in reported profits. These fast changes in the nominal valuation of companies can create an illusion that the rate of profit in different industries is narrowly clustered around an average profit rate.

It is more realistic to assume that it is the interest rate that firms use as a criterion of the viability of investment.

We suggest that average rate of interest for industrial borrowers is a function of the reserve ratio of the banks, such that

$$r = f(\frac{m}{Q}) \qquad f'(\frac{m}{Q}) < 0 \tag{12.23}$$

where m denotes the cash reserves of banks and Q denotes total bank deposits.

---

[3] This failing is not restricted to explicit Marxians. Sraffa and his followers share this assumption as do critics of Marx such as SamuelsonSamuelson (1973).

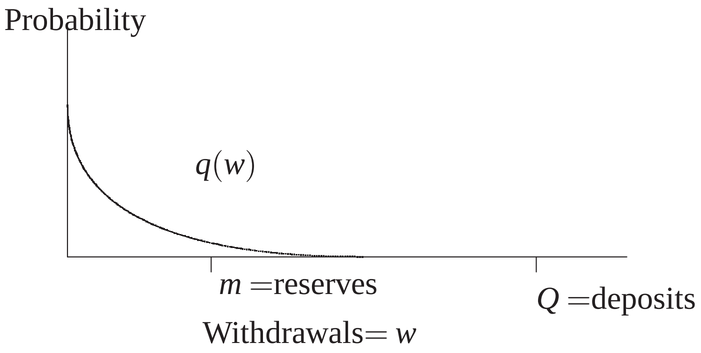


Figure 12.9: The function $q(w)$ shows the probability of withdrawals exceeding w. This reproduces Figure 11.5 in stylized form.

As discussed in Chapter 11.1, the possibility of profitable banking is based upon the fact that banks face a probability function for net withdrawals in any small time period of the form $q(w)$ as shown in Figures 11.5 and 12.9. These show that the probability of a large net withdrawal in any given time period is smaller than the probability of a small net withdrawal. This enables banks to keep reserves that make up only a fraction of total deposits, since the probability of withdrawals exceeding reserves, though finite, is small. In the event of withdrawals temporarily exceeding reserves, the bank will be forced to borrow from other banks to meet its obligations to them (the most significant portion of a bank's liabilities is always to other banks where credit money predominates). If forced to borrow to meet obligations it has to pay interest on the sum borrowed. This enables us to calculate the probable cost to a bank of such a loan, l.

$$li + r\int_{m-l}^{m} q(w)w\,dw \tag{12.24}$$

We know that q is a decreasing function of w, so it follows that the cost of making a loan will be a decreasing function of cash reserves m, but an increasing function of r, the rate payable to other banks for short term loans. Since the bank has to make a profit on its loan, the rate of interest it charges must be sufficient to cover the cost of the loan. So as l tends to zero we get the following inequality:

$$r' > i + rq(m) \tag{12.25}$$

where r' denotes rate of interest charged to industrial borrowers, i is rate of interest paid on deposits, and r is rate of interest paid on short term loans from other banks.

If we assume that i is fixed by inter-bank competition and that there is a going rate of interest charged on inter-bank loans, then it follows that the rate of interest charged to borrowers will be a decreasing function of bank reserves for each individual bank. This still leaves the global values of i and r undefined. Where a central bank exists it can be assumed that it will fix the rate r thereby exercising a control over r'. Otherwise we could treat r as a linear function of $\overline{r}'$, with the premium $\overline{r}' - r$ determined by the relative probabilities of banks and commercial borrowers defaulting (where $\overline{r}'$ is the market rate for commercial borrowers). We can assume that i will be related to r' in some lagged fashion, since if the difference between the two rates became too high, rentiers could by-pass banks by lending directly to industrial borrowers.

It follows from the above that a fall in reserve ratios will lead first to a rise in the profits of banks as a wider gap develops between rates charged to industrial borrowers and paid to depositors, followed by a general rise in interest rates as rates paid to depositors are adjusted upwards.

There is some empirical evidence for this sort of relationship between bank reserve ratios and the rate of interest. The evidence is particularly compelling for the earlier periods in the USA before active intervention by the Federal Reserve became a major factor determining the rates Cagan (1969).

12.7 Dominance of the financial sector

Why does the financial sector come to be so dominant in mature capitalist economies like Britain?

What causes it to replace manufacturing as the bedrock of the economy?

The supposed role of the financial sector is to fund investment. Savings are meant to be channelled through the banks, investment trusts and the stock market into firms that want to carry out investment in new capital stock. This process obviously does occur, but it is by no means obvious why, in the face of continuing improvements in information processing technology, the sector which carries out this channeling of funds should, over time, absorb a larger and larger portion of national resources, and appear to contribute an increasing share of national income.

Figure 12.10: Flows into and out of the financial sector.

Channelling funds is manipulation of information. The ‘funds’ are records kept by the banking system and their channelling is a sequence of transfers between records. The records have long ago moved from paper to computer databases. The power of computers has improved by leaps and bounds. One would have thought that the labour required to manage this system would have declined. The mechanisation of agriculture eliminated the peasantry, but computers have not laid waste to the City of London. Why?

The key to this paradox is to realise that despite the modern jargon of a financial services ‘industry’ that offers financial ‘products’ to customers, the financial sector is not a productive industry in the normal sense. It’s structural position in capitalistic information flows ensures its continued command over resources despite changes in technology that would decimate any other industry.

Consider Figure 12.10, it shows in summary form the flows of funds into and out of the financial sector. Savings by individual capitalists, by firms, and also from the pension schemes of employees enter the system. Funding flows out to firms carrying out capital investment, and also typically to the state to fund the public debt. However money also flows out as costs: the income of the financial sector itself. This comprises wages of its employees, the bonuses it pays, the distributed dividends of financial companies, and the costs of buildings and equipment that the sector uses. Let us denote savings by σ, bonuses and costs by β and funding of investment by ϕ.

The residual, which we will denote by δ is made up by the change in the money balances of the finacial sector itself: $\delta = \sigma - \beta - \phi$. We need to explain why β, the costs/income of the financial sector rise as a share of national income over time.

We have argued (section 12.6) that the real rate of return on capital tends to decline over the course of capitalist development. If the rate of interest foes not fall at a corresponding rate then the level of voluntary fundraising by firms will decline, since a diminishing portion of firms will be making enough profits to cover the rate of interest. However the level of savings will not necessarily decline at a corresponding rate.

The level of employees pension scheme savings changes relatively slowly - though recently British firms have been trying to reduce these. We showed in Chapter 9 that the distribution of income in capitalist societies will be highly uneven. A large proportion of income goes to a small part of the population. People with very high incomes tend to save most of it. A decline in the rate of profit on capital will not alter this. It just means that the book value of the assets of those on very high incomes rises. So savings going into the financial system will not decline. The slack can be taken up in three ways:

(1) A build up in the reserves of the financial sector :δ in our equation.
(2) An increase in borrowing by the state.
(3) A rise in the income/costs of the financial sector itself.

We can view these as short, medium and long term consequences. The immediate consequence of a fall in ϕ relative σ, will be that the reserves of financial institutions rise. If a single bank gets more deposits than it makes loans, then its cash reserves rise. But all financial institutions will have a target for the proportion of their assests that they wish to keep as cash. This target will vary over time and in response to conditions on the stock market. But their immediate response to a rise is to attempt to shift cash into other assets.

A fall in investment by non-financial companies, does often lead to a rise in state borrowing. During recessions, the state gets in less tax whilst expenditure on social security climbs. More state bonds become available as assets, this allows financial sector to limit the growth of its cash balances. But in the longer term there are political pressures to limit government budget deficits. The position of the dollar as an international reserve currency has allowed the US goverment great leeway in the accumulation of public debt, but the EU Stability Pact imposes much more stringent tests on European states.

Although over the longer term, the growth of public debt has been constrained, financial institutions can still balance their portfolios by bidding up the prices of assets. Share prices and land prices will rise until the financial sector reaches its desired cash reserve ratio. Although real capital investment may be sluggish, this is hidden from savers. They see the book price of their holdings in investment trusts etc, rise.

But there remains a conceptual problem here. A rise in the aggregate book price of shareholdings is a stock phenomenon measured in £ billion, but the variables ϕ and σ denotes flows : £ per year. One can not redress an imbalance between the flows of savings and investment by a change in stock prices. There has to be a corresponding outflow of funds. Closure is provided by the charging practices of the investment trusts. These typically charge a management fee rated a fixed proportion of the assets they manage - for example 0.5% per year. As average asset prices rise so do the management fees. The income/costs of the financial sector then rises to ensure that $\beta \approx \sigma - \phi$.

There is an inbuilt tendancy for the costs of administration to absorb uninvested surplus value.

In a young capitalist economy, like contemporary China, the financial sector exists to transfer funds into real capital investment. It allows hundreds of millions of workers to be employed in the construction of capital assets whilst recording the claims on those assets held by individual capitalists. In an old capitalist country like Britain, the financial sector abandons its old role of ‘intermediation’ and increasingly becomes a consumer of the surplus product.

The vast bureacracy of financial sector administers ever more distant claims on real assets. As the share of the population constructing these assets fell, the costs of administration rose. Between generations, millions of people shifted from making real capital goods to administering claims. Whilst the steelworks of Motherwell and Redcar yielded to the wrecker’s crane, the glass towers of the City and Canary Warf rose. This is the secret behind our famous shift from a productive to a service economy.

Burgeoning bonuses made City financial analysts the second highest paid group of employees (after CEOs). Their *average* salaries in 2005 were £80,000 a year. Such largesse generated in its wake new servant classes - nannies, cleaners, restaurant workers. Openings grew for every trade that caters to luxury: lifestyle consultants, designers, home decorators etc. House prices escalate. Television became obssesed with house makeover shows and guides to property speculation.

This vast cost was unproductive. Although it grew at the very point when its original social function atrophied, this was not obvious. Its bonuses were a form of self affirmation. They seemed a testament to productivity. In reality they were an inadvertent side effect of economic conditions way beyond the control of their recipients.

CHAPTER 13

HAYEK ON INFORMATION AND KNOWLEDGE

Cottrell, Cockshott

The fact that the present authors have, on balance, a positive view of the economic work of Karl Marx, will not have escaped the reader. Although this book, with its grounding in the physical and information sciences is not a work of orthodox Marxism, its presentation of economic issues is certainly influenced by Marx and the school of economics that follows him. Examination of the economics of information is, however, more associated with a very different school of economics: that of Hayek. Friedrich August von Hayek (1899–1992) was an Austrian economist and political philosopher, noted for his defense of liberal democracy and free-market capitalism against socialist and collectivist thought in the mid-20th century. Hayek's ideas acquired a practical relevance from their political adoption, first by the Thatcher government in Britain in the 1980s and later by post-Soviet governments in Russia and Eastern Europe. We consider that he made fundamental errors in his analysis of economic information, errors which when they became the basis for practical policy, had catastrophic effects on economic co-ordination and performance.

Prices, according to Hayek, provide the telecoms system of the economy, a means by which knowledge is diffused and disseminated.

Whereas the present authors strongly believe in the applicability of the methods of natural science to the study of social phenomena, Hayek (1955) was concerned to distinguish radically between the two domains of investigation. In the natural sciences, advances involve recognizing that things are not what they seem. Science dissolves the immediate categories of subjective experience and replaces them with underlying, often hidden, causes.

The study of society on the other hand has to take as its raw material the ideas and beliefs of people in society. The facts studied by social science

> differ from the facts of the physical sciences in being beliefs or opinions held by particular people, beliefs which as such are our data, irrespective of whether they are true or false, and which, moreover, we cannot directly observe in the minds of people but which we can recognize from what they say or do merely because we have ourselves a mind similar to theirs. (Hayek, 1955, p. 28)

He argues that there is an irreducible subjective element to the subject mater of the social sciences which was absent in the physical sciences.

> [M]ost of the objects of social or human action are not "objective facts" in the special narrow sense in which the term is used in the Sciences and contrasted to "opinions", and they cannot at all be defined in physical terms. So far as human actions are concerned, things *are* what the acting people think they are. (Hayek, 1955, pp. 27–27)

His paradigm for the social or moral sciences is that society must be understood in terms of men's conscious reflected actions, it being assumed that people are constantly consciously choosing between different possible courses of action. Any collective phenomena must thus be conceived of as the unintended outcome of the decisions of individual conscious actors.

This imposes a fundamental dichotomy between the study of nature and of society, since in dealing with natural phenomena it may be reasonable to suppose that the individual scientist can know all the relevant information, while in the social context this condition cannot possibly be met.

We believe that Hayek's objection is fundamentally misplaced. Even Laplace, who is famously cited as an advocate of determinism argued that although the universe was in principle predictable to the smallest detail, this was in practice impossible because of limited knowledge and that thus science had to have recourse to probability theory. Certainly since Boltzmann it has been understood how collective phenomena arise as 'unintended' or emergent outcomes of a mass of uncoordinated processes. Our work in Chapter 6 shows how the law of value arises in a similar way. But we did not have to model consciousness on the part of the economic actors to get this result.

In Hayek's view, there were two knowledge forms: scientific knowledge (understood as knowledge of general laws) versus "unorganized knowledge" or "knowledge of the particular circumstances of time and place". The former, he says, may be susceptible of centralization via a "body of suitably chosen experts" (Hayek (1945), p. 521) but the latter is a different matter.

> [P]ractically every individual has some advantage over others in that he possesses unique information of which beneficial use might be made, but of which use can be made only if the decisions depending on it are left to him or are made with his active cooperation. (Hayek (1945), pp. 521–22)

Hayek is thinking here of "knowledge of people, of local conditions, and special circumstances" (Hayek (1945), p. 522), e.g., of the fact that a certain machine is not fully employed, or of a skill that could be better utilized. He also cites the sort of specific, localized knowledge relied upon by shippers and arbitrageurs. He claims that this sort of knowledge is often seriously undervalued by those who consider general scientific knowledge as paradigmatic. But this leaves out of account whole layer of knowledge that is crucial for economic activity, namely knowledge of specific technologies, knowledge captured in designs, knowledge captured in software[1]. Such knowledge is not reducible to general scientific law (it is generally a non-trivial problem to move from a relevant scientific theory to a workable industrial innovation), but neither is it so time- or place-specific that it is non-communicable. The licensing and transfer of technologies in a capitalist context shows this quite clearly. It also misses out the tendency of capitalist society to capture ever human knowledge in objective form:

> once a worker's knowledge is captured as structural capital, you can then do away with the worker. In industrial capitalism the worker's surplus labor was expropriated, but you had to retain the worker as long as you wanted to make use of his labor. The worker still owned his labor power, and sold it for his wages. But in the new economy, knowledge is both labor and the means of production, both of which are expropriated and turned into structural capital for the exclusive use

---

[1] It would be anachronistic to accuse Hayek of not seeing knowledge in software, but in his day knowledge already existed in the control programs for automatic machines, for instance piano-la rolls.

> of the corporation. Thus, intellectual capital can be totally alienated from the worker. Not only is the value of the labor stolen, but the labor itself. Harris (1996)

Hayek's notion of knowledge existing solely 'in the mind' is an obstacle to understanding. It is by now all but universal practice for firms to keep records of their inputs and outputs in the form of some sort of computer spreadsheet. These computer files form an image of the firm's input–output characteristics, an image which is readily transferable.[2]

Further, even the sort of 'particular' knowledge which Hayek thought too localized to be susceptible to centralization is now routinely centralized. Take his example of the information possessed by shippers. In the 1970s American Airlines achieved the position of the world's largest airline, to a great extent on the strength of their development of the SABRE system of computerized booking of flights Gibbs (1994). Since then we have come to take it for granted that either we will be able to tap into the Internet to determine where and when there are flights available from just about any A to any B across the world. Hayek's appeal to localized knowledge in this sort of context may have been appropriate at the time of writing, but it is now clearly outdated.

13.1 INADEQUACY OF THE PRICE FORM

Prices, according to Hayek, provide the telecoms system of the economy. But how adequate is this telecoms system and how much information can it really transmit?

While insisting that very specific, localized knowledge is essential to economic decision making, Hayek clearly recognizes that the "man on the spot" needs to know more than just his immediate circumstances before he can act effectively. Hence there arises the problem of "communicating to him such further information as he needs to fit his decisions into the whole pattern of changes of the larger economic system" (Hayek, 1945, p. 525) How much does he need to know? Fortuitously, only that which is conveyed by prices. Hayek constructs an example to illustrate his point:

---

[2] Admittedly, such an image does not of itself provide any information on how, for instance, a particularly favorable set of input–output relations can be *achieved*, only that it is *possible*.

> Assume that somewhere in the world a new opportunity for the use of some raw material, say tin, has arisen, or that one of the sources of supply of tin has been eliminated. It does not matter for our purpose and it is very significant that it does not matter which of these two causes has made tin more scarce. All that the users of tin need to know is that some of the tin they used to consume is now more profitably employed elsewhere, and that in consequence they must economize tin. There is no need for the great majority of them even to know where the more urgent need has arisen, or in favor of what other uses they ought to husband the supply. (Hayek, 1945, p. 526)

Despite the absence of any such overview, the effects of the disturbance in the tin market will ramify throughout the economy just the same.

> The whole acts as one market, not because any of its members survey the whole field, but because their limited individual fields of vision sufficiently overlap so that through many intermediaries the relevant information is communicated to all. (*ibid*.)

Therefore the significant thing about the price system is “the economy of knowledge with which it operates” (Hayek, 1945, pp. 526–7). He drives his point home thus:

> It is more than a metaphor to describe the price system as a kind of machinery for registering change, or a system of telecommunications which enables individual producers to watch merely the movement of a few pointers, as an engineer might watch the hands of a few dials, in order to adjust their activities to changes of which they may never know more than is reflected in the price movements. (Hayek, 1945, p. 527)

He admits that the adjustments produced via the price system are not perfect in the sense of general equilibrium theory, but they are nonetheless a “marvel” of economical coordination. (*ibid*.)

Hayek’s example of the tin market bears careful examination. Two preliminary points should be made.

First, the market system does manage to achieve a reasonable degree of coordination of economic activities. The “anarchy of the market” is far from total chaos. In the end, through the fluctuation of prices the law of value acts. Fluctuations of prices about values do function to regulate the allocation of labour between branches of production.

Second, even in a planned economy there will always be scope for the disappointment of expectations, for projects that looked promising ex ante to turn out to be failures and so on. Failures of coordination are not confined to market systems.

That said, it is nonetheless clear that Hayek grossly overstates his case. In order to make rational decisions relating to changing one's usage of tin, one has to know whether a rise in price is likely to be permanent or transient, and that requires knowing *why* the price has risen. The current price signal is never enough in itself. Has tin become more expensive temporarily, due, say, to a strike by the tin miners? Or are we approaching the exhaustion of readily available reserves? Actions that are rational in the one case will be quite inappropriate in the other.

Prices *in themselves* provide adequate knowledge for rational calculation only if they are at their long-run equilibrium levels, but of course for Hayek they never are. On this point it is useful to refer back to Hayek's own theory of the trade cycle[3], in which the 'misinformation' conveyed by disequilibrium prices can cause very substantial macro-economic distortions. In Hayek's cycle theory, the disequilibrium price that can do such damage is the rate of interest, but clearly the same sort of argument applies at the micro level too. Decentralized profit-maximizing responses to unsustainable prices for tin or RAM chips are equally capable of generating misinvestment and subsequent chaos.

At minimum, prices may be said to carry information regarding the terms on which various commodities may currently be exchanged, via the mediation of money (so long as markets markets clear, which is not always the case). It does not follow, however, that a knowledge of these exchange ratios enable agents to calculate the profitability, let alone the social usefulness, of producing various commodities. A commodity can be produced at profit if its price exceeds the sum of the prices of the inputs required to produce it, using the production method which yields the least such sum, but the use of current prices in this calculation is legitimate only in a static context: either prices are unchanging or production and sale take zero time. Hayek, of course, stresses constant change as the rule, so he is hardly in a position to entertain this sort of assumption. Whether production of com-

---

[3] Hayek (1935); see also Lawlor and Horn (1992) and Cottrell (1994)

modity *x* will in fact prove profitable or not depends on future prices as well as current prices. And whether production of *x* currently appears profitable depends on current expectations of future prices.

If we start from the assumption that prices will almost certainly not remain unchanged in future, how are agents supposed to form their expectations?

One possibility is that they are able to gather sufficient relevant information to make a definite forecast of the changes that are likely to occur. This clearly requires that they know much more than just current prices. They must know the process whereby prices are formed, and form forecasts of the evolution of the various factors (at any rate, the more important of them) that bear upon price determination. Hayek's informational minimalism is then substantially breached. A second possibility is that described by Keynes (1936), (esp. chapter 12): agents are so much in the dark on the future that, although they are sure that some (unspecified) change will occur, they fall back upon the convention of assuming that tomorrow's prices will equal today's. This enables them to form a conventional assessment of the profitability of producing various commodities, using current price information alone; but the cost of this approach (from the standpoint of a defender of the efficiency of the market) is the recognition that those ex ante assessments will be regularly and perhaps substantially wrong.

Prices do convey objective information about the social costs of production, through the noise of their fluctuations the signal of value shines through. Because of this they may well function as a regulator of production. Divergences of prices above or below values could serve to attract or repel labour resources into and from branches of production. It is one thing to recognize that this is possible, another to assess its importance in regulating the economy. Posted prices are not the only telecoms system the economy has. Actual orders for commodities are another. Firms set prices and then get orders which are specified in quantities. If a business manager paid attention only to the prices she sold things at and ignored the quantities being ordered, the firm would not survive long. Apriori one can not say whether the price system or the quantity system is more significant in regulating the economy.

One has to know how flexible firms actually are in adjusting their prices in response to sales and then to compare this with how often they adjust their actions in response to changes in orders. Consider a supermarket, how many price adjustments does it make in a day compared to the number of new quantitative orders it places with its warehouse?

Or consider a TV factory: how often does the factory respond to orders with a change in price as compared to how often it responds by adjusting the current level of production?

Consider a design engineer deciding what components to use in a new Set Top Box for digital TV. Should the engineer base their choice solely on component price, or should they take into account information such as availability, what stocks held by suppliers, the existence of second sources?

The relative importance of the price channel and the quantity channel in inter-firm communication is an open question. One could answer it either by empirical studies of business practice or by multi-agent simulations similar to those described earlier in the book, but which had been extended to incorporate input/output tables coding the flows between industries. Given such a model one could vary the rules used by firms to respond to orders between variants in which the firms responded primarily to quantity signals and ones in which the firms responded primarily to price signals. Initial investigations by one of the authors seem to indicate that are more reliant upon price signaling can be subject to catastrophic instabilities. Fluctuations in deliveries can lead to key industries collapsing and the whole economy shutting down.

13.1.1 Information loss

Hayek is certainly right to say that prices involve an economy of information, since the process by which a price is formed is entropy reducing. If we consider an input/output table like Table 7.2, we see that it is a square matrix. A full input output table of an economy with n products would contain n^2 numbers. But the prices of these products can be encoded in a vector of only n distinct numbers. Let us assume that the entropy of interconnection of an economy H_I is encoded in the input/output table, then the entropy of the price vector H_P grows according to the law

$$H_P \approx \sqrt{H_I}$$

We will see later that this treatment somewhat overestimates the entropy of interconnection, but it is clear that there is a very substantial information reduction going on here.

How then can such a reduced information structure function to regulate the economy?

How can it work if it allows “individual producers to watch merely the movement of a few pointers”?

We will leave aside for now the relative importance of the price and quantity channels in economic information flows, and concentrate on how a single vector of prices might act as a regulator for a complex matrix of inter-sector flows. There seem to be two basic reasons why it could work:

(1) The universality of human labour means that it is possible to associate with each commodity a single scalar number - price - which indirectly represents the amount of labour that was used to make it. Deviations of relative prices from relative values can then allow labour to move from where it is less socially necessary to where it is more necessary. But this is only possible because all economic activity comes down in the end to human activity. Were that not the case, a single indicator would not be sufficient to regulate the consumption of inputs that were fundamentally of different dimensions. It is only because the dimension of all inputs is ultimately labour - direct or indirect that prices can regulate activity.

(2) Another answer lies in the computational tractability of systems of linear equations.

Consider the method that we gave in Chapter 7.3.1 for computing the labour values of commodities from an input output table. We made an initial estimate of the value of each commodity and then used the I/O table to make successively more precise estimates. What we have here is an iterative functional system where we repeatedly apply a function to the value vector to arrive at a new value vector. Because the mapping is what is termed a contractive affine transform the functional system has an attractor to which it converges. For a discussion of such systems see Barnsley (1988), in particular Chapter 3.. This attractor is the system of labour values. The system must constitute a contractive transform because any viable economy must have a net surplus product in its basic sector. Hence an initial error in the estimate of the value of an input commodity is spread over a larger quantity of the commodity on output and thus after an iteration the percentage error must decline.

> The process that we described algorithmically in Chapter 7.3.1 is what happens in a distributed manner in a real economy as prices are being formed. Firms add up wage costs and costs of other commodity inputs, add a mark-up and set their prices accordingly. This distributed algorithm, which is nowadays carried out by a combination of people and company computers, is structurally similar to that we described. It too, constitutes a contractive affine transform which converges on a price vector. Empirical evidence indicates that the price vector that it converges on lies somewhere in-between the vector of labour values and the vector of Sraffian prices. The exact attractor is not relevant at this point, what is relevant is that the iterative functional system has a stable attractor.
>
> It has this because the process of economic production can be well approximated by a piecewise contractive linear transform on price or value space. Were it the case that production processes were strongly non linear such that the output of say corn were a polynomial like:
>
> $$C_{out} = aC_{in} + bC_{in}^2 + dC_{in}^3 + dL + eL^2 + fL^3 + gI + hI^2$$
>
> with C representing corn, L labour and I iron, then the iterative functional system would be highly unstable, and the evolution of the entire price system would be completely chaotic and unpredictable. Prices would then be useless as a guide to economic activity. For the instability of such systems see Becker and Dorfler (1989) or Baker and Gollub (1990).

Neither of the two factors above are specific to a market economy. Labour is the key universal resource in any society prior to full robotisation. By the full version of the Church-Turing thesis if a problem coud be solved by a distributed collection human computers, then it can be solved by a Universal Computer. If it is tractable for a distributed collection of humans it is also algorithmically tractable when calculated by the computers of a socialist planning agency. The very factors which make the price system relatively stable and useful are the factors which make socialist economomic calculation tractable. Computing the labour value of each product is tractable, as argued in Digression 7.2, hence labour values could be used as a basis for pricing in a planned economy - transmitting basically the same information as is transmitted in prices.

Having argued that the centralized processing of much economic information is tractable, we now consider its desirability. When economic calculation is viewed as a computational process, the advantages of calculation on a distributed or decentralized basis are far from evident; the question hinges on how a multiplicity of facts about production possibilities in different branches of the economy interrelate. The interrelation of facts is, partially, an image in the field of information of the real interrelation of the branches of the economy. The outputs of one activity act as inputs for another: this is the *real* interdependence. In addition, there are *potential* interactions where different branches of production function as alternative users of inputs.

It is important to distinguish the two types of interaction. The first represents real flows of material and is a static property of a snapshot of the economy. The second, the variation in potential uses for goods, is not a property of the real economy but of the phase space of possible economies. The latter is part of the economic problem insofar as this is considered to be a search for optimal points within this phase space. According to neo-classical economic theory, the evolution of a real market economy—the real interdependencies between branches—provides the search procedure by which these optima are sought. The economy describes a trajectory through its phase space. This trajectory is the product of the trajectories of all of the individual economic agents, with these individual agents deciding upon their next position on the basis of the information they get from the price system.

Following up on Hayek's metaphor of the price system as telecoms system or machinery for registering changes, the market economy as a whole acts as a single processor[4]. A single processor, because at any one point in time it can be characterized by a single state vector that defines its position in the phase space of the economic problem. Moreover, this processor operates with a very slow cycle time, since the transmission of information is bounded by the rate of change of prices. To produce an alteration in prices, there must be a change in the real movement of goods (we are abstracting

---

[4] If we take neo-classical theory in its own terms the processor would have to be an analogue processor, since the maths of neo-classical theory is cast in terms of real variables. According to Velupillai (2003) this fundamentally undermines many of its conclusions. However, as we have argued previously, analogue computation with real numbers is, for physical reasons a fantasy. Moreover all economic transactions are done in integer quantities of money.

here from the small number of highly specialized futures markets). Thus the speed of information transmission is tied to the speed with which real goods can be moved or new production facilities brought on line. In sum, a market economy performs a single-threaded search through its state space, with a relatively slow set of adjustments to its position, the speed of adjustments being determined by how fast the real economy can move.

Contrast this now with what can potentially be done if the relevant facts can be concentrated, not in one place—that would be impossible—but within a small volume of space. If the information is gathered into one or more computing machines, these can search the possible state space without any change in the real economy.

Here the question of whether to concentrate the information is very relevant. It is a basic property of the universe that no portion of it can affect another in less time than it takes for light to propagate between them. Suppose one had all the relevant information spread around a network of computers across the country. Assume any one of these could send a message to any other. Suppose that this network was now instructed to simulate *possible* states of the economy in order to search for optima. The evolution from one simulated state to another could proceed as fast as the computers could exchange information regarding their own current state. Given that electronic signals between them travel at the speed of light this will be far faster than a real economy can evolve.

But the speed of evolution will be much faster still if we bring all of the computers into close proximity to one another. Massively parallel computers attempt to place all the processors within a small volume, thereby allowing signals moving at the speed of light to propagate around the machine in a few nanoseconds, compared to the hundredths of a second required for telecoms networks. Hence, in general, if one wishes to solve a problem fast, the information required should be placed in the smallest possible volume.

It may be objected that the sheer scale of the economic problem is such that although conceivable in principle, such computations would be unrealisable in practice (Hayek (1955);[5] see also Nove (1983)). Given modern computer technology this is far from the case as we show in section 13.3.

---

[5] The specific reference here is to p. 43, and more particularly to note 37 on pp. 212–213, of *The Counter-Revolution of Science*. In the note, Hayek appeals to the judgment of Pareto and Cournot, that the solution of a system of equations representing the conditions of general equilibrium would be practically infeasible. This is perhaps worth emphasizing

However neo-classical economists and the Austrian school have a very different concept of equilibrium from us. Our concept is of statistical equilibrium as discussed in section 6.1.2. Statistical equilibrium is not a point in phase space, but a region defined by certain macroscopic variables, such that there is a large set of microscopic conditions that are compatible with it. The concept of equilibrium with which Hayek was familiar was that of a mechanical equilibrium, a unique position in phase space at which all forces acting on the economy come into balance. Arrow and Debreu (1954) supposedly established the existence of this sort of equilibria for competitive economies, but as Velupillai (2003) showed, their proof rested on theorems that are only valid in non-constructive mathematics.

Why does it matter whether Arrow used constructive or non-constructive mathematics?

Because only constructive mathematics has an algorithmic implementation and is guaranteed to be effectively computable. But even if

(1) a mechanical economic equilibrium can be proven to exist,
(2) it can be shown that there is an effective procedure by which this can be determined : i.e., the equilibrium is in principle computable,

there is still the question of its computation tractability. What complexity order governs the computation process that arrives at the solution?

Suppose that an equilibrium exists, but that all algorithms to search for it are NP-hard, that is, the algorithms may have a running time that is exponential in the size of the problem. This is just what has been shown by Deng and Huang (2006). Their result might at first seem to support Hayek's contention that the problem of rational economic planning is computationally intractable. In Hayek's day, the notion of NP-hardness had not been invented, but he would seem to have been retrospectively vindicated. Problems with a computational cost that grows as $\mathbf{O}e^n$ soon become astronomically difficult to solve.

We mean astronomical in a literal sense. One can readily specify an NP-hard problem that involves searching more possibilities than there are atoms in the universe before arriving at a definite answer. Such problems, although in principle finite, are beyond any practical solution.

But this knife cuts with two edges. On the one hand it shows that no planning computer could solve the neo-classical problem of economic equi-

---

in view of the tendency of Hayek's modern supporters to play down the computational issue.

librium. On the other it shows that no collection of millions of individuals interacting via the market could solve it either. In neo-classical economics, the number of constraints on the equilibrium will be proportional, among other things, to the number of economic actors n. The computational resource constituted by the actors will be proportional to n but the cost of the computation will grow as e^n. Computational resources grow linearly, computational costs grow exponentially. This means that a market economy could never have sufficient computational resources to find its own mechanical equilibrium.

It follows that the problem of finding the neo-classical equilibrium is a mirage. No planning system could discover it, but nor could the market. The neo-classical problem of equilibrium misrepresents what capitalist economies actually do and also sets an impossible goal for socialist planning.

If you dispense with the notion of mechanical equilibrium and replace it with statistical equilibrium one arrives at a problem that is much more tractable. The simulations described in Chapter 6 show that a market economy can rapidly converge on this sort of equilibrium. But as we have argued above, this is because regulation by the law of value is computationally tractable. This same tractability can be exploited in a socialist planning system. Economic planning does not have to solve the impossible problem of neo-classical equilibrium, it merely has to apply the law of value more efficiently.

13.1.2 Prices, efficiency and 'know how'

It is one of the progressive features of capitalism that the process of competition forces some degree of convergence upon least-cost methods of production (even if the cost in question is monetary cost of production, which reflects social cost in a partial and distorted manner). Hayek reminds us, and rightly so, that this convergence may in fact be far from complete. Firms producing the same commodity (and perhaps even using the same basic technology) may co-exist for extended periods despite having quite divergent costs of production. If the law of one price applies to the products in question, the less efficient producers will make lower profits and/or pay lower wages.

The question arises whether convergence on best practice could be enforced more effectively in a planned system. This may be the case. If all workers are paid at a uniform rate for work done, it will be impossible for inefficient producers to mask their inefficiency by paying low wages. Indeed, with the sort of labour-time accounting system advocated elsewhere (Cottrell and Cockshott (1989), (1993)), differentials in productive efficiency will be immediately apparent. Not only that, but there should be a broader range of mechanisms for eliminating differentials once they are spotted. A private firm may realize that a competitor is producing at lower cost, but short of industrial espionage may have no way of finding out how this is achieved. Convergence of efficiency, if it is attained at all, may have to wait until the less efficient producer is driven out of business and its market share usurped by more efficient rivals. In the context of a planned system, on the other hand, some of the managers or technical experts from the more efficient enterprises might, for instance, be seconded as consultants to the less efficient enterprises. One can also imagine—in the absence of commercial secrecy—economy-wide wikipedia on which the people concerned with operating particular technologies, or producing particular products, share their tips and tricks for maximizing efficiency.

13.2 INFORMATION FLOWS UNDER MARKET AND PLAN

One of Hayek's most fundamental arguments is that the efficient functioning of an economy involves making use of a great deal of distributed information, and that the task of centralizing this information is practically impossible.

In what follows we attempt to put this argument to a quantitative test. We compare the information transmission costs implicit in a market system and a planned system, and examine how the respective costs grow as a function of the scale of the economy. Communications cost is a measure of work done to centralize or disseminate economic information: we will use the conceptual apparatus of algorithmic information theory (Chaitin (1999)) to measure this cost.

Our strategy is first to consider the dynamic problem of how fast, and with what communications overhead, an economy can stabilize. We will demonstrate that this can be done faster and at less communications cost by the planned system. We consider initially the dynamics of convergence on a fixed target, since the control system with the faster impulse response will also be faster at tracking a moving target.

Consider an economy $E = [\mathbf{A}, \mathbf{c}, r, w]$ with n producers each producing distinct products using technology matrix $\mathbf{A}$, with a well defined vector of final consumption expenditure $\mathbf{c}$ that is independent of the prices of the n products, an exogenously given wage rate w and a compatible rate of profit r. Then there exists a possible Sraffian solution $e = [\mathbf{U}, \mathbf{p}]$ where $\mathbf{U}$ is the commodity flow matrix and $\mathbf{p}$ a price vector. We will assume, as is the case in commercial arithmetic, that all quantities are expressed to some finite precision rather than being real numbers. How much information is required to specify this solution?

The argument that follows is relatively insensitive to the exact way we have specified the starting condition from which a solution is to be sought. This is because we consider convergence in information space. Recall that we have in Section 12.1.1 expressed scepticism about the existence of a given rate of profit r as assumed in Sraffian theory. We are not concerned with showing that a capitalist economy does converge towards a solution, that can be left to the neo-classical and neo-ricardian economists. Whether or not such a convergence tendency actually exists, let us concede that it does for the sake of the current argument.

Assuming that we have some efficient binary encoding method and that $I(s)$ is a measure in bits of the information content of the data structure s using this method, then the solution can be specified by $I(e)$, or, since the solution is in a sense given in the starting conditions, it can be specified by $I(E) + I(p_s)$ where p_s is a program to solve an arbitrary system of Sraffian equations. In general we have $I(e) \leq I(E) + I(p_s)$. In the following we will assume that $I(e)$ is specified by $I(E) + I(p_s)$.

Let $I(x|y)$ be the conditional or relative information (Chaitin (1987)) of x given y. The conditional information associated with any arbitrary configuration of the economy, $k = [\mathbf{U}_k, \mathbf{p}_k]$, may then be expressed relative to the solution, e, as $I(k|e)$. If k is in the neighbourhood of e we should expect that $I(k|e) \leq I(k)$. For instance, suppose that we can derive $\mathbf{U}_k$ from $\mathbf{A}$ and an intensity vector $\mathbf{u}_k$ which specifies the rate at which each industry operates then

$$I(k|e) \leq I(\mathbf{u}_k) + I(\mathbf{p}_k) + I(p_u)$$

where p_u is a program to compute $\mathbf{U}_k$ from some $\mathbf{A}$ and some $\mathbf{u}_k$. Since $\mathbf{U}_k$ is a matrix and $\mathbf{u}_k$ a vector, each of scale n, we can assume that $I(\mathbf{U}_k) > I(\mathbf{u}_k)$.

As the converges on a solution the conditional information required to specify it will shrink, since $\mathbf{u}_k$ starts to approximate to $\mathbf{u}_e$.[6] Intuitively we only have to supply the difference vector between the two, and this will require less and less information to encode, the smaller the distance between $\mathbf{u}_k$ and $\mathbf{u}_e$. A similar argument applies to the two price vectors $\mathbf{p}_k$ and $\mathbf{p}_e$. If we assume that the system follows a dynamic law that causes it to converge towards a solution then we should have the relation $I(k_{t+1}|e) < I(k_t|e)$.

Now construct a model of the amount of information that has to be transmitted between the producers of a market economy in order to move it towards a solution. Make the simplifying assumptions that all production process take one time step to operate, and that the whole process evolves synchronously. Assume the process starts just after production has finished, with the economy in some random non-equilibrium state. Further assume that each firm starts out with a given selling price for its product. Each firm i carries out the following procedure.

(1) It writes to all its suppliers asking them their current prices.

(2) It replies to all price requests that it gets, quoting its current price $\mathbf{p}_i$.

(3) It opens and reads all price quotes from its suppliers.

(4) It estimates its current per-unit cost of production.

(5) It calculates the anticipated profitability of production.

(6) If this is above r it increases its target production rate $\mathbf{u}_i$ by some fraction. If profitability is below r a proportionate reduction is made.

(7) It now calculates how much of each input j is required to sustain that production.

(8) It sends off to each of its suppliers j, an order for amount $\mathbf{U}_{ij}$ of their product.

(9) It opens all orders that it has received and

    (a) totals them up.

---

[6] Note that this information measure of the distance from equilibrium, based on a sum of logarithms, differs from a simple Euclidean measure, based on a sum of squares. The information measure is more sensitive to a multiplicity of small errors than to one large error. Because of the equivalence between information and entropy it also measures the conditional entropy of the system.

  (b) If the total is greater than the available product it scales down each order proportionately to ensure that what it can supply is fairly distributed among its customers.

  (c) It dispatches the (partially) filled orders to its customers.

  (d) If it has no remaining stocks it increases its selling price by some increasing function of the level of excess orders, while if it has stocks left over it reduces its price by some increasing function of the remaining stock.

(10) It receives all deliveries of inputs and determines at what scale it can actually proceed with production.

(11) It commences production for the next period.

Experience with computer models of this type of system indicates that if the readiness of producers to change prices is too great, the system could be grossly unstable. We will assume that the price changes are sufficiently small to ensure that only damped oscillations occur. The condition for movement towards solution is then that over a sufficiently large ensemble of points k in phase space, the mean effect of an iteration of the above procedure is to decrease the mean error for each economic variable by some factor $0 \leq g < 1$. Under such circumstances, while the convergence time in vector space will clearly follow a logarithmic law—to converge by a factor of D in in vector space will take time of order $\log_{\frac{1}{g}}(D)$—in information space the convergence time will be *linear* because of the logarithmic nature of information measures. Thus if at time t the distance from equilibrium is $I(k_t|e)$, convergence to within a distance ε will take a take a time of order

$$\frac{I(k_t|e) - \varepsilon}{\delta \log(\frac{1}{g})}$$

where δ is a constant related to the number of economic variables that alter by a mean factor of g each step. The convergence time in information space, for small ε, will thus approximate to a linear function of $I(k|e)$ which we can write as $\Delta I(k|e)$.

We are now in a position to express the communications costs of reducing the conditional entropy of the economy to some level ε. Communication takes place at steps 1, 2, 8 and 9c of the procedure. How many messages does each supplier have to send, and how much information must they contain?

Letters through the mail contain much redundant pro-forma information: we will assume that this is eliminated and the messages reduced to their bare essentials. The whole of the pro forma will be treated as a single symbol in a limited alphabet of message types. A request for a quote would thus be the pair $[R,H]$ where R is a symbol indicating that the message is a quotation request, and H the home address of the requester. A quote would be the pair $[Q,P]$ with Q indicating the message is a quote and P being the price. An order would similarly be represented by $[O,\mathbf{U}_{ij}]$, and with each delivery would go a dispatch note $[N,U_{ij}]$ indicating the actual amount delivered, where $U_{ij} \leq \mathbf{U}_{ij}$.

If we assume that each of n firms has on average m suppliers, the number of messages of each type per iteration of the procedure will be nm. Since we have an alphabet of message types (R,Q,O,N) with cardinality 4, these symbols can be encoded in 2 bits each. We will further assume that $(H,P,\mathbf{U}_{ij},U_{ij})$ can each be encoded in binary numbers of b bits. We thus obtain an expression for the communications cost of an iteration of $4nm(b+2)$. Taking into account the number of iterations, the cost of approaching the equilibrium will be $4nm(b+2)\Delta I(k|e)$.

Let us now contrast this with what would be required in a planned economy. Here the procedure involves two distinct procedures, that followed by the (state-owned) firm and that followed by the planning bureau. The model of socialist economy we are describing is roughly that given in Lange (1938) or Cottrell and Cockshott (1992). The firms do the following:

(1) In the first time period:

  (a) They send to the planners a message listing their address, their technical input coefficients and their current output stocks.

  (b) They receive instructions from the planners about how much of each of their output is to be sent to each of their users.

  (c) They send the goods with appropriate dispatch notes to their users.

  (d) They receive goods inward, read the dispatch notes and calculate their new production level.

  (e) They commence production.

(2) They then repeatedly perform the same sequence replacing step 1a with:

    (a) They send to the planners a message giving their current output stocks.

The planning bureau performs the complementary procedure:

(1) In the first period:
    (a) They read the details of stocks and technical coefficients from all of their producers.
    (b) They compute the equilibrium point *e* from technical coeffients and the final demand.
    (c) They compute a turnpike path (Dorfman et al. (1958)) from the current output structure to the equilibrium output structure.
    (d) They send out for firms to make deliveries consistent with moving along that path.
(2) In the second and subsequent periods:
    (a) They read messages giving the extent to which output targets have been met.
    (b) They compute a turnpike path from the current output structure to the equilibrium output structure.
    (c) They send out for firms to make deliveries consistent with moving along that path.

We assume that with computer technology the steps b and c can be undertaken in a time that is small relative to the production period (see below section 13.3).

Comparing the respective information flows, it is clear that the number of orders and dispatch notes sent per iteration is invariant between the two modes of organization of production. The only difference is that in the planned case the orders come from the center whereas in the market they come from the customers. These messages will again account for a communications load of $2nm(b+2)$. The difference is that in the planned system there is no exchange of price information. Instead, on the first iteration there is a transmission of information about stocks and technical coefficients. Since any coefficient takes two numbers to specify, the communications load per firm will be: $(1+2m)b$. For n firms this approximates to the $nm(b+2)$ that was required to communicate the price data.

The difference comes on subsequent iterations, where, assuming no technical change, there is no need to update the planners' record of the technology matrix. On $i-1$ subsequent iterations, the planning system has therefore to exchange only about half as much information as the market system. Furthermore, since the planned economy moves on a turnpike path to equilibrium, its convergence time will be less than that of the market economy. The consequent communications cost is $2nm(b+2)(2+(i-1))$ where $i < \Delta I(k|e)$.

The consequence is that, contrary to Hayek's claims, the amount of information that would have to be transmitted in a planned system is substantially lower than for a market system. The centralized gathering of information is less onerous than the commercial correspondence required by the market. Hayek's error comes from focusing on the price channel to the exclusion of the quantity channel. In addition, the convergence time of the market system is slower. The implication of faster convergence for adaptation to changing rather than stable conditions of production and consumption are obvious.

In addition, it should be noted that in our model for the market, we have ignored any information that has to be sent around the system in order to make payments. In practice, with the sending of invoices, cheques, receipts, clearing of cheques etc., the information flow in the market system is likely to be several times as high as our estimates. The higher communications overheads of market economies are reflected in the numbers of office workers they employ, which in turn leaves its mark on the architecture of cities—as witnessed by the skylines of Moscow and New York in the 1980s.

13.3 THE ARGUMENT FROM DYNAMICS

Does Hayek's concentration on the dynamic aspect of prices, price as a means of dynamically transmitting information, make any sense?

In one way it does. In section 8.1 we showed that the information content of a price in the UK was less than 14 bits. If we consider todays price of a cup of coffee as an example, then yesterday's price was probably the same. If the price changes only once a year, then for 364 days the only information that it conveys is that the price has not changed. The information content of this, $-\log_2 \frac{364}{365}$, is about 0.0039 of a bit. Then when the price does change its information content is $-\log_2 \frac{1}{365} + b$ where b is the number of bits to encode the price increase. For a reasonable value of the increase, say 10 pence, the whole amounts to some 12 bits. So on the day the price changes, it conveys some 3000 times as much information as it did every other day of the year.

So it is almost certainly true that most of the information in a price series is encoded in the price changes. From the standpoint of someone observing and reacting to prices, the changes are all important. But this is a viewpoint internal to the dynamics of the market system. One has to ask if the information thus conveyed has a more general import. The price changes experienced by a firm in a market economy can arise from many different causes, but we have to consider which of these represent information that is independent of the social form of production.

We can divide the changes into those that are direct results of events external to the price system, and those which are internal to the system. The discovery of new oil reserves or an increase in the birth rate would directly impinge upon the price of oil or of baby clothes. These represent changes in the needs or production capabilities of society, and any system of economic regulation should have means of responding to them. On the other side, we must count a fall in the price of acrylic feed stocks and a fall in the price of acrylic sweaters, among the second- and third-order internally generated changes consequent upon a fall in oil prices. In the same category would go the rise in house prices that follows an expansion of credit, any fluctuation in share prices, or the general fall in prices that marks the onset of a recession. These are all changes generated by the internal dynamics of a market system, and as such irrelevant to the consideration of non-market economies.

Hayek is of course right that the planning problem is greatly simplified if there are no changes, but it does not follow from this that all the changes of a market economy are potential problems for a planned one. If we assume that the economy retains some form of market for consumer goods as proposed by Lange to provide information on final requirements then the process of deriving a balanced plan is tractable.

Let us take a very simple example, an economy with 4 types of goods which we will call bread, corn, coal and iron. In order to mine coal, both iron and coal are used as inputs. To make bread we need corn for the flour and coal to bake it. To grow the corn, iron tools and seed corn are required.

Table 13.1: Convergence of gross production on that required for the final net product

```
 iron  coal  corn bread      labour
    0 20000     0  1000           0  Net output
 2000 24500  1500  1000       61000  1st estimate gross usage
 2580 29400  1650  1000      129500
 3102 31540  1665  1000      157300
 3342 33012  1666  1000      174310
 ..   ..     ..    ..        ..      hidden steps
 3708 34895  1667  1000      196510
 3708 34895  1667  1000      196515
 3708 34896  1667  1000      196517 20th estimate gross usage
```

The making of iron itself demands coal and more iron implements. We can describe this as a set of four processes:

1 ton iron ← 0.05 ton iron + 2 ton coal + 20 days labour
1 ton coal ← 0.2 ton coal + 0.1 ton iron + 3 days labour
1 ton corn ← 0.1 ton corn + 0.02 ton iron + 10 days labour
1 ton bread ← 1.5 ton corn * 0.5 ton coal + 1 days labour

Assume, following Lange (1938), that the planning authorities have a current estimate of consumer demand for final outputs. The planners start with the required net output. This is shown on the first line of Table 13.1. We assume that 20000 tons of coal and 1000 tons of bread are the consumer goods required.

They estimate how much iron, corn, coal, and labour would be directly consumed in producing the final output: 2000 tons of iron, 1500 tons of corn and 4500 additional tons of coal.

They add the intermediate inputs to the net output to get a first estimate of the gross usage of goods. Since this estimate involved producing more iron, coal and corn than they had at first allowed for, they repeat the calculation to get a second estimate of the gross usage of goods.

Each time they repeat the process they get different total requirement of iron, coal corn and labour, as shown in Table 13.1. Does this confirm the claims of Hayek that the equations necessary for socialist planning are unsolveable?

No, it does not. The answers differ each time round, but the differences between sucessive answers get smaller and smaller. Eventually, after 20 attempts in this example, the planners get a consistent result: if the population is to consume 20000 tons of coal and 1000 tons of bread, then the gross output of iron must be 3708 tons, coal must be 34896 tons and that that of corn 1667 tons.

Is it feasible to scale this up to the number of goods produced in a real economy?

Whilst the calculations would be impossibly tedious to do by hand, they are readily automated. Table 13.1 was produced by running the computer algorithm given in Appendix D. If detailed planning is to be feasible, we need to know:

(1) How many types of goods an economy produces.
(2) How many types of inputs are used to produce each output.
(3) How fast a computer program running the algorithm would be for the scale of data provided in (1) and (2).

Table 13.2 illustrates the effect of running the planning algorithm on a cheap personal computer of 2004 vintage. We determined the calculation time for economies whose number of industries ranged from one thousand to one million. Two different assumptions were tested for the way in which the mean number of inputs used to make a good depends on the complexity of the economy.

It is clear that the number of direct inputs used to manufacture each product is only a tiny fraction of the range of goods produced in an economy. It is also plausible that as industrial complexity develops, the mean number of inputs used to produce each output will also grow, but more slowly. In the first part of Table 13.2 it is assumed that the mean number of inputs (M) grows as the square root of the number of final outputs (N). In the second part of the table the growth of M is assumed to follow a logarithmic law.

It can be seen that calculation times are modest even for very big economic models. The apparently daunting million equation foe, yields gracefully to the modest home computer. The limiting factor in the experiments is computer memory. The largest model tested required 1.5 Gigabytes of memory. Since the usable data space of a P4 processor is at most 2 Gigabytes larger models would have required a more advanced 64-bit computer.

The experiment went up to 1 million products. The number of industrial products in the Soviet economy was estimated by Nove (1983) to be around

Table 13.2: Timings for applying the planning algorithm in Appendix D to model economies of different sizes. Timings were performed on a 3 Ghz Intel Zeon running Linux, with 2 GB of memory.

	Industries N	Mean Inputs M	CPU Time seconds	Memory Requirement bytes
Law $M = \sqrt{N}$				
	1,000	30	0.1	150KB
	10,000	100	3.8	5MB
	40,000	200	33.8	64MB
	160,000	400	77.1	512MB
	320,000	600	166.0	1.5G
Law $M \approx \log N$				
	1,000	30	0.1	150KB
	10,000	40	1.6	2.4MB
	100,000	50	5.8	40MB
	1,000,000	60	68.2	480MB

of 10 million. Nove believed this number was so huge as to rule out any possibility of constructing a balanced disagregated plan. This may well have been true with the computer technology available in the 1970s, but the situation is now quite different. A single PC could compute a disaggregated plan for a smallish ecoomy like Sweden in a couple of minutes.

Suppose we want to plan a continental scale economy. It might have 10 million products. Let us assume that the average number of inputs required to produce each output is, a very large, 2000. On the basis Table 13.2 this would require a computer with 80 Gigabytes of memory: Euro 6000 at 2006 prices. processor. Using a single 2006 vintage 64-bit processor the computation would take of the order of an hour.

The algorithm we have presented is for a single processor, but the problem lends itself well to parallelisation. A Beowulf cluster of PCs, costing perhaps Euro 40,000 could probably cut the compute time to under 10 minutes. More sophisticated algorithms capable of allocating fixed capital stocks have comparable complexities and running times.[7]

The compute time required is sufficiently short for a planning authority, should it so wish, to be able to perform the operation on a daily basis.

---

[7] Cockshott (1990), Cottrell and Cockshott (1992)

In performing this calculation the planners arrive at the various scales of production that the market economy would operate at were it able to attain equilibrium. Faced with an exogenous change, the planners can compute the new equilibrium position and issue directives to production units to move directly to it. This direct move will involve the physical movement of goods, laying of foundations, fitting out of buildings etc, and will therefore take some considerable time.

We now have two times, the time of *calculation* and the time of *physical adjustment*. If we assume that the calculation is performed with an iterative algorithm, we find that in practice it will converge acceptably within a dozen iterations. Since each of these iterations would take a few minutes on a supercomputer the overall time would probably be under an hour. In a market economy, even making the most favourable assumptions about its ability to adjust stably to equilibrium, the individual iterations will take a time proportional to the physical adjustment time. The overall relaxation period would be around a dozen times as long as that in the planned system (assuming a dozen convergence steps).

13.4 THE ECONOMICS OF INFORMATION

NOTE: this may not be in the right place yet.

This book is about the role—or rather, the multiple roles—played by information in the economy. Starting from some seminal articles published in the 1970s, a field of economics called the “Economics of Information” has come to increasing prominence. This field was brought to public attention in 2001, when George Akerlof, Michael Spence and Joseph Stiglitz were awarded won the Nobel Prize in Economics “for their analyses of markets with asymmetric information”.[8] The question arises: What is the relationship between our work and the Economics of Information?

The short answer is that the concerns of our book and the work that was recognized by the 2001 Nobel are largely quite distinct, although there is a degree of overlap on certain points. We’ll explain this by reference to a survey article on the Economics of Information by one of the Nobel laureates (Stiglitz, 2000). We select three main themes from Stiglitz’s article for comment:

---

[8] `http://nobelprize.org/economics/laureates/2001/public.html`

- The critique of standard neoclassical economics from the standpoint of information.
- The concept of information as a “special commodity”.
- The principal–agent problem.

On the first point, neoclassical economics has long assumed “perfect information” in constructing its analyses of the working of a market economy, up to and including what is generally regarding as the most sophisticated variant of this analysis, namely the general equilibrium theory of Kenneth Arrow and Gerard Debreu (Debreu, 1959). Practitioners of neoclassical economics have, of course, known that the perfect information assumption was unrealistic, literally false. But it was possible to believe that this assumption—which greatly simplifies the analysis—yielded a good approximation to the results of a more realistic treatment. If the competitive market system could be shown to achieve Pareto-optimal results under perfect information (see Digression 13.1), it was reasonable to suppose that it would get “close” to such optimal results when information falls a bit short of perfection.

One contribution of Stiglitz and others working in the same tradition has been to show that the hunch that “slight” imperfection of information will make only a small difference to market outcomes is wrong: even small departures from perfect information can make a substantial difference. In some instances, one can show that, as a theoretical matter, there *is* no competitive equilibrium under imperfect information. In other cases, a competitive equilibrium may exist, but be far from Pareto optimality.

This is an important finding, but it does not relate very closely to argument made in this book.

The second theme mentioned above was the idea of information as a special commodity, “fundamentally different” from other commodities (Stiglitz, 2000, p. 1448). The statements Stiglitz makes on this point strike us as valid up to a point, but not rigorously thought out (and not founded on the scientific concept of information as outlined in previous chapters of this book).

For example, Stiglitz says that the “consumption” of information is “nonrivalrous”, and that information “possesses many of the properties of a public good” (p. 1448). These terms of art mean that my making use of a piece of information does not in any way diminish your ability to make use of the same piece of information. (Compare the way in which my eating a piece of cake makes it unavailable to you—consumption of cake is “rivalrous”.) At

Digression 13.1 Pareto optimality

Pareto optimality (named after the Italian economist and political scientist Vilfredo Pareto (1848–1923)) is a concept much employed in neoclassical economics. It gives a precise (but limited) content to the idea of an "optimal" allocation of resources. Suppose the resources of a particular economy are allocated in some definite way: so much labour and so much capital are allocated to each firm, and hence each industry; the economy as a whole is producing definite quantities of each commodity; and each person is receiving some definite income. Now we ask: Is is possible to reallocate resources (e.g. take some labour away from this or that industry and give it to another) in such as way that at least one person is better off after the change *and nobody is any worse off*? If the answer is No, the original allocation was Pareto-optimal; if the answer is Yes, we were not at a Pareto optimum. There is a intuitive sense in which, if an economy is not at a Pareto optimum, it has some "slack"; and conversely if it's at a Pareto optimum, all the slack is taken up. Arguments of the sort made by Arrow and Debreu show that, under certain assumptions (including perfect information), the equilibrium state of a competitive market system will be a Pareto optimum.

The Pareto concept is limited in the following sense. We might want to argue that a state of the economy in which resources are allocated relatively equally across people is, other things equal, "better" than a state in which resources are allocated very unequally. This judgment falls outside of the scope of Pareto optimality: if moving from the unequal state to the more equal one makes *anyone* worse off—if it even slightly dents the privileges of the super-rich—it is not a "Pareto improvement". Or in other words, a particular state of the economy may be Pareto optimal, yet quite undesireable on grounds of economic justice or equality.

one level this is obviously true, and quite problematic for the efficiency of the capitalist economic mechanism. You have a piece of software on your computer. I can make a copy of that software at negligible cost, and use the copy without hurting you. So why shouldn't I make the free copy? But if it's proprietary software, and if I want to stay on the right side of the law, I may have to pay hundreds of pounds for the privilege of getting a copy that costs virtually nothing.

If we delve a little deeper, however, we see that this aspect of information in the modern economy is not actually an inherent feature of information as such. Rather, it is an outcome of a particular stage of *information technology*. Think of medieval monks laboriously copying the gospels or the writings of Aristotle. The prized information was embodied in manuscripts whose use was indeed “rivalrous” (a manuscript held in one monastery was not available to others) and which could be duplicated only at considerable cost in labour time. With printing, matters become a lot easier but the marginal cost of a printed book is still non-negligible (paper, ink, binding and so on). The present situation, where we have come to think of information as something that can be shared at close to zero cost, is a specific outcome of a complex of technologies supporting the digitization of information (so that copies are perfect), the cheap storage of information on magnetic media such as computer hard disks, and the cheap transfer of information via fibre optic cable across computer networks or via laser (burning CDs).

In the same context—the uniqueness of information as a commodity—Stiglitz writes

> A piece of information cannot be purchased like a chair. An individual can look at a chair and ascertain its properties before purchasing it. But if the seller of information tells the information that he wishes to sell to the buyer (before he has bought it), there is no reason that the individual will pay for it.

There are two points to be made here. First, Stiglitz’s chair is at one end of the spectrum, in terms of what the prospective purchaser can ascertain before buying. One can not only look at a chair, but sit in it, and arrive at an informed estimate of all its relevant properties. A little further on, Stiglitz says it’s “not only that the market for information is markedly different from the market for apples or oranges or chairs. . . ”. Not so fast. If not apples and oranges, then melons, avocados and packaged meals are rather different from the chair in this regard. One can smell the melon, gently squeeze the avocado, and read the list of ingredients of the packaged meal, but after buying these things one may still find the melon tasteless, the avocado brown and stringy inside, and the packaged meal unpalatable.

Second, the difficulty with regard to buying information is overstated. How much information are we talking about? Stiglitz’s point has full force only if a single bit of information is at issue. Is the King dead? Will a general election be called this month? To give a preview of this information is to give the information away. But the consumer doesn’t *need* a preview in the one-bit case: if I'm looking for a yes/no answer to a question, then presumably I'm well aware of the significance of the answer and well able to judge how much it's worth to me. On the other hand, consumers generally buy information in bulk, not by the bit, and in that case it's easy—in fact, routine—to give them some sort of preview or sample of the information on offer: a demo version of a piece of software; time-limited access to a subset of census data; a snippet of a movie or song.

FIXME need to write a segue to the P–A problem

A "principal" is a person or organization wanting to get some task accomplished: a homeowner who wants to get a house sold, a capitalist who wants a business run at a profit, a local authority seeking to have a school run efficiently. An "agent" is a person or organization hired, or otherwise contracted, to accomplish the given task on the principal's behalf: a real estate agent to sell the house, a manager to run the business. It is in the nature of this set-up that the agent will have more information than the principal about the possible means of accomplishing the task in hand and the trade-offs involved. In itself this is not a problem: the principal engages an agent precisely because she doesn't want to, or doesn't have time to, concern herself with the all the details of the job. But a problem arises if the agent has distinct interests, not fully congruent with those of the principal. For example, the real estate agent, whose commission is a relatively small percentage of the selling price, may favour a quick sale where the homeowner would prefer to hold out for a better price. The manager may prefer to increase the sales (and hence the size) of the business rather than maximizing its profits. In general, we may have a situation where the agent makes certain decisions on the basis of the information at his disposal, while if the principal had that same information, she would find different decisions more consonant with her interests. The structure of the situation is such that the agent may have an interest in withholding information from the principal, or even falsifying information passed to the principal, so as to leave open the opportunity of pursuing his own interests.

The principal is faced with the problem of designing the contract with the agent in such a way as to minimize the divergence of interest, or in other words, to make it worthwhile for the agent to act more or less as the principal herself would act if she took on the task herself. For a simple example, consider the real estate agent again. Suppose the agent were paid a flat fee on the sale of the house. Then he'd clearly have an incentive to sell the house as quickly as possible, which would in general be against the homeowner's interest. Paying a percentage commission in this case is a means of bringing the agent's calculations more in line with the principal's interest (but not perfectly so). Besides the payment scheme (and hence the incentives faced by the agent), a *monitoring* mechanism may play a role here.

CHAPTER 14

THE BIG PICTURE, ECONOMIC TRAJECTORIES IN BRITAIN AND CHINA

Using the model developed in section **??** and the analysis of demographic constraints in section 12.4 we can now look at the likely future trajectory of the world economy. We will look at this in terms of the interaction of Britain and China.

Figure 14.1: The evolution of accumulation as a share of profit over a century in the UK: Michaelson et al. (1995).

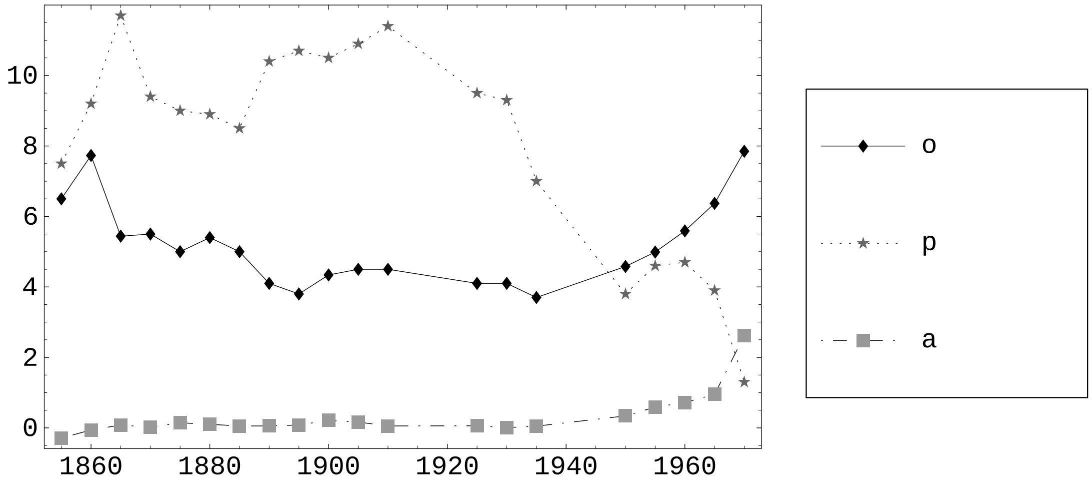


Figure 14.2: The evolution of organic composition of capital, the value rate of profit and accumulation as share of profit in the UK 1855..1970: Michaelson et al. (1995).

14.1 BRITAIN

Although history does not repeat exactly in different countries one can see in British history the effects that followed from the demographic transition. The migration from country to town was effectively complete 100 years ago. As the proletarian population became more stable and hereditary, trades union organisation spread, strikes disputes became more common. It became harder for employers to expand their workforce at the old level of wages. This process was already underway in the 'Belle Epoche' just before the first world war, a period that saw the rapid spread of general trades unions. Earlier unions had be craft based, organising skilled labour. It now proved possible to organise unions among the the bulk of the working class, not just an aristocracy of skilled artisans.

We have shown how the dynamic interaction of industrial capital and the banks polarises capital and precipitates out a class of rentiers. By the late Victorian era this process was well underway. A capitalist class whose grandfathers had been the pioneering cotton masters or iron masters of the industrial revolution had been transformed into a rentier class. Where fru-

Digression 14.1 Capital Exports

What does it mean to say that a country is a capital exporter?

In part the notion of capital export is simply a conceptual confusion. It is play on words associating the notion of export of goods with something else - the acquisition of capital assets in an other country. In certain circumstances capital 'exports' and exports of goods coincide. Consider Japan in the 1980s or '90s. It ran a large trade surplus exporting cars and electronic goods to the USA. The dollar surpluses built up by Japanese firms allowed them to invest in factories in the USA and buy up US companies - Sony buying into Hollywood for example.

The trade surplus Japan runs with the USA is part of the surplus value produced by Japanese workers - value they produce but do not consume as part of their real wage. The productive assets aquired by Japanese firms in the USA are the capitalisation of this export surplus.

Late Victorian Britain is often refered to as another capital exporter. But unlike Japan a century later, it ran an almost continuous trade deficit, indicating that it was a net consumer of surplus value. Despite this British rentiers continued to build up their overseas portfolios. Britains trade deficit was financed by the repatriation of profits on these portfolios. In order for their portfolios to grow, it was enough merely to refrain from repatriating the entirety of the profit. The flow imperial of profit in and out of the City of London made the reflux look like 'capital export' even though it was just a reinvestment of surplus produced abroad.

---

gality and accumulation had once been their watchwords, they now increasingly aped the lifestyles of their former political enemies, the landed aristocracy. Fortunes were spent constructing stately homes in the country and on employing retinues of domestic servants. With so much going on luxury consumption less was left for investment. The late Victorian rate of investment was low. Typically, less than 15% of profit was reinvested in new plant and equipment within the UK (see Figure 14.1). Another factor offsetting domestic accumulation was overseas investment in the empire or semi-colonial countries like Argentine. British rentiers were indifferent as to where their investments occured so 15% underestimates their total capital accumulation.

Empire provided Britain's rentiers with a way to sidestep the demographic transition in the home country. New populations were brought into the cash economy as described in Digression 10.1. The Queen's African subjects could be exploited even more ruthlessly than those at home. By investing overseas rather than at home, the organic composition of capital in Britain (see Figure 14.2) was kept down. This prevented the main mechanism that Marx had forseen as bring down the rate of profit. Empire also allowed a more intensive exploitation of the domestic working class even while living standards rose.

Canada, Australia, New Zealand and Argentine provided cheap grain, meat, butter, and wool. West Africa provided edible oils, cacao and kola. The replacement of less productive British agriculture by Empire suppliers cut the number of hours of labour needed to produce the weekly groceries of the average working class family. As a result a wage sufficient to maintain the same, or even an enhanced living standard represented fewer hours labour. In 1875, the wage that the average worker got for a 10 hour day was able to purchase the product of 4 hours 40 minutes of labour. By 1910, the wage for 10 hours would purchase the product of only 4 hours and 5 minutes labour. The difference of 35 mins labour per day had been transfered to property income.

As a consequence of Empire, late Victorian and Edwardian profitability remained high through the British demographic transition. These economic gains came at a huge social and political cost. From the 1890s there was increasing rivalry between the main industrial powers of Europe : Britain, France and Germany. This was not a simple trade rivalry, but a rivalry over the control of African colonies. The ensuing arms race culminated in the catastrophe of the Great War.

Britania emerged victorious from the war. Her two main historical rivals Russia and Germany were abased, their military and economic power in ruins. George V was still on his throne as King Emperor whilst his fellows William and Nicholas lost their thrones or lives. But the prosperity of the pre-war period had vanished. As Keynes (1920) described the extent to which the war had disrupted the network of trade on which that prosperity had rested, and predicted the dire consequences of excessive reparations on the defeated powers.

The post war period saw an intense struggle between financial and industrial interests in Britain. The rentier interest had been weakened by the loss of overseas assets to fund the war. They were keen to restore their international position revaluing the pound. During the war the pound had been

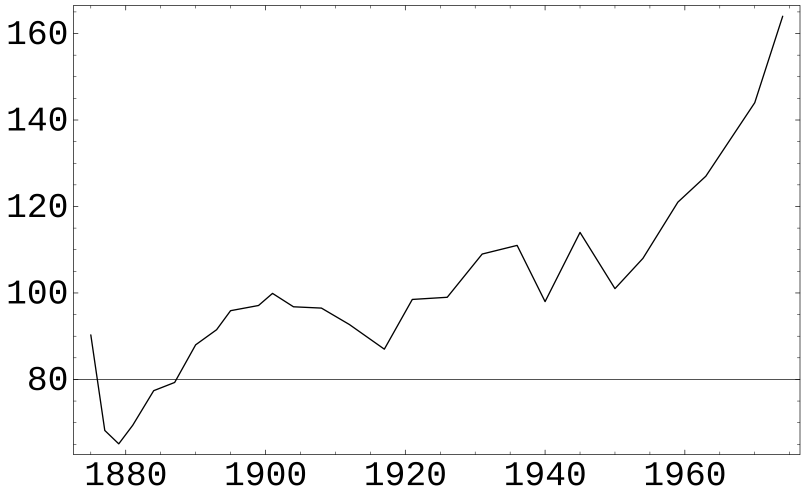


Figure 14.3: Index of real wages in the UK 1874..1974. Year 1900 wage rate taken as 100 : Michaelson et al. (1995).

devalued and convertibility to gold suspended. The rentier interest, personified in Montague Norman, head of the Bank of England were determined that the gold standard be restored. We have argued earlier that money was an irreducibly political institution, an outgrowth of the state's commuting labour taxes to symbolic taxes in coin. The metallist school of economists stressed instead the need for money to be backed by gold or silver. This theoretical doctrine disguised a real economic constituency - that of creditors in general and banks in particular. We have argued that the rate of interest depends on the ratio of bank liabilities to monetary reserves. If reserves of state money are constrained to grow slowly relative to total bank deposits, then interest rates will be high, a larger share of total profits will be appropriated by banks and rentiers. If the state issues money more rapidly, bank reserves increase and interest rates fall.

By the 1844 Bank Act, the Bank of England had been prohibited from issuing notes faster than the growth of its gold reserves. The commercial banks held Bank of England notes and deposits as their own reserves. Thus the Bank Act indirectly limited the total reserves of the banking system. This tended to hold interest rates higher than they would otherwise have been. During the war, the Bank Act had been suspended, and Sterling ceased to be convertible to gold at the Bank. The economics and politics of the inter-war period were dominated by the struggle over the gold standard. The attempt by Churchill to restore the link to gold, led inevitably [1] both to recession and to hightened class conflict. The General Strike of 1926 followed as a direct consequence.

If the value of the pound was raised, prices had to be reduced. If industrial capital was to remain profitable in the face of high interest rates and falling prices, then wages had to be cut. The miners struck to resist these wage cuts. The TUC called a General Strike in solidarity. When the timid TUC backed down in the face of Churchill's opposition, and the miners eventually lost, there was a big shift in the share of national income going as profit. Between 1924 and 1930 the rate of surplus value rose from 113% to 145%. Deflation and high interest rates combined to crush accumulation (see Figure 14.1).

The British rentier interest met a nemesis in the German National Socialism to whose birth its punative reparations had contributed. An even longer and costlier war ushered in the end of Empire. 1945 saw the election of a socialist government commited to Indian independence and a strongly progressive system of taxation. A large share of private overseas holdings had been requisitioned by the state during the war to finance imports from the USA. In consequence it was no longer possible for the country to cover a trade deficit by the repatriated profits on overseas holdings. State economic policy was directed at restoring industrial production above all else, the interests of the City of London came much lower down the list of priorities than in the past. The deliberate devaluation of the Pound against the Dollar, a more rapid growth of the state monetary base, and a deliberate policy of stimulating industrial investment transformed the development of the economy. As Figure 14.1 shows, the rate of accumulation accelerated to previously unseen levels. For some 30 years from 1945 the state had a deliberate policy of maintaining full employment, attempting to stimulate demand and investment whenever unemployment threatened to rise. We can view labour, industrial capital, and the rentier interests as three vectors pulling in different directions, as shown in Figure 14.4. During the 1950s and 1960s the combined effects of labour and industrial capital predominated. Even a conservative leader like Macmillan was, due to his background in the industrial North, and his memories of the depression, more sympathetic to the indus-

---

[1] As Keynes (1925) showed.

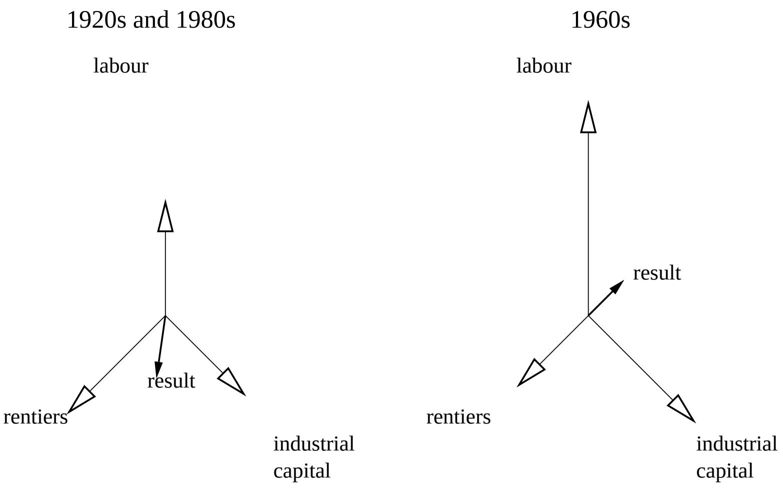


Figure 14.4: The political and economic forces acting in Britain at different points in time.

trial/labour axis than that of finance. Conservative and Labour governments continued to impose exchange controls limiting the outflow of capital.

Given that there were no significant reserves of agricultural population left - aside from marginal migration from Ireland, the rapid accumulation of capital combined with a much slower growth of the working population created social conditions that favoured the working class. Trades unions became much stronger and were able to gain consistent increases in real wages. Figure 14.3 shows how the rapid increase in wages during these 30 years was unlike anything that had gone before. If we compare Figures 14.3 and 14.1 we see that the rise in real wages almost exactly followed the rise in accumulation rates. Rapid accumulation increased the demand for labour and also raised labour productivity a combination highly favourable to militant trades unionism. It also raised the capital labour ratio which, by the Marxian law of declining profits, caused the profit rate to plummet (see Figure 14.2).

In Britain, the effects of demographic transition were delayed by half a century during which Empire and imperial wars held back domestic capital accumulation. When these brakes were removed, the law of the falling rate of profit re-asserted itself.

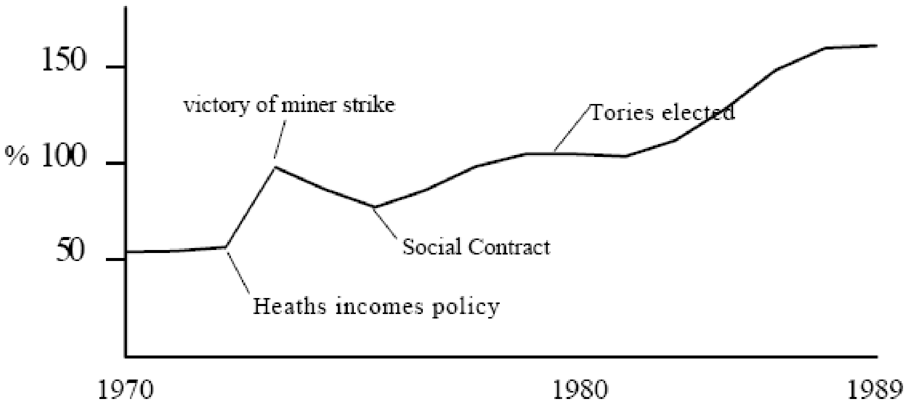


Figure 14.5: The trajectory of the rate of surplus value through and after the social crisis of the 1970s : Michaelson et al. (1995).

The falling rate of profit had, by the late 1970s, caused a general economic and social crisis. The bankruptcy of large sections of industry meant that major companies (Rolls Royce, British Leyland, the steel and shipbuilding industries) had to be nationalised to keep production going. By 1973, capital accumulation stood at more than 400% of current profits. Profits were so low that accumulation relied heavily on state funded investment and on borrowing at what were, due to inflation, effectively negative interest rates. The capitalist system of production had reached its internal limits and Britain was sleepwalking towards a socialist economy. The old model of capital accumulation funded by rentiers via the stock market had ceased to operate. A politically directed process of accumulation funded out of taxation was replacing it.

This was a real historical example of the process we examined theoretically in Figure 12.5. It is the end point of capitalist development once the growth in the working population stagnates.

Of course we know that Britain today is quite different. The election of a conservative government in 1979 ushered in a determined attempt by the propertied interest to reverse this process. Monetary policies that had been discredited since the 1930s were reintroduced. Interest rates rose steeply and large sections of industry went bankrupt. Unemployment was now welcomed to weaken the power of labour. Anti Trades Union laws further reduced the bargaining position of workers. The combined effect was to increase the rate of surplus value, reduce the organic composition of capital and increase the rate of profit (Figure 14.5). The question is, why did these policies suceed in shifting the balance of social and political forces in British society?

There are several key reasons:

(1) The global labour supply.
(2) The discovery of oil in the North Sea.
(3) Liberalisation of international capital movements.

Global Labour Supply Whilst labour supply had reached its limits in Britain, and similar demographic transitions were taking place in the other major economies of Western Europe, the same was not true of Asia. East Asia had huge reserves of labour. But these reserves were not effectively employable by European firms. Up until about 1980, China was completely off limits to Western investment. South East Asia was unattractive for investment because of the wars that had lasted from the Japanese invasion to the end of the 1970s. Communist insurgencies in Malaysia lasted down to the end of the 1950s, conflict between Britain and Indonesia lasted down to the late 1960s. South Korea and Japan although capitalist, were hostile to European investment. Only the remaining Imperial Colony of Hong Kong was open.

The 1980s saw huge changes. The Chinese government became pro-capitalist. The wars in South East Asia ended. It at last became possible to shift manufacturing from Britain, and later from other European countries to the East.

North Sea Oil In the period from 1945 to the 1980s, Britain had a constant struggle to maintain a positive balance of payments in international trade. Governments had to pay keen attention to fostering industrial production, as export industries were essential. The discovery of domestic supplies of oil transformed the balance of payments. The government could look with relative equanimity on the decline of key industries. Their exports were no longer needed.

Oil fueled trade surpluses allowed the City of London to once again build up its overseas investments. It began to have much more influence in shaping economic policy. The City also had indirect control over part of the oil surpluses of Gulf producers, who, lacking a sophisticated financial infrastructure of their own, deposited funds in London.

Capital Movements City lobbying led to the scrapping of UK exchange controls. From the ’80s British governments launched an international diplomatic campaign to get other countries to follow suit. Capital became free to move to the new labour reserves.

Under these circumstances, industry and the industrial working class became effectively surplus to requirements. The City financial institutions on the other hand became even more important and influential than in the heyday of Empire.

14.2 CHINA

The big story about the current world system is the entry of China and India as fully fledged capitalist economies. In these countries we see the same sort of rapid exponential growth that astonished continental observers of Britain in the early 19th century. These are economies undergoing a teenage growth spurt. They are economies currently free from the demographic constraint.

Suppose we have a ’typical’ firm in China. Let us suppose that the firm makes a 10% profit on turnover. Suppose half of the profit is consumed by the owner and the other half retained for internal investment. Then ideally the firm should be able to grow at 5% a year. The capitalist sector of the economy can show sustained growth rates of this order for a few decades. As the typical firm grows, it takes on more staff, buys additional stocks of raw materials and purchases larger premises. Let us suppose that the number of workers it employs grows in line with its turnover at 5%. Now if something grows at 5% a year, it doubles in size roughly every 14 years. Suppose that in 1990 there were 200 million people employed in such Chinese firms. By 2004 it would have grown to 400 million.

Clearly even in the most populous country in the world this kind of growth rate could not continue much longer. Such rapid growth in employment depends upon the existence of a surplus population drawn in from agriculture. Historically peasant populations have had the relatively high birthrate necessary for survival in the face of severe infant mortality. The first phases of modernisation in China, under the Communist government

Table 14.1: Growth in Chinese life expectancy over 50 years

Year	1955	1960	1970	1980	1990	2000	2005
Life expectancy at birth	40.8	44.6	59.6	65.3	67.1	69.7	71.0

Table 14.2: China's population by age groups, 1950 - 2050

	1950	1995	2010	2025	2050
Total	556.7	1,226.7	1,380.5	1,488.1	1,484.4
0 - 4	76.2	103.7	92.7	86.3	78.1
5 - 19	165.0	319.6	290.4	278.1	245.6
20 - 49	228.4	594.7	665.0	597.9	529.7
50 +	87.1	208.8	332.4	525.8	631.0

Future years derived from United Nations Population Projection, 1998 Revision, Medium Variant (million)

of Mao were accompanied by public health measures : the provision of ’barefoot doctors’, innoculation campaigns, measures to restrict inset pests, provision of clean water supplies. These produced massive improvements in infant survival rates and increased life expectancies (see Table 14.1).

This created an enormous surplus population that could potentially be drawn into industrial employment (see Table 14.2). Under the Mao government the policy had been to use this surplus labour in comunally owned rural agro/industrial complexes: the so called ’Peoples Communes’. These were run on socialist rather than capitalist lines, with members being paid in ’work points’ for labour performed, rather than money.

Under subsequent governments the communes were dissolved and there was large scale migration to the cities and absorbtion of the surplus population into an expanding capitalist sectors. As people move into cites and become wage workers instead of peasants there are changes in the family structure. The family is no longer a unit of production that for whom children are additional labour. Industrial society demands that children go to school and be financially supported by their parents. After a generation or so working-class families end up being smaller, the population growth

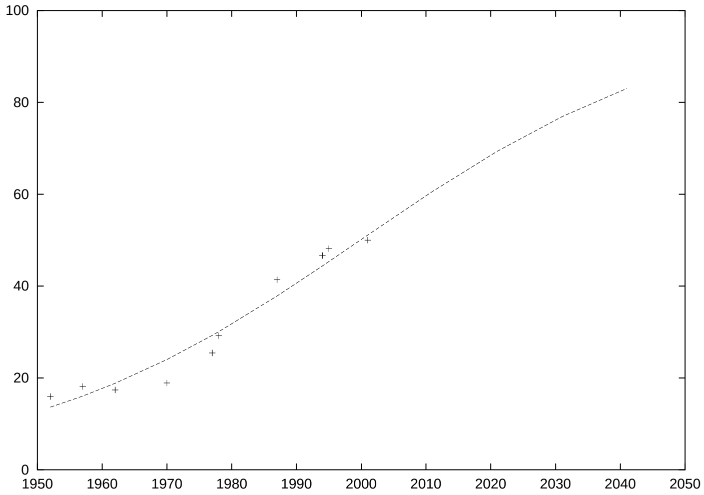


Figure 14.6: Growth of the non rural percentage of the Chinese working population. Figures after 2001 are projections using a logistic curve. Original data from: http://www.eco.rug.nl/~maddison/china_book/chap_3_tables/table3.17.pdf.

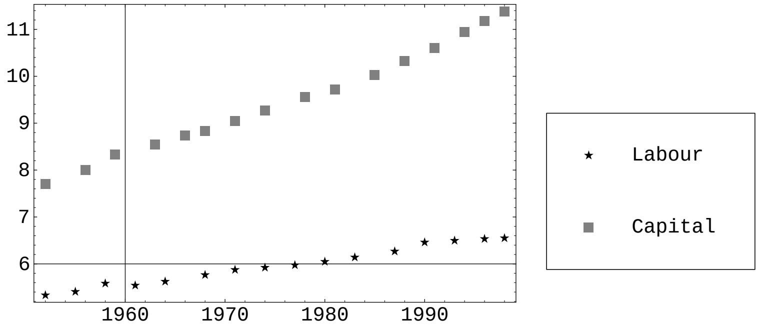


Figure 14.7: Growth labour force and capital stock in China on a log scale. Note that the growth slope of the capital stock is higher than that of the labour force. Derived from Li (2003).

slows down and migration to the cities becomes less significant. Table 14.2 and Figure 14.6 shows how this process is developing in China. The demographic transition there has been accelerated by the one-child family policy, but earlier capitalist countries went through an analogous demographic transition. Notice how in China the most economically active section of the population will peak within the next decade. Migration to the towns will continue well into the middle of the century, but it should be noted that the degree of urbanisation of a country is just a proxy for its degree of proletarianisation. A portion of the rural population in China is employed in village enterprises as described for instance by Pan and Park (1998). In 2002 of the 489 million people economically active in rural areas 132 million worked in village enterprises. These enterprises are successors to the industrial parts of the communes and are still collectively owned. This portion of the rural population is not available to the capitalist economy, so Figure 14.6 understates the degree to which the labour reserve has been used up. In 2002 some 324 million Chinese were still farmers some 50% of the workforce by 2004 this had dropped to 47%. If the shift from agriculture to industry continues at this rate, and the economically peaks by 2025, then latent reserves of labour in China will be substantially exhausted in the next 20 years.

The long term rate of profit in an economy is determined by the relative rates of growth of the labour supply and the capital stock. Figures for capital stock are not given in the China Statistical Yearbook, but they can be infered from published data on investment. Li (2003) has estimated time series for the relative growth of capital stock and the labour force in China up to 1998. Figure 14.7 plots his data on a log scale to show the comparative rates of growth of the two variables. It can be seen that the trend rate of growth of capital stock was higher than that for the labour supply[2].

Year	Capital Per Worker in 1978 Yuan
1958	1350
1978	3537
1998	12537

---

[2] Li's data gives what Marx called the 'mass of capital' employed per worker since it measures capital in constant 1978 prices. This rise in the mass of capital need not imply a rise in the organic composition of capital, since the amount of capital invested as wages may have gone up comparably.

Since 1998 the rate of capital formation has accelerated. In 1998 37% of GDP was going as capital investment, by 2004 that had fisen to 44%. In the first four months of 2006, investment was 1800 Billion Yuan against a total value of industrial output of 2461 Billion Yuan for the same period. Thus investment comprised an extraordinary 73% of industrial output. The rate of growth of industrial out put was 17%. This is a remarkable rate of growth of output by any measure, but a declining return on capital is evident. In 2004 capital formation was had grown by a factor of 10.1 over 15 years, national income had grown by a factor of 8.2. Higher rates of investment were not brining proportionate growth in the value of output. By 2006, to maintain that annual growth rate of 17% in output, capital investment was having to grow at an annual rate of 30%. China's rate of capital formation can not go much higher. China's rapid growth incidentally verifies Kalecki's thesis that investment is self financing. As the rate of investment has risen so too have the profits necessary to fund it. The role of foreign capital and of state appropriations as sources of funds has shrunk.

As the Chinese economy exhausts its supplies of peasant labour, the widespread but relatively isolated labour militancy of today can be expected to coalese into a powerful trades union movement. Real wages have been rising fast already, and this will continue. The very rapid high share of profits being accumulated will depress the proportional rate of return on capital. The profitability margin attracting capital from Europe to China will then become less marked. Faced with declining rates of return at home Chinese firms will look abroad for investment opportunities in the comming decade. The Chinese purchase of IBMs PC division, and of the remains of the UK car industry are early harbingers. China's trade surpluses mean that it is already in a position to be a substantial capital exporter. The process that occured with Britain in the 1880s or Japan from the 1980s onwards as these countries labour reserves were used up, shows us what to expect.

The critical difference between successive capital exporters is in their sizes. Japan was bigger than Britain, China is an order of magnitude larger than Japan. When it tries to become a major capital exporter the remaining underdeveloped parts of the world able to absorb that capital will be relatively small. China will of course export capital to Africa - probably comming to dominate it as much as Europe once did. South America too, will be a destination. As was the case with European imperialism a century ago, these exports will initially concentrated on the production of the raw materials needed to feed Chinese industry. Textiles, garments and toys will follow.

Since the 1980s the threat of jobs being exported to China has been held over European and American unions. In the 2020's Chinese labour unions would be less intimidated by the threat of jobs moving to Africa. Compared to Asia, Africa's labour reserves are modest. The international weakness of the labour movement since the '80s has stemmed from a glut of labour power on the world market. China's stupendous capital accumulation is rapidly reversing the balance. Capital not labour will soon be abundant. The processes that led to the European social crisis of the 1970s will, half a century later start to be replicated across Asia. Wages will tend to rise as a share of national income. The rate of return of capital will fall.

We explained in section **??** that a fall in profits relative to interest rates provokes a polarisation of capital into debtor and creditor firms. A proportion of firms is pushed towards formal insolvency. If enough do go insolvent, there is a big recession.

China is governed by a Communist Party. It is a Communist Party that allows large scale capitalist development, and one whose leading members are closely linked to capitalist business. But it is also a Communist Party that has delivered sustained increases in wages and employment. There has never been a recession in the years since they abandoned Maoism and turned to the market. A big recession like the 1930's in America would be political dynamite. To prevent it the Chinese central bank will be under pressure to hold down interest rates.

Japanese experience shows the ultimate limits of such a course. Firms whose debts would otherwise have led them to collapse, remain trading. Bad or irrecoverable debts have come to dominate the balance sheets of Japanese banks. The same will happen in China. The state will, as happened in Britain, will have to nationalise, or in the Chinese case, re-nationalise, leading industries.

With wage rates are no longer held down by competition from peasant migrants, one could see the sort of wage inflation that was common in the West during the 1970s. Back in the 19th century British wages were regulated by cyclical recessions. Each recession drove down wages, which then rose again during the boom period. But this was in a very class divided society, one where the working class did not have the vote. With a universal franchise, severe recessions give opportunites to those who promise remedial action. Think of Roosevelt and Hitler. The alternative to recession is political regulation of wages. Many European economies experimented with prices and incomes policies during the 1960-70s The state set a maximum rate of wage increase and some cases regulated the prices of key consumer goods. The aim was to increase the share of national income going in profits.

Such incomes policies are highly political. They are only imposed because of the real economic power of labour. They can only be justified by appeals to fairness and social solidarity. Profit incomes, which in the case of unregulated capitalism appear the natural return to risk or enterprise, are shown up as being inversely related to wages. The distribution relations that underpin capitalist society are exposed on the surface of politics. This raises political demands for the taxation or regulation of profit incomes. The only social justification for profit is its role in funding investment. Where, as in China today, profits are associated with rapid economic growth and generally rising incomes, their legitimacy will be at least grudgingly accepted by the majority. If firms stop investing, and at the same time wages are regulated by the state, capitalism's political credit begins to run out. As the current cycle of global economic development, the cycle that started with the opening up of China, draws to a close, the capitalist social order will face economic and political crisis on a global scale. The issue of alternative ways of regulating the economy, alternatives forgotten since the end of the USSR will have to be faced.

APPENDIX A

PROOFS

Proof of lemma 1. Substituting (6.5) into (6.8) and considering a single sector gives:

$$\dot{a_j} = \psi\gamma M(b_j - a_j \sum_{k=1}^{L} \frac{l_k b_k}{c_k a_k}) \tag{A.1}$$

which is coupled with:

$$\dot{b_j} = -\omega N(\frac{a_j}{l_j} - \frac{1}{c_j}) \tag{A.2}$$

Setting $\dot{\mathbf{a}} = \dot{\mathbf{b}} = \mathbf{0}$ yields the unique equilibrium point of the system. (A.2) implies $a_j = l_i/c_j$ and (A.1) implies $a_j = b_j$ as $\sum_{k=1}^{L} b_k = 1$. This solution is valid and unique for economies with reproduction coefficient $\eta = 1$, such that the equalities $\sum_{k=1}^{L} a_k = \sum_{k=1}^{L} b_k = \sum_{k=1}^{L} l_k/c_k = 1 = \eta$ hold. □

Proof of lemma 2. The non-linear sum in (A.1) can be eliminated as follows. Summing over all sectors:

$$\sum_{j=1}^{L} \dot{a_j} = \sum_{j=1}^{L} [\psi\gamma M(b_j - a_j \sum_{k=1}^{L} \frac{l_k}{c_k}\frac{b_k}{a_k})]$$

but given that

$$\sum_{j=1}^{L} a_j = 1 \implies \dot{a_1} + \dot{a_2} + \ldots + \dot{a_L} = 0$$

then

$$\sum_{j=1}^{L} [\psi\gamma M(b_j - a_j \sum_{k=1}^{L} \frac{l_k}{c_k}\frac{b_k}{a_k})] = 0 \implies \gamma M \sum_{j=1}^{L} [\psi b_j - a_j \sum_{k=1}^{L} \frac{l_k}{c_k}\frac{b_k}{a_k}] = 0$$

As $\gamma M \neq 0$ then

$$\sum_{j=1}^{L} \psi b_j = \sum_{j=1}^{L} a_j \sum_{k=1}^{L} \frac{l_k}{c_k} \frac{b_k}{a_k}$$

Recalling that $\sum_{j=1}^{L} a_j = 1$ and $\sum_{j=1}^{L} b_j = 1$ then

$$\sum_{j=1}^{L} \frac{l_j}{c_j} \frac{b_j}{a_j} = 1$$

Substitution into (A.1) yields a linear form of the labour equation:

$$\dot{a}_j = \psi\gamma M(b_j - a_j) \tag{A.3}$$

A change of variables, $x_j = a_j - \frac{l_j}{c_j}$ and $y_j = b_j - \frac{l_j}{c_j}$ translates the equilibrium point to the origin. Given that $\dot{\mathbf{x}} = \dot{\mathbf{a}}$ and $\dot{\mathbf{y}} = \dot{\mathbf{b}}$ the transformed linear system is:

$$\begin{aligned} \dot{\mathbf{x}} &= \psi\gamma M(\mathbf{y} - \mathbf{x}) \\ \dot{\mathbf{y}} &= -\omega N \mathbf{X}\mathbf{l} \end{aligned}$$

where $\mathbf{X}$ is the L by L diagonal matrix with (i,i) entry equal to x_i and the (i,j) $(i \neq j)$ entry zero.

The x_j and y_j represent production and income errors respectively. Consider the function

$$\begin{aligned} &V : \mathbb{R}^{2L} \to \mathbb{R} \\ &V(x_1,\ldots,x_L,y_1,\ldots,y_L) = \frac{1}{2\psi\gamma M} \sum_{j=1}^{L} x_j^2 + \frac{1}{2\omega N} \sum_{j=1}^{L} l_j y_j^2 \end{aligned}$$

that associates a scalar error measure with each possible state of the simple commodity system. In fact, V defines an error potential.

Global stability is now deduced by Lyapunov's direct method (e.g. see Brauer and Nohel (1989)). V is positive definite as $V(\mathbf{0}) = 0$ and $V(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$. Hence, V is a Lyapunov function. V is now shown to be strictly decreasing on all state trajectories:

$$\begin{aligned} V^* &= \frac{\partial V}{\partial x_1}\dot{x}_1 + \frac{\partial V}{\partial x_2}\dot{x}_2 + \ldots + \frac{\partial V}{\partial x_L}\dot{x}_L + \frac{\partial V}{\partial y_1}\dot{y}_1 + \frac{\partial V}{\partial y_2}\dot{y}_2 + \ldots + \frac{\partial V}{\partial y_L}\dot{y}_L \\ &= \frac{1}{\psi\gamma M} \sum_{j=1}^{L} x_j \dot{x}_j + \frac{1}{\omega N} \sum_{j=1}^{L} l_j y_j \dot{y}_j \end{aligned}$$

Substituting for $\dot{x}_i$ and $\dot{y}_i$ gives

$$\begin{aligned} V^* &= \sum_{j=1}^{L} x_j(y_j - x_j) - \sum_{j=1}^{L} x_j y_j \\ &= \sum_{j=1}^{L} x_j y_j - \sum_{j=1}^{L} x_j^2 - \sum_{j=1}^{L} x_j y_j \\ &= -\sum_{i=j}^{L} x_j^2 \\ &\leq 0 \end{aligned}$$

with $V^* = 0$ only when $\mathbf{x}^* = \mathbf{0}$. In other words, as time progresses the simple commodity system always follows an error-reducing trajectory that approaches the origin. By Lyapunov's Theorem the equilibrium point is asymptotically stable. Stability properties for linear systems are global. Therefore the equilibrium point is globally asymptotically stable. □

Derivation of equation (6.13). The exchange value is given by

$$\langle \varepsilon_j \rangle = \frac{\langle p_j \rangle}{\lambda} = \frac{\gamma M}{\lambda N} \frac{b_j}{a_j} l_j = \frac{b_j}{a_j} l_j$$

Rearrange to give $b_j = a_j \langle \varepsilon_j \rangle / l_j$ and substitute into equation (A.3):

$$\begin{aligned} \dot{a}_j &= \psi\gamma M(b_j - a_j) \\ &= a_j\Big(\frac{\langle \varepsilon_j \rangle}{l_j} - 1\Big)\psi\gamma M \end{aligned}$$

□

APPENDIX B

EXPERIMENTAL DETAILS

L	3	4	5	6	7	8	9	10
corr.	1.0	0.99	0.98	0.96	0.98	0.88	0.96	0.96
	0.99	1.0	0.99	0.97	0.98	0.97	0.96	0.99
	0.98	0.94	0.99	0.99	0.97	0.94	0.96	0.96
	0.98	0.99	0.93	0.99	0.95	0.97	0.97	0.91
	0.99	0.99	0.99	0.99	0.94	0.91	0.86	0.92
	0.96	0.84	0.99	0.93	0.98	0.99	0.95	0.95
	1.0	0.95	0.99	0.99	0.96	0.93	0.95	0.98
	0.99	0.97	1.0	0.98	0.96	0.94	0.94	0.95
	1.0	0.96	0.96	0.95	0.95	0.94	0.93	0.99
	1.0	1.0	0.95	0.97	0.95	0.95	0.99	0.93
mean	**0.99**	**0.96**	**0.98**	**0.97**	**0.96**	**0.94**	**0.95**	**0.95**

Table B.1: Labour value/market price correlations from random samples of the **SCE**, with parameter settings N:200, L:n ($n = 3,\ldots,10$), M:500, R:20, C:2. Each parameter setting is sampled 10 times. Results are rounded to 2 decimal places. The current implementation runs out of memory when the number of commodities exceeds 10 (and is also prohibitively slow). If $L \to N$ (i.e. the number of commodities approaches the number of actors) then the economy is unlikely to sustain production rates and correlations will decrease.

The **SCE** is defined to have reached a state of statistical equilibrium when the rate of change (sampled every 1000 time steps) of the labour value/market price vector correlation is lower than a small threshold. When this convergence condition is met the simulation continues for a further 5000 time steps in order to sample the stationary distributions. (An alternative convergence condition is to check when the rate of change of entropy of every commodity price distribution is lower than a small threshold, but this was not tried). An upper-limit of 200000 time steps is set in case convergence is not achieved within a reasonable time period. In almost every cases convergence is reached before the upper-limit. Market clearing rule $\mathbf{M}_1$ cycles until there are either no buyers or no sellers for every commodity. With a large number of actors the clearing loop takes a prohibitively long time, therefore, in practice, an upper limit of the maximum number of transaction attempts per actor is set. Once the number of maximum transactions is reached the actor is neither a buyer nor seller for any commodity. This can be interpreted as a ‘time limit’ on the market period.

APPENDIX C

COMMODITY AMPLITUDE SPACE

We will now develop the concept of an underlying space, commodity amplitude space, which can model commodity exchanges and the formation of debt. Unlike commodity space itself, this space, is a true vector space whose evolution can be modeled by the application of linear operators. The relationship between commodity amplitude space and observed holdings of commodities by agents is analogous to that between amplitudes and observables in quantum theory.

Let us consider a system of n agents and m commodities, and represent the state of this system at an instance in time by a matrix **A**, where a_{ij} represents the amplitude of agent i in commodity j. The actual value of the holding of commodity j by agent i , we denote by h_{ij} an element of the holding matrix **H**. This is related to a_{ij} by the equation $a_{ij} = \sqrt{h_{ij}}$.

C.1 COMMODITY SALES

Suppose we start off with Table 10.1 as our holding matrix H. We can generate the matrix **A** as shown in Table C.1.

Table C.1: Table of money and commodity amplitudes of agents following Table 10.1

Agent	Coin	Casava	Kola
State	$\sqrt{2}$	0	0
Alande	0	$\sqrt{6}$	$\sqrt{6}$
Tunde	0	$\sqrt{2}$	$\sqrt{5}$
Femi	$\sqrt{7}$	0	0

This gives us a matrix pair of the form

$$\mathbf{A} = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{6} & \sqrt{6} \\ 0 & \sqrt{2} & \sqrt{5} \\ \sqrt{7} & 0 & 0 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 6 & 6 \\ 0 & 2 & 5 \\ 7 & 0 & 0 \end{bmatrix} \tag{C.1}$$

Commodity sales have to respect two conservation laws:

(1) The total quantity of a commodity in existence is unchanged by the act of sale.

(2) The value of each agents holdings of money plus commodities are unchanged by an act of sale.

After a sale has taken place commodities may appreciate or be consumed so that neither of these constraints holds outside the sale itself.

The transfer of 6 kola nuts from Alande to Femi changes the column 3 of the **A** matrix as follows shown below:

$$\begin{bmatrix} 0 \\ \sqrt{6} \\ \sqrt{5} \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 0 \\ \sqrt{5} \\ \sqrt{6} \end{bmatrix}$$

This implies that the sum of the squares of the amplitudes in the column remain unchanged at 11 before and after, so a transfer can be regarded as a unitary rotation of one of the amplitude columns.

At the same time we have payment of 3 coins from Femi to Alande which can be expressed in coin amplitude space as:

$$\begin{bmatrix} \sqrt{2} \\ 0 \\ 0 \\ \sqrt{7} \end{bmatrix} \rightarrow \begin{bmatrix} \sqrt{2} \\ \sqrt{3} \\ 0 \\ 2 \end{bmatrix}$$

Since the sum of the squares in coin amplitude space remain equal to 9, we also have a unitary rotation in this space.

The transfer and payment operation affects two rows, those refering to the asset holdings of Femi and Alande (rows 2 and 4). Can we represent this as a unitary rotation as well?

Since only Alande and Femi's rows are affected and only the kola and money columns are involved, we will simplify the argument by looking at the 2 by 2 sub matrix of these rows and columns. We have the transform

$$\begin{bmatrix} 0 & \sqrt{6} \\ \sqrt{7} & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \sqrt{3} & 0 \\ 2 & \sqrt{6} \end{bmatrix}$$

It is clear that the tranformation is not a unitary rotation on the rows of the matrix. The length of the first row is $\sqrt{6}$ before the sale and $\sqrt{3}$ afterwords. But this is because our matrix is in terms of disparate units - kola nuts and coins. To we need to change the original holdings matrix so that instead of being denominated in material units it is denominated in money units. If the price of a kola nut is $\frac{1}{2}$ a coin, we must multiply the kola holding column by a half prior to obtain a value matrix **V**.

Let us illustrate this with a new and simpler example. We have two columns, column 1 for money, column 2 for kola nuts.

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$$

Where agent one has 1d of coin and no kola, and agent two has no coin and 8 kolas worth 4d. We can model the purchase of 2 kolas worth 1d by agent one from agent two by the evolution of **A** to:

$$\mathbf{A2} = \begin{bmatrix} 0 & 1 \\ 1 & \sqrt{3} \end{bmatrix}$$

which corresponds to final holdings of:

$$\mathbf{V2} = \begin{bmatrix} 0 & 1 \\ 1 & 3 \end{bmatrix}, \mathbf{H2} = \begin{bmatrix} 0 & 2 \\ 1 & 6 \end{bmatrix}$$

Note that the operation on amplitude space is a length preserving rotation on both the rows and the columns. The lengths of the row zero and column zero in **A2** are 1 the lengths of row and column one is 2 just as it was for **A.** This operation can be effected by the application of an appropriate rotation matrix so that $\mathbf{A2} = \mathbf{M}.\mathbf{A}$. A matrix which produces this particular set of rotations is:

$$\mathbf{M} = \begin{bmatrix} 0 & \frac{1}{2} \\ 1 & \frac{\sqrt{3}}{2} \end{bmatrix}$$

This form of amplitude preserving operation is characteristic of a conservation law. It enforces a particular form of symetry - that of rotation in the appropriately defined space.

C.2 PRICE CHANGES

Price movements are equivalent to the application of scaling operations which can be modeled by the application of diagonal matrices. Thus a 50% fall in the price of kola in our model would be represented by the application of the matrix $\begin{matrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{matrix}$ to the current commodity amplitude matrix. Scaling operations are not length preserving.

MONEY LOANS

Let us look at the original loan from Femi to Tunde as a whole considering both the Holding and the Debt matrices as merged into a single Worth matrix $W = [H|D]$, so that:

$$\mathbf{W} = \left[\begin{array}{ccc|cccc} 2 & 0 & 0 & 0 & 2 & 2 & 2 \\ 0 & 6 & 6 & -2 & 0 & 0 & 0 \\ 0 & 2 & 5 & -2 & 0 & 0 & 0 \\ 7 & 0 & 0 & -2 & 0 & 0 & 0 \end{array}\right] \quad \rightarrow \quad {}^{1}\mathbf{W} = \left[\begin{array}{ccc|cccc} 2 & 0 & 0 & 0 & 2 & 2 & 2 \\ 0 & 6 & 6 & -2 & 0 & 0 & 0 \\ 2 & 2 & 5 & -2 & 0 & 0 & -2 \\ 5 & 0 & 0 & -2 & 0 & 2 & 0 \end{array}\right]$$

Looking only at the last two rows we have the original worth positions of Femi and Tunde as :

Agent	Worth Vector	Net Worth
Femi	$= [7,0,0,-2,0,0,0]$	5
Tunde	$= [0,2,5,-2,0,0,0]$	5
Tunde-Femi	$= [-7,2,5,0,0,0,0]$	0

there is no difference between the net worth of the two agents. After the loan the we have

^{1}Femi	$= [5,0,0,-2,0,2,0]$	5
^{1}Tunde	$= [2,2,5,-2,0,0,-2]$	5
Difference	$= [-3,2,5,0,0,-2,-2]$	0

Again there is no difference between the net worth of the agents.

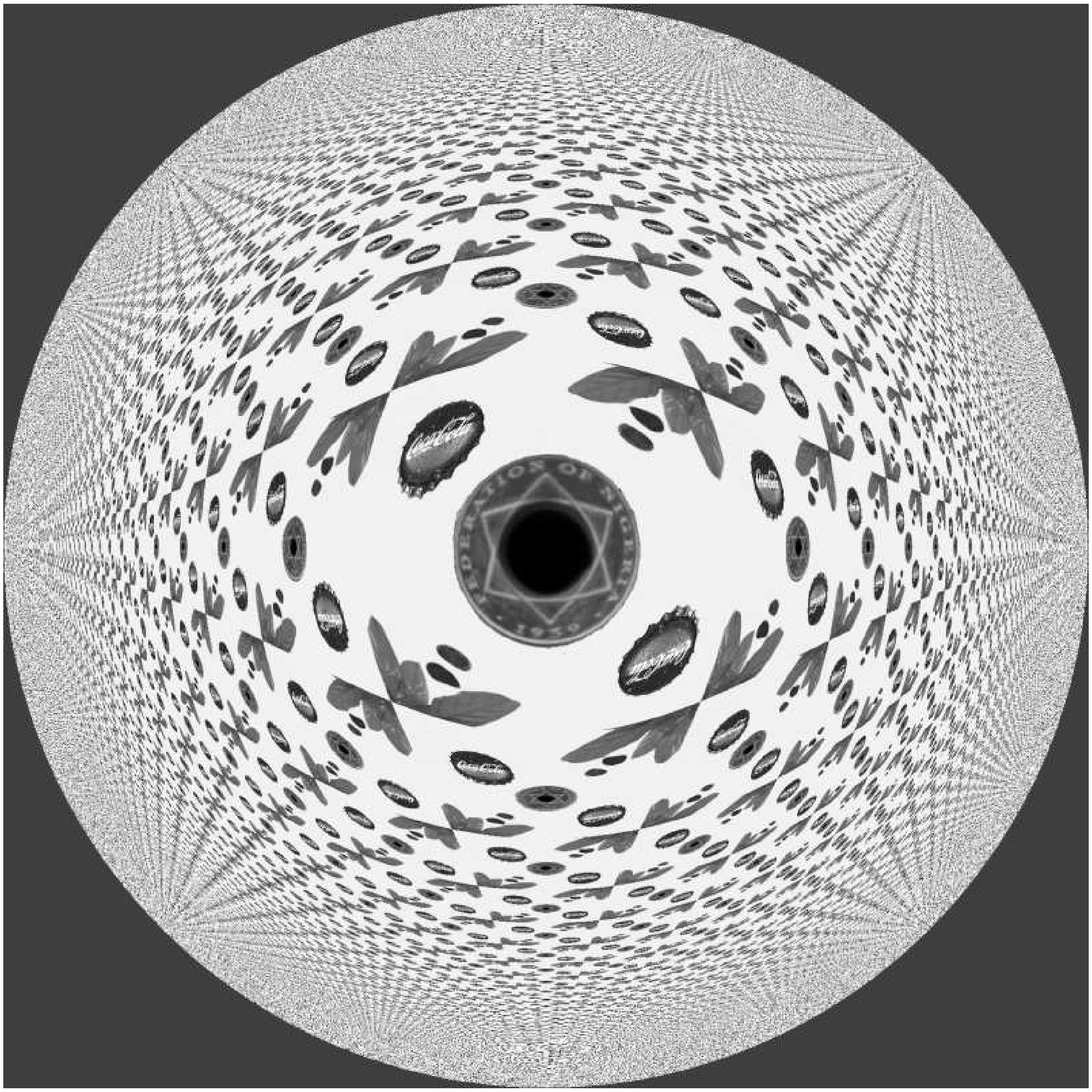


Figure C.1: Symetries and exchange relations. An artwork on the subject of this chapter, featuring the monetary system introduced by the British colonialists to Nigeria, the Kola nut, and its most famous derivative commodity.

But if we look at the worth vectors in amplitude space we get a different picture. We introduce a pair of new columns. One shows the amplitude vectors and the other shows the norms or lengths of these vectors[1].

Agent	Worth Vector	Net Worth	Amplitude	Norm
Femi	$= [7,0,0,-2,0,0,0]$	5	$[\sqrt{7},0,0,i\sqrt{2},0,0,0]$	$\sqrt{9}$
Tunde	$= [0,2,5,-2,0,0,0]$	5	$[0,\sqrt{2},\sqrt{5},i\sqrt{2},0,0,0]$	$\sqrt{9}$
Tunde-Femi	$= [-7,2,5,0,0,0,0]$	0	$[i\sqrt{7},\sqrt{2},\sqrt{5},0,0,0,0]$	$\sqrt{14}$

Note that the two agents have amplitude vectors of the same length. This expresses the fact that they have the same net worth and are in consequence an equal distance from the origin, the position where an agent has nothing of every posssible type of asset. On the other hand, although the difference in their net worth is zero, the norm of the amplitude of the difference vector is non-zero. This is because, although they have the same net worth, their asset positions are not identical. They are separated in amplitude space. The norm of the amplitude of their differences in asset holdings measures how far away from each other they are.

Now look at the effect of the loan:

^{1}Femi	$= [5,0,0,-2,0,2,0]$	5	$[\sqrt{5},0,0,i\sqrt{2},0,\sqrt{2},0]$	$\sqrt{9}$
^{1}Tunde	$= [2,2,5,-2,0,0,-2]$	5	$[\sqrt{2},\sqrt{2},\sqrt{5},i\sqrt{2},0,0,i\sqrt{2}]$	$\sqrt{13}$
Difference	$= [-3,2,5,0,0,-2,-2]$	0	$[i\sqrt{3},\sqrt{2},\sqrt{5},0,0,i\sqrt{2},i\sqrt{2}]$	$\sqrt{14}$

Note that when we look at the norms of the agents' vectors in amplitude space we find that the lengths of the vectors have not been conserved. This means that although loans conserve the net asset positions of agents, they are not unitary rotations in amplitude space in the way that commodity exchanges are. However the lengths (norms) of the amplitudes of the differences in the Worth vectors of Femi and Tunde are preserved by the loan. But this is because Femi lent money to someone poorer than him.

Suppose the situation at the start had been:

---

[1] The norm of a complex vector is the length of the vector computed by the formula

$$\mathbf{norm}(v) = \sqrt{\sum_i v_i^* \times v_i} \tag{C.2}$$

where x^* denotes the complex conjugate of x, that is, the value resulting from multiplying the imaginary part of x by -1. The effect of this is to cancel out any imaginary terms and yield a real valued result.

Because of the relationship between amplitude vectors and monetary vectors

Agent	Worth Vector	Net Worth	Amplitude	Norm
Femi	$= [7,0,0,-2,0,0,0]$	5	$[\sqrt{7},0,0,i\sqrt{2},0,0,0]$	$\sqrt{9}$
Tunde	$= [10,2,5,-2,0,0,0]$	15	$[\sqrt{10},\sqrt{2},\sqrt{5},i\sqrt{2},0,0,0]$	$\sqrt{19}$
Tunde-Femi	$= [3,2,5,0,0,0,0]$	10	$[\sqrt{3},\sqrt{2},\sqrt{5},0,0,0,0]$	$\sqrt{10}$

A loan of 2 from Femi to Tunde now results in the situation of:

^{2}Femi	$= [5,0,0,-2,0,2,0]$	5	$[\sqrt{5},0,0,i\sqrt{2},0,\sqrt{2},0]$	$\sqrt{9}$
^{2}Tunde	$= [12,2,5,-2,0,0,-2]$	15	$[\sqrt{12},\sqrt{2},\sqrt{5},i\sqrt{2},0,0,i\sqrt{2}]$	$\sqrt{23}$
Difference	$= [7,2,5,0,0,-2,-2]$	10	$[\sqrt{7},\sqrt{2},\sqrt{5},0,0,i\sqrt{2},i\sqrt{2}]$	$\sqrt{18}$

In this case the norm of Tunde's vector grows and consequently the distance vector between Femi and Tunde grows from $\sqrt{10}$ to $\sqrt{18}$. This represents the creation of net credit that we described in terms of manhattan distances in section 10.3.2.

Note that Manhattan distance in commodity debt space is related to norms in amplitude space by the formula

$$\sum |c_i| = (\mathbf{norm}(\mathbf{v}))^2 \tag{C.3}$$

where **c** is a vector in commodity debt space and **v** is the corresponding vector in amplitude space.

Purchase on credit

Consider the example of credit purchase in section 10.3.1, starting from holdings

Agent	money	kola	debts		total	Manhattan length
0	1d	0d	0	0	1d	1d
1	0d	4d	0	0	4d	4d
totals	1d	4d			5d	

agent zero buys 2d of kola from agent one. Since agent zero only has 1d in money to pay for it, the transaction leaves the following holdings:

Agent	money	kola	debts		total	Manhattan length
0	0	2d	0	-1d	1d	3d
1	1d	2d	1d	0	4d	4d
totals	1d	4d			5d	

We see that

- the totals for both money and kola are conserved,

- the total assets of each person do not change.
- the Manhattan lengths of the asset vectors were not conserved.

Thus by equation C.3 this implies that credit purchase is a non-unitary operator.

C.3 DEBT REPAYMENT - AN ANNIHILATION PROCESS

Suppose we have an initial situation in worth space with two agents with mutual debts which are then repaid we can show this as:

$$\mathbf{W} = \left[\begin{array}{c|cc} 3 & 0 & -2 \\ 2 & 2 & 0 \end{array}\right] \rightarrow {}^{1}\mathbf{W} = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 4 & 0 & 0 \end{array}\right]$$

We have a corresponding evolution of vector lengths in manhattan space as:

$$\left[\begin{array}{c} 5 \\ 4 \end{array}\right] \rightarrow \left[\begin{array}{c} 1 \\ 4 \end{array}\right]$$

It is clear that the Manhattan lengths of the vectors are not conserved, and that by equation C.3 this implies that debt repayment is a non-unitary operator.

Note that a sale in cash terms can be temporally decomposed into a sale on credit and a repayment of debt. The unitary sale operator thus corresponds to paired, individually non-unitary, debt creation and annihilation operators. When the operators corresponding to sale and purchase are temporally separated, then their non-unitary character creates real economic effects.

APPENDIX D

A SIMPLE PLANNING PROGRAM.

This algorithm performs the planning calculations that are presented in Chapter 13.3.

```
program plan ;
type
    good =( iron ,coal ,corn ,bread ,labour );
    consv =array [good ] of real ;
const
    usage: array [good ,1..3] of real =( ( 0.05,2.0,20.0),
    ( 0.2,0.1,3.0),
    ( 0.1,0.02,10.0),
    ( 1.5,0.5,1.0),
    ( 0,0,0));
    inputs: array [good ,1..3] of good =( ( iron ,coal ,labour ),
    ( coal ,iron ,labour ),
    ( corn ,iron ,labour ),
    ( corn ,coal ,labour ),
    ( corn ,coal ,labour ));
    demand :consv =( 0,2e4,0,1e3,0);
var
    Let used, previous ∈ consv;
procedure calcstep ; (see Section D.1 )
var
    Let l ∈ integer;
begin
```

```
    used ← demand;
    previous ← 0;
    writeln(iron, coal, corn, bread, labour);
    write(round(used));
    for l ← 1 to 20 do
        calcstep;
end .
```

D.1 CALCSTEP

```
procedure calcstep ;
```

This performs one step of the plan balancing by adding up the ingredients used to make the previous step of the iteration

```
var
    Let i, k ∈ good;
    Let j ∈ integer;
    Let temp ∈ consv;
begin
    temp ← 0;
    for i ← iron to labour do
        for j ← 1 to 3 do
        begin
            $k \leftarrow inputs_{i,j}$;
            $temp_k \leftarrow temp_k + (used_i - previous_i) \times usage_{i,j}$;
        end ;
        previous ← used;
        used ← used + temp;
        write(round(used));
end ;
```

APPENDIX E

PROFITS IN THE SA MODEL

.

The SA model provides the opportunity to deduce an analytical form for the industrial rate-of-profit distribution given additional assumptions on capital investment. As a step toward this goal an approximation to the firm-weighted rate-of-profit distribution in the SA model is now derived.

Consider a single firm that trades for a single year and has an average size of s employees during this period. The firm samples the market on average $12s$ times during a year. This is a simplification, as during a year, firms are created and destroyed, and therefore do not necessarily interact with the marketplace over the whole year. The value of each market sample, M_i, is a function of the instantaneous money distribution, which is mixture of exponential and Pareto forms. Assume each M_i is independent and identically distributed (iid) with mean μ_m and variance σ_m^2. During a month the same employee may be repeatedly selected, or not selected, due to the causal slack introduced by rule **1M**. Therefore the value generated per employee per month, V_i, is some function f of M_i independent of the firm size s. Simplifying further to avoid detailed consideration of the distribution of market samples per employee, assume that $f(x) = x + v$, where v is a constant. Hence each V_i is idd with mean $\mu_1 = \mu_m + v$ and variance $\sigma_1^2 = \sigma_m^2$. By the Central Limit Theorem the sum of the firm's market samples in a year, which constitutes the total revenue, R, can be approximated by a normal distribution $R = \sum V_i \approx N(\mu_r, \sigma_r^2)$, where $\mu_r = 12s\mu_1$ and $\sigma_r^2 = 12s\sigma_1^2$.

The firm's total wage bill for the year, W, is the sum of $12(s-1)$ individual wage payments, ω_i. Note that the capitalist owner does not receive wages. Each ω_i is iid according to a uniform distribution, $\omega_i \sim U(\omega_a, \omega_b)$, with mean $\mu_2 = (\omega_a + \omega_b)/2$ and variance $\sigma_2^2 = (\omega_b - \omega_a)^2/12$. By the Central Limit Theorem the wage bill, W, can be approximated by a normal distribution $W = \sum \omega_i \approx N(\mu_w, \sigma_w^2)$, where $\mu_w = 12(s-1)\mu_2$ and $\sigma_w^2 = 12(s-1)\sigma_2^2$.

Define the ratio of revenue to the wage bill as $X = R/W$ and assume that R and W are independent. X is the ratio of two normal variates and its pdf is derived by the transformation method to give:

$$
\begin{aligned}
f_X(x \mid s) \quad &= \quad \frac{\exp\left[-\frac{1}{2}\left(\frac{\mu_r^2}{\sigma_r^2} + \frac{\mu_w^2}{\sigma_w^2}\right)\right]}{4\pi k_1^{3/2}} \\
&\left(2\sigma_r\sigma_w\sqrt{k_1} + \exp^{\Lambda(x)}\sqrt{2\pi}\left(1 + \Phi\left(\sqrt{\Lambda(x)}\right)\right)\left(\mu_w\sigma_r^2 + x\mu_r\sigma_w^2\right)\right)
\end{aligned}
\tag{E.1}
$$

where

$$
\begin{aligned}
k_1 &= \sigma_r^2 + x^2\sigma_w^2 \\
\Lambda(x) &= \frac{(\mu_w\sigma_r^2 + x\mu_r\sigma_w^2)^2}{2\sigma_r^2\sigma_w^2(\sigma_r^2 + x^2\sigma_w^2)} \\
\Phi(x) &= \frac{2}{\sqrt{\pi}}\int_0^x \exp^{-t^2} \underset{\cdot}{t}
\end{aligned}
$$

(8) is the pdf of the rate-of-profit conditional on the firm size s. The unconditional rate-of-profit distribution can be obtained by considering that firm sizes are distributed according to a Pareto (power-law) distribution:

$$
f_S(s) = \frac{\alpha\beta^\alpha}{s^{\alpha+1}}
$$

where α is the shape and β the location parameter. Firm sizes in the model range between 1 (a degenerate case of an unemployed worker) to a maximum possible size N. Therefore the truncated Pareto distribution

$$
g_S(s) = f_S(s \mid 1 < S \le N) = \frac{f(s)}{F(N) - F(1)} = \frac{s^{-(1+\alpha)}\alpha}{1 - N^{-\alpha}}
$$

where

$$
f(s) = F'(s)
$$

is formed to ensure that all the probability mass is between 1 and N. By the Theorem of Total Probability the unconditional distribution $f_X(x)$ is given by:

$$f_X(x) = \int_2^N f_X(x \mid s) g_S(s)\text{ṣ} \tag{E.2}$$

where the range of integration is between 2 and N as firms of size 1 are a degenerate case that do not report profits. Expression (9) defines the $g_S(s)$ parameter-mix of $f_X(x \mid S = s)$. The rate-of-profit variate is therefore composed of a parameter-mix of a ratio of independent normal variates each conditional on a firm size s that is distributed according to a power-law. Writing (9) in full yields:

$$\begin{aligned} f_X(x) \quad = \quad & \int_2^N \frac{\exp\left[-6\left(\frac{s\mu_1^2}{\sigma_1^2} + \frac{(s-1)\mu_2^2}{\sigma_2^2}\right)\right]}{2\pi\Theta^{3/2}(x)} \\ & \left(k_2\sqrt{\Theta(x)} + \sqrt{6\pi}\Psi(x)\exp\left[\frac{6\Psi^2(x)}{k_2^2\Theta(x)}\right]\left(1 + \Phi\left(\frac{\sqrt{6}\Psi(x)}{k_2\sqrt{\Theta(x)}}\right)\right)\right) \\ & \frac{s^{-(1+\alpha)}\alpha}{1 - N^{-\alpha}}\,\text{ṣ} \end{aligned} \tag{E.3}$$

where

$$\begin{aligned} k_2 &= \sqrt{s\sigma_1^2}\sqrt{(s-1)\sigma_2^2} \\ \Theta(x) &= s\sigma_1^2 + (s-1)x^2\sigma_2^2 \\ \Psi(x) &= (s-1)s(\mu_2\sigma_1^2 + x\mu_1\sigma_2^2) \end{aligned}$$

(10) is the pdf of $X = R/W$ but the rate-of-profit in the simulation is measured as $P = 100(X-1)$. The pdf of P is therefore a linear transform of X:

$$f_P(x) = \frac{1}{100} f_X\left(1 + \frac{x}{100}\right) \tag{E.4}$$

(11) defines a distribution with 6 parameters: (i) the mean employee market sample μ_1, (ii) the variance of the employee market sample σ_1^2, (iii) the mean wage μ_2, (iv) the wage variance σ_2, (v) the Pareto exponent, α, of the firm size distribution, and (vi) the number of economic actors N.

(11) is solved numerically to compare the distribution of the theoretically derived variate P with the profit data generated by the SA model. The

values of the parameters are measured from the simulation. In this particular case $\mu_1 = \mu_m + v \approx 50 + v$, $\sigma_1^2 = \sigma_m^2 \approx 55000$, $\mu_2 \approx 50$, $\sigma_2^2 \approx 533.3$, $\alpha \approx 1.04$ and $N = 1000$. The best fit is achieved with $v = 25$ coins. Figure 12 plots the pdf $f_P(x)$ with the rate-of-profit frequency histogram of Fig. 11 and shows a reasonable fit between the derived distribution and the data.

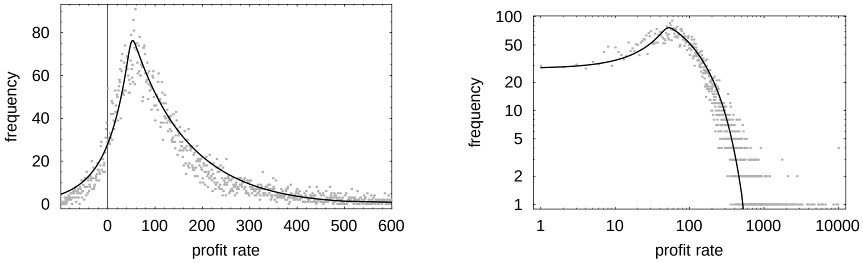


Figure E.1: Theoretical fit to the firm-weighted rate-of-profit distribution. The solid lines plot the theoretical pdf $f_P(x)$ scaled by a constant in the frequency axis. The RHS graph plots the function and data in log-log scale and extends the range of the plot to the super-profit range. All profits in excess of 10000 are truncated, which accounts for the outlier at the maximum profit rate.

With some further work the 6-parameter distribution $f_P(x)$ could be fitted to empirical rate-of-profit measures and compared against other candidate functional forms. Although (11) ignores effects due to capital investment the interpretation of the parameters μ_1, σ_1^2, μ_2 and σ_2^2 can be extended to refer to the means and variances of revenue and cost per employee Wright (2004). A testable consequence of the SA model is the conjecture that the empirical rate-of-profit distribution will be consistent with a parameter-mix of a ratio of normal variates with means and variances that depend on a firm size parameter that is distributed according to a power law.

BIBLIOGRAPHY

Althusser, L. and Balibar, É.: 1970, *Reading Capital*, New Left Books.

Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, F. and Salinger, M. A.: 1997, Scaling behavior in economics: I. empirical results for company growth, *Journal de Physique I France* **7**, 621–633.

Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Maass, P., Salinger, M. A., Stanley, H. E. and Stanley, M. H. R.: 1997, Scaling behavior in economics: the problem of quantifying company growth, *Physica A* **244**, 1–24.

Amaral, L. A. N., Gopikrishnan, P., Plerou, V. and Stanley, H. E.: 2001, A model for the growth dynamics of economic organizations, *Physica A* **299**, 127–136.

Arrow, K. and Debreu, G.: 1954, Existence of an Equilibrium for a Competitive Economy, *Econometrica* **22**(3), 265–290.

Ausloos, M., Miskiewicz, J. and Sanglier, M.: 2004, The duration of recession and prosperity: does their distribution follo a power or an exponential law?, *Physica A* pp. 548–558.

Axtell, R.: 1999, The emergence of firms in a population of agents: local increasing returns unstable nash equilibria, and power law size distributions. CSED working paper No.3.

Axtell, R. L.: 2001, Zipf distribution of u.s. firm sizes, *Science* **293**, 1818–1820.

Babbage, C.: 1832, *The Economy of Machinery and Manufactures*, London.

Bachelard, G.: 1970, *La philosophie du non: essai d'une philosophie du nouvel esprit scientifique*, Presses Universitaires de France, Paris.

Badger, W. W.: 1980, An entropy-utility model for the size distribution of income, *in* B. J. West (ed.), *Mathematical models as a tool for social science*, Gordon and Breach, New York, pp. 87–120.

Baker, G. L. and Gollub, J. P.: 1990, *Chaotic Dynamics*, Cambridge University Press.

Barnsley, M.: 1988, *Fractals Everywhere*, Academic Press.

Becker, K. H. and Dorfler, M.: 1989, *Dynamical Systems and Fractals*, Cambridge University Press.

Bennett, C. H.: 1988, Logical depth and physical complexity, *The Universal Turing Machine: a Half Century Survey*, Oxford University Press.

Bernstein, P. A., Shipman, D. W. and James B. Rothnie, J.: 1980, Concurrency control in a system for distributed databases (sdd-1), *ACM Trans. Database Syst.* **5**(1), 18–51.

Berry, G.: 1988, *Seventeenth Century England: Traders and their Tokens*, Seaby, London.

Blanchard, O., Fischer, S. and Blanchard, O.: 1989, *Lectures on Macroeconomics*, Mit Pr.

Bolin, S.: 1958, *State and Currency in the Roman Empire to 300 A.D*, Stockholm.

Bottazzi, G. and Secchi, A.: 2003, Explaining the distribution of firms growth rates.

Bouchaud, J.-P. and Mezard, M.: n.d., Wealth condenstation in a simple model of economy. http://www.arXiv.org/abs/cond-mat/0002374.

Brillouin, L.: 1951, Maxwell's demon cannot operate: Information and entropy, *Journal of Applied Physics* **22**, 334–7.

Brynjolfsson, E.: 1993, The productivity paradox of information technology, *Commun. ACM* **36**(12), 66–77.

Cagan, P.: 1969, Interest rates and bank reserves, *in* J. Guttentag and P. Cagan (eds), *Essays on Interest Rates*, National Bureau of Economic Research, New York.

Canning, D., Amaral, L. A. N., Lee, Y., Meyer, M. and Stanley, H. E.: 1998, Scaling the volatility of gdp growth rates, *Economics Letters* **60**, 335–341.

Carradice, I. and Price, M.: 1988, *Coinage in the Greek World*, Seaby, London.

Chaitin, G.: 1987, *Information, Randomness and Incompleteness*, World Scientific.

Chaitin, G.: 1999, Information and randomness: A survey of algorithmic information theory, *The Unknowable*, Springer, Signapore.

Chomsky, N.: 1956, Three models for the description of language, *IRE Trans. on Information Theory* **2(3)**, 113–124.

Cockshott, P. and Renfrew, K.: 2004, *SIMD programming for Windows and Linux*, Springer.

Cockshott, W. P.: 1990, Application of artificial intelligence techniques to economic planning, *Future Computing Systems* **2**, 429–443.

Cockshott, W. P. and Cottrell, A. F.: 1998, Does marx need to transform?, *in* R. Bellofiore (ed.), *Marxian Economics: A Reappraisal*, Vol. 2, Basingstoke, pp. 70–85.

Cockshott, W. P. and Cottrell, A. F.: 2003, A note on the organic composition of capital and profit rates, *Cambridge Journal of Economics* **27**, 749–754.

Cook, W. and Ormerod, P.: 2003, Power law distribution of the frequency of demises of us firms, *Physica A* **324**, 207–212.

Cottrell, A.: 1994, Hayek's early cycle theory re-examined, *Cambridge Journal of Economics* **18**, 197–212.

Cottrell, A. and Cockshott, P.: 1992, *Towards a New Socialism*, Vol. Nottingham, Bertrand Russell Press.

Cottrell, A. and Cockshott, W. P.: 1989, Labour value andsocialist economic calculation, *Economy and Society* **18**, 71–99.

Davies, R.: 1989, *Service in the Roman Army*, Edinburgh University Press.

Debreu, G.: 1959, *Theory of value – an axiomatic analysis of economic equilibrium*, Yale University Press, New Haven and London.

Deng, X. and Huang, L.: 2006, On the complexity of market equilibria with maximum social welfare, *Information Processing Letters* **97**(1), 4–11.

DiVincenzo, D.: 1995, Quantum computation, *Science* **270**, 255–261.

Dorfman, R., Samuelson, P. and Solow, R.: 1958, *Linear Programming and Economic Analysis*, McGraw Hill, New York.

Dragulescu, A. A.: 2003, *Applications of Physics to economics and finance: money, income, wealth and the stock market*, PhD thesis, Department of Physics. http://arXiv.org/abs/cond-mat/0307341.

Dragulescu, A. and Yakovenko, V. M.: 2000, Statistical mechanics of money, *The European Physical Journal B* **17**, 723–729.

Dragulescu, A. and Yakovenko, V. M.: 2002, Statistical mechanics of money, income and wealth: a short survey. http://arXiv.org/abs/cond-mat/0211175.

Duménil, G.: 1983, Beyond the transformation riddle: a labor theory of value, *Science and Society* **47**(4), 427–450.

Duménil, G.: 2002, The profit rate: Where and how much did it fall? did it recover? (usa 1948-2000), *Review of Radical Political Economy* **34**, 437–461.

Edvinsson, R.: 2003, A tendancy for the rate of profit to fall, *Paper presented at the economic-historical meeting in Lund*.

Edvinsson, R.: 2005, *Growth, Accumulation, Crisis: With New Macroeconomic Data for Sweden 1800-2000*, PhD thesis, Stockholm University.

Fabritiis, G. D., Pammolli, F. and Riccaboni, M.: 2003, On size and growth of business firms, *Physica A* **324**, 38–44.

Farjoun, E.: 1984, *"Production of commodities by means of what?"*.

Farjoun, E. and Machover, M.: 1983, *Laws of Chaos, a Probabilistic Approach to Political Economy*, Verso, London.

Feynman, R. P.: 1999, Simulating physics with computers, pp. 133–153.

Fisher, D.: 1963, *The Epic of Steel*, harper and Row, New Yourk.

Foley, D. K.: 1982, The value of money, the value of labor power, and the marxian transformatio problem, *Review of Radical Political Economics* **14**(2), 37–47.

Foley, D. K.: 1983, On marx's theory of money, *Social Concept* pp. 5–19.

Foley, D. K.: 2003, Sraffa's legacy, *Cambridge Journal of Economics* **27**(2), 225–238. available at http://ideas.repec.org/a/oup/cambje/v27y2003i2p225-238.html.

Foley, D. K. and Michl, T. R.: 1999, *Growth and Distribution*, Harvard University Press, Cambridge.

Forstater, M.: 2003, Taxation: A secret of colonial capitalist (so-called) primitive accumulation, *Technical Report 25*, Center for Full Employment and Price Stability.

Frisch, U. B., Hasslacher, B. and Pomeau, Y.: 1986, Lattice gas automata for the navier stokes equation, *Physical Review Letters* **56**, 1505–08.

Fujiwara, Y., Guilmi, C. D., Aoyama, H., Gallegati, M. and Souma, W.: 2004, Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms, *Physica A* **335**, 197–216.

Gaffeo, E., Gallegati, M. and Palestrini, A.: 2003, On the size distribution of firms: additional evidence from the g7 countries, *Physica A* **324**, 117–123.

Gatti, D. D., Guilmi, C. D., Gaffeo, E., Giulioni, G., Gallegati, M. and Palestrini, A.: 2003, A new approach to business fluctuations: heterogenous interacting agents scaling laws and financial fragility. To appear in Journal of Economic Behaviour and Organisation.

Gibbs, W. W.: 1994, Software's chronic crisis, *Scientific American* **271**, 86–95.

Gode, D. and Sunder, S.: 1993, Allocative efficiency of markets with zero intelligence traders: market as a partial substitute for individual rationality, *Journal of Political Economy* **101**, 119–137.

Goguen, J. A. and Meseguer, J.: 1982, Completeness of many-sorted equational logic, *SIGPLAN Not.* **17**(1), 9–17.

Hales, T. C.: 2001, The honeycomb conjecture, *Discrete and Computational Geometry* **25**, 1–22.

Harris, J.: 1996, From das capital to dos capital: A look at recent theories of value, *Technical report*, Chicago Third Wave Study Group.

Hayek, F. A.: 1935, *Prices and Production*, Routledge, London.

Hayek, F. A.: 1945, The use of knowledge in society, *American Economic Review* pp. 519–530.

Hayek, F. A.: 1955, *The Counter-Revolution of Science*, The Free Press, New York.

Hodges, A.: 1983, *Alan Turing the enigma*, Touchstone.

Ingham, G.: 2004, *The Nature of Money*, London.

Innes, A. M.: 1913, What is money?, *Banking Law Journal* pp. 377–408.

Itoh, M. and Lapavitsas, C.: 1999, *Political Economy of Money and Finance*, Macmillan.

Kalecki, M.: 1954, *Theory of Economic Dynamics*, Rinehart and Company Inc., New York.

Kalecki, M.: 1971, *Selected Essays on the Dynamics of the Capitalist Economy*, Cambridge University Press.

Kapur, J. N.: 1989, *Maximum Entropy Models in Science and Engineering*, Wiley Eastern Limited.

Kapur, J. N. and Kesavan, H. K.: 1992, *Entropy optimization principles with applications*, Academic Press Inc.

Keynes, J. M.: 1920, *The Economic Consequences of the Peace*, Harcourt Brace and Howe.

Keynes, J. M.: 1925, *The economic consequences of Mr Churchill*, Hogarth Press.

Keynes, J. M.: 1936, *The General Theory of Employment Interest and Money*, Macmillan, London.

Khinchin, A. I.: 1949, *Mathematical foundations of statistical mechanics*, Dover Publications.

Knapp, G. F.: 1973, *The State Theory of Money*, Augustus M. Kelley, Clifton.

Kuhn, T.: 1970, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.

Landauer, R.: 1961, Irreversibility and heat generation in the computing process, *IBM Journal of Research and Development* **5**, 183–91.

Lange, O.: 1938, *On the Economic Theory of Socialism*, University of Minnesota Press.

Lavington, S.: 1980, *Early British Computers*, Manchester University Press, Manchester.

Lavington, S. H.: 1978, The manchester mark i and atlas: a historical perspective, *Commun. ACM* **21**(1), 4–12.

Lawlor, M. S. and Horn, B. L.: 1992, Notes on the Sraffa–Hayek Exchange, *Review of Political Economy* **4**.

Lee, Y., Amaral, L. A. N., Canning, D., Meye, M. and Stanley, H. E.: 1998, Universal features in the growth dynamics of complex organizations, *Physical Review Letters* **81**(15), 3275–3278.

Levy, M. and Solomon, S.: 1997, New evidence for the power-law distribution of wealth, *Physica A* **242**, 90–94.

Levy, M. and Solomon, S.: n.d., Of wealth power and law: the origin of scaling in economics. citeseer.nj.nec.com/63648.html.

Li, K.-W.: 2003, China's Capital and Productivity Measurement, *Technical Report 851*, Economic Growth Center, Yale, P.O. Box 208269 New Haven, CT 06520-8269.

Marx, K.: 1954, *Capital*, Vol. 1, Progress Publishers, Moscow. Original English edition published in 1887.

Marx, K.: 1971, *Capital*, Vol. 3, Progress Publishers, Moscow.

Matteo, T. D., Aste, T. and Hyde, S. T.: 2004, *Exchanges in complex networks: income and wealth distributions*, IOS Press, Amsterdam.

Maxwell, J. C.: 1875, *The Theory of Heat*, London.

Michaelson, G., Cockshott, W. P. and Cottrell, A. F.: 1995, Testing marx: some new results from uk data, *Capital and Class* pp. 103–129.

Mirowski, P.: 1989, *More Heat Than Light: Economics as Social Physics, physics as Nature's Economics*, Cambridge University Press.

Montroll, E. W. and Shlesinger, M. F.: 1983, Maximum entropy formalism, fractals, scaling phenomena, and 1/f noise: tale of tails, *Journal of Statistical Physics* **32**, 209–230.

Nirei, M. and Souma, W.: 2003a, Income distribution and stochastic multiplicative process with reset events. http:// www.santanfe.edu/ makato/ papers/ income.pdf.

Nirei, M. and Souma, W.: 2003b, Income distribution and stochastic multiplicative process with reset events. http:// www.santanfe.edu/ makato/ papers/ income.pdf.

Norretranders, T.: 1998, *The User Illusion*, Harmondsworth: Penguin.

Nove, A.: 1983, *The Economics of Feasible Socialism*, George Allen and Unwin, London.

Ochoa, E. M.: 1989, Values, prices, and wage–profit curves in the us economy, *Cambridge Journal of Economics* **13**, 413–29.

Okishio, N.: 1961, Technical change and the rate of profit, *Kobe University Economic Review*, **7**, 86–99.

Okishio, N.: 1990, Constant and variable capital, *in* J. Eatwell, M. Milgate and P. Newman (eds), *The New Palgrave – Marxian Economics*, W. W. Norton and Company, New York and London, pp. 91–103.

Ormerod, P.: 2002, The us business cycle: power law scaling for interacting units with comple internal structure, *Physica A* **314**, 774–785.

Ormerod, P. and Mounfield, C.: 2001, Power law distribution of the duration and magnitude of recessions in capitalis economies: breakdown of scaling, *Physica A* **293**, 573–582.

Pan, S. and Park, A.: 1998, Collective ownership and privatization of china’s village enterprises, *William Davidson Institute Working Papers Series 141*, William Davidson Institute at the University of Michigan Stephen M. Ros Business School. available at http://ideas.repec.org/p/wdi/papers/1998-141.html.

Pashukanis, E. B.: 1989, *Law and Marxism: A General Theory Towards a Critique of the Fundamenta Juridical Concepts*, Pluto Publishing, Worcester.

Pasinetti, L. L.: 1977, *Lectures on the theory of production*, Columbia University Press, New York.

Petrovic, P.: 1987, The deviation of production prices from labour values: some methodolog and empirical evidence, *Cambridge Journal of Economics* **11**, 197–210.

Polanyi, K., Arensberg, C. and Pearson, H.: 1957, *Trade and Market in the Early Empires*, Free Press, new York.

Porter, A. W.: 1946, *Thermodynamics*, Methuen, London.

Reed, W. J.: 2000, The pareto law of incomes – an explanation and an extension.

Reed, W. J.: 2001, The pareto, zipf and other power laws, *Economics Letters* **74**, 15–19.

Roemer, J.: 1982, *A General Theory of Exploitation and Class*, Harvard University Press, Cambridge.

Roemer, J.: 1986, *Should Marxists be interested in exploitation?*, Cambridge University Press, Cambridge.

Rubin, I. I.: 1973, *Essays on Marx's Theory of Value*, Black Rose Books. Russian edition published in 1928.

Samuelson, P. A.: 1973, Reply on marxian matters, *Journal of Economic Literature* **11**, 64–68.

Shaikh, A. M.: 1984, The transformation from marx to sraffa, *in* E. Mandel and A. Freeman (eds), *Ricardo, Marx, Sraffa – the Langston Memorial Volume*, Verso, London, pp. 43–84.

Shaikh, A. M.: 1998, The empirical strength of the labour theory of value, *in* R. Bellofiore (ed.), *Marxian Economics: A Reappraisal*, Vol. 2, Macmillan, pp. 225–251.

Shannon, C.: 1948, A mathematical theory of communication, *The Bell System Technical Journal* **27**, 379–423 and 623–56.

Smith, A.: 1974, *The Wealth of Nations*.

Software, N. R.: 1988, *Numerical Recipies in C*, Cambridge University Press.

Souma, W.: 2000, Physics of personal income, *in* H. Takayasu (ed.), *Empirical science of financial fluctuations: the advent of econophysics*, Nihon Keizai Shimbun, Tokyo. http://arxiv.org/abs/cond-mat/0202388.

Sraffa, P.: 1960, *Production of commodities by means of commodities*, Cambridge University Press, Cambridge.

Stanley, M. H. R., Amaral, N. L. A., Buldyrev, S. V., an H. Leschhorn, S. H., Maass, P., Salinger, M. A. and Stanley, H. E.: 1996, Scaling behavior in the growth of companies, *Nature* **379**, 804–806.

Ste-Croix, G. E. M.: 1981, *Class Struggle in the Ancient Greek World*, Duckworth.

Steedman, I.: 1981, *Marx after Sraffa*, Verso, London.

Steindl, J.: 1952, *Maturity and Stagnation in American Capitalism*, Oxford University Press, Oxford.

Strand, S.: 1979, *Machines, an Illustrated History*, AB Nordbok, Göteborg.

Szilard, L.: 1964, On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings, *Behavioral Science* **9**, 301–310.

Tann, J.: 1981, *Selected Papers of Boulton and Watt*, London.

Traiger, I. L., Gray, J., Galtieri, C. A. and Lindsay, B. G.: 1982, Transactions and consistency in distributed database systems, *ACM Trans. Database Syst.* **7**(3), 323–342.

Turing, A.: 1937, On Computable Numbers, With an Application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society* **42**, 230–65.

Velupillai, K.: 2003, Essays on Computable Economics, Methodology and the Philosophy of Science, *Technical report*, Universita' Degli Studi di Trento - Dipartimento Di Economia.

von Glahn, R.: 1996, *Fountain of Fortune: Money and Monetary Policy in China, 1000-1700*, Berkeley.

von Glahn, R.: 2004, Revisiting the song monetary revolution: a review essay, *International Journal of Asian Studies* **1**, 159–178.

Wells, J.: 2001, What is the distribution of the rate of profit?, *IWGVT mini-conference at the Eastern Economic Association*, New York NY.

Wolfram, S.: 2002, *A New Kind of Science*, Wolfram Media, Inc., Illinois.

Wray, R.: 2004, The credit money and state money approaches, *Credit and State Theories of Money: the contributions of A. Mitchell Innes*, Edward Elgar, Cheltenham, pp. 79–98.

Wright, I.: 2003a, The duration of recessions follows an exponential not a power law. Submitted for publication.

Wright, I.: 2003b, Simulating the law of value. Submitted for publication.

Wright, I.: 2004, A conjecture on the distribution of firm profit. To appear in Economía: Teoría y P'ráctica.

Zachariah, D.: 2006, Labour value and equalisation of profit rates, *Indian Development Review* **4**(1), 1–21.

Ziv, J. and Lempel, A.: 1978, Compression of individual sequences via variable-rate coding, *IEEE Trans. Inform. Theory* **24**(5), 530–536.