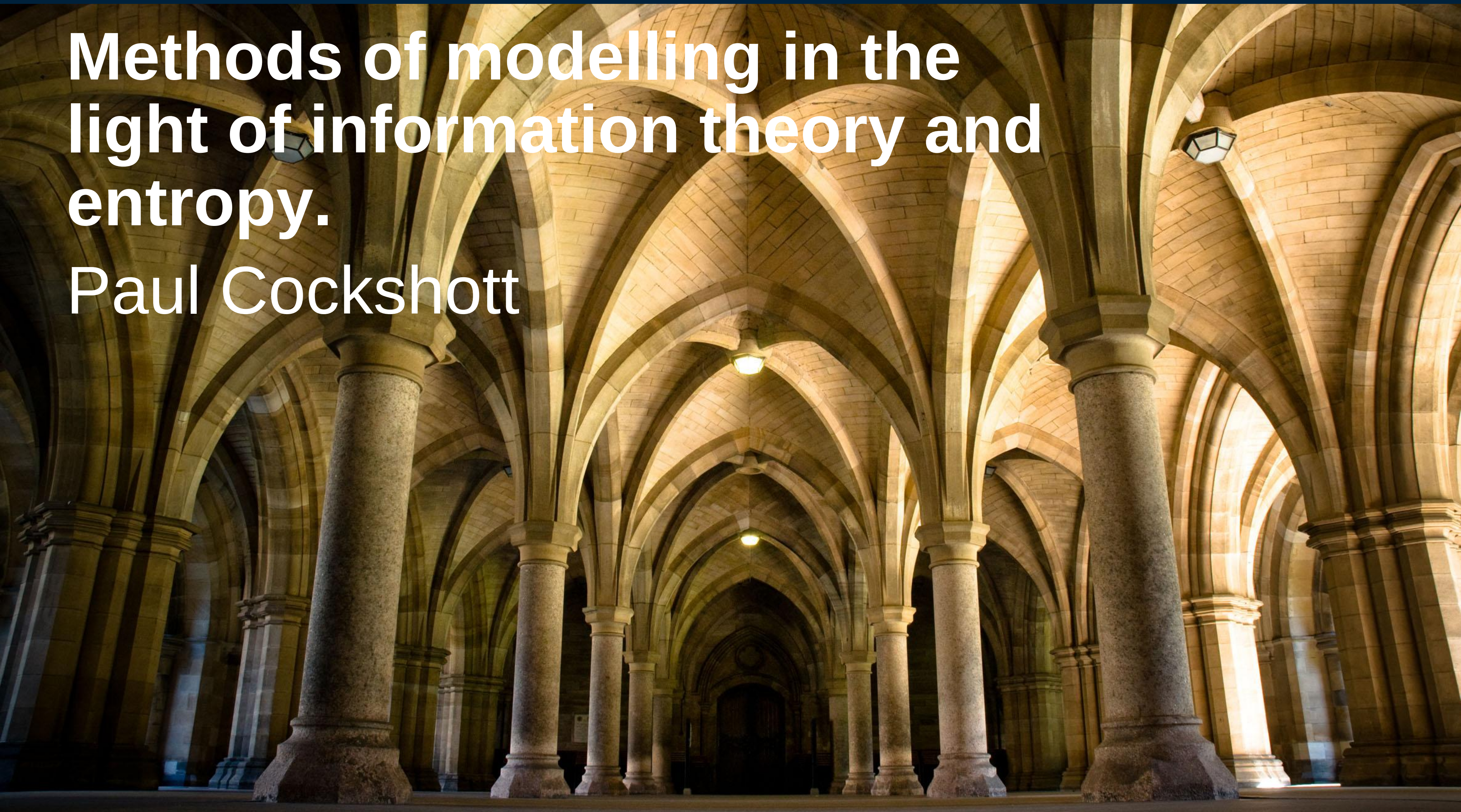
Methods of modelling in the light of information theory and entropy.

Paul Cockshott



A scientific model is a machine for predicting how part of reality will behave.

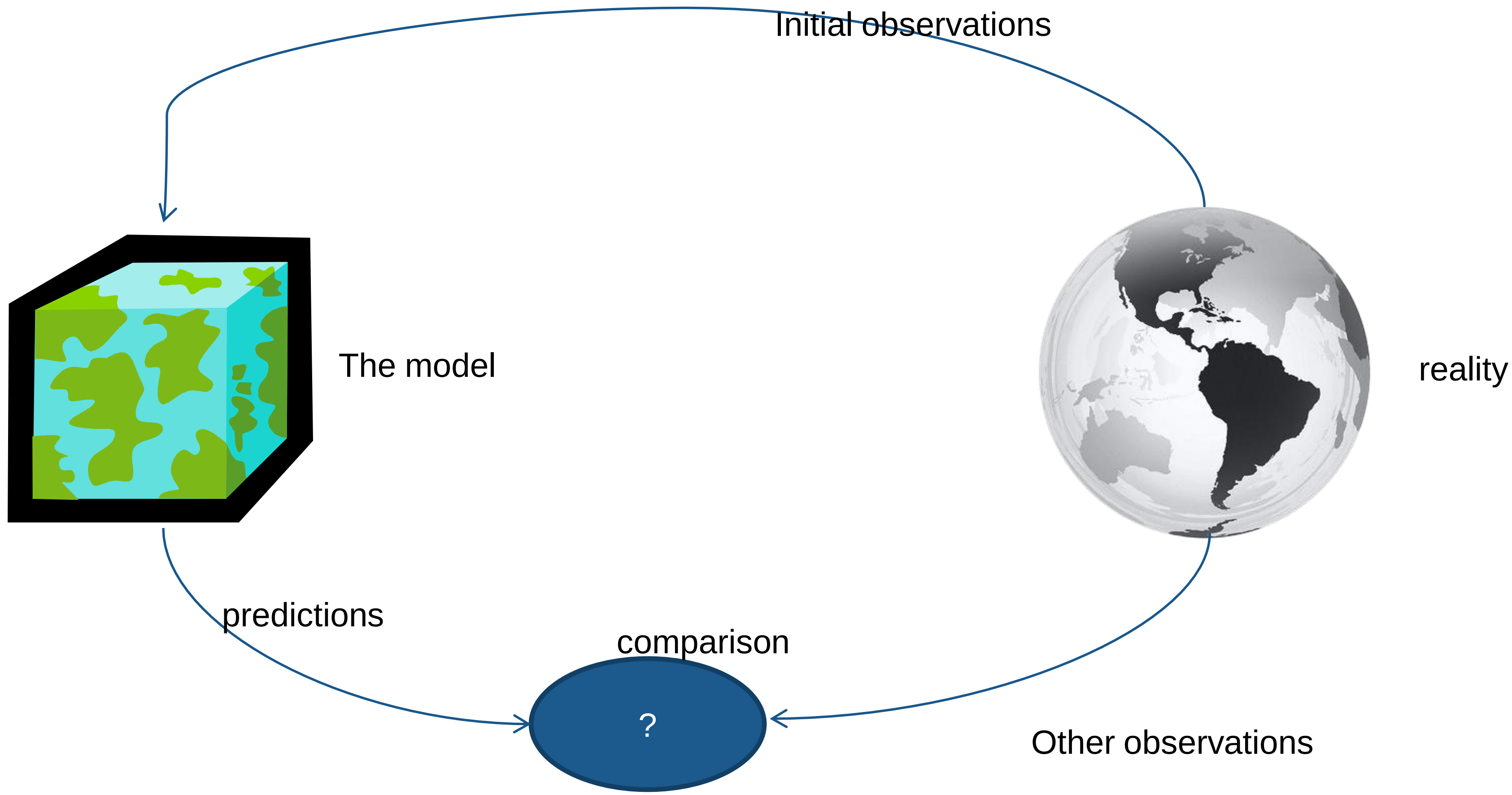
We tend to think of models in an abstract conceptual sense, I want to argue that we should look at them in a very concrete material sense.

I will present historical examples, and then formulate criteria for the ‘goodness’ of models, before applying these general principles to economic modeling.

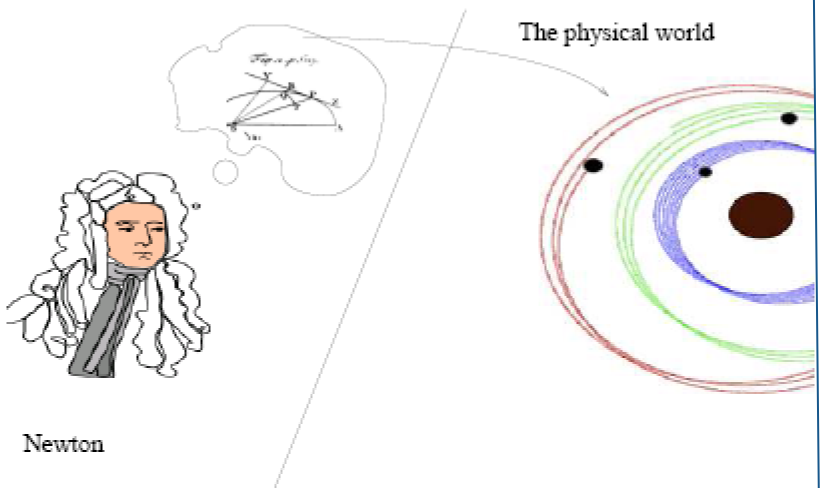
The whole approach is very computational.

University of Glasgow

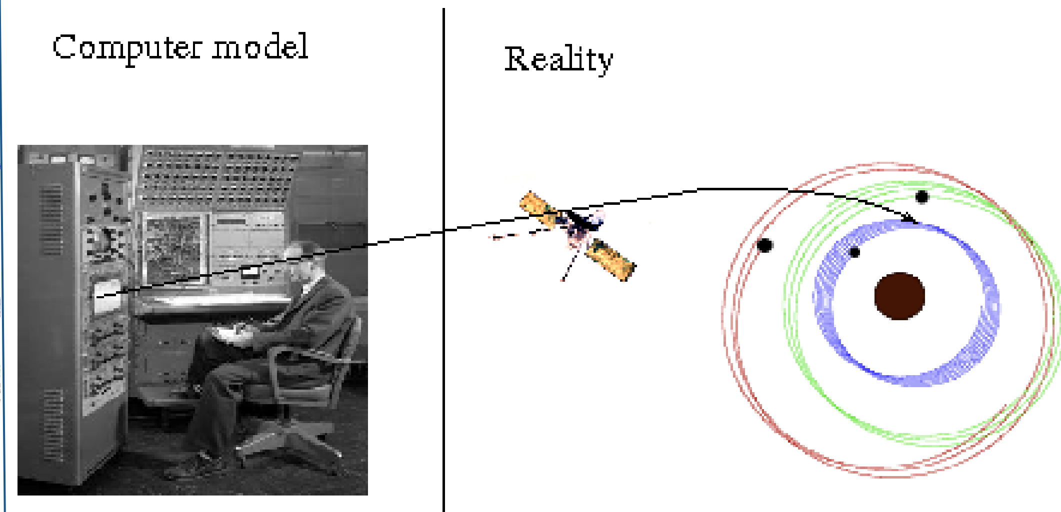
- # The basic modelling process



- **Are models ‘ideas’ or are they machines?**



Newton Thinks….



NASA computes …

- **But were they thoughts?**

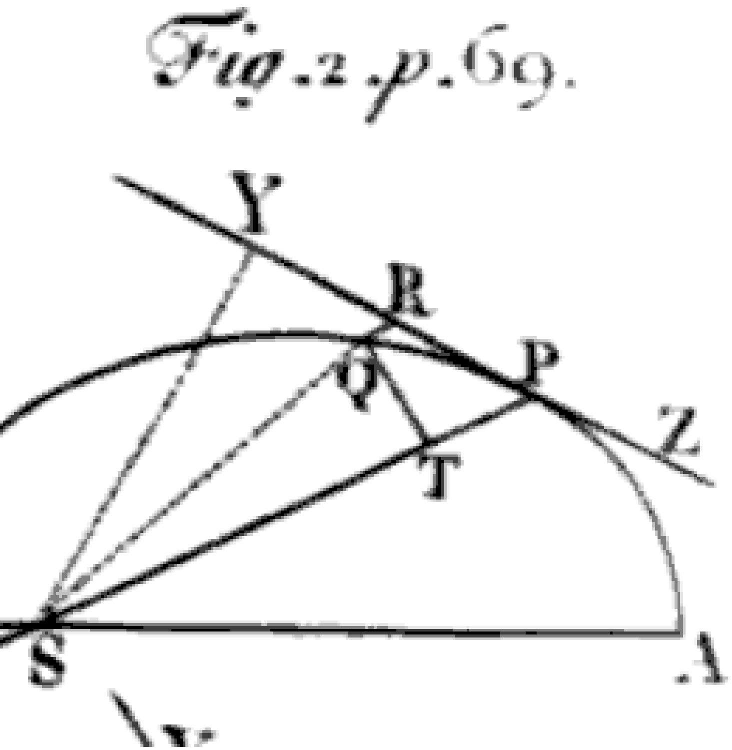
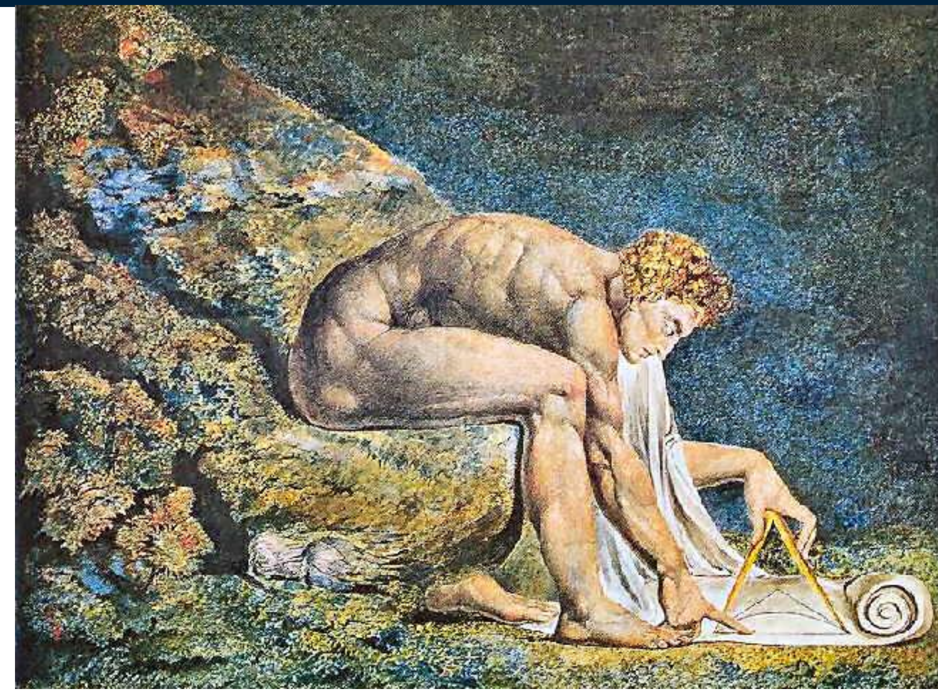


Illustration from the *Principia*



Blake's 'Newton'

Or were they always something material, produced by physical work using physical tools?

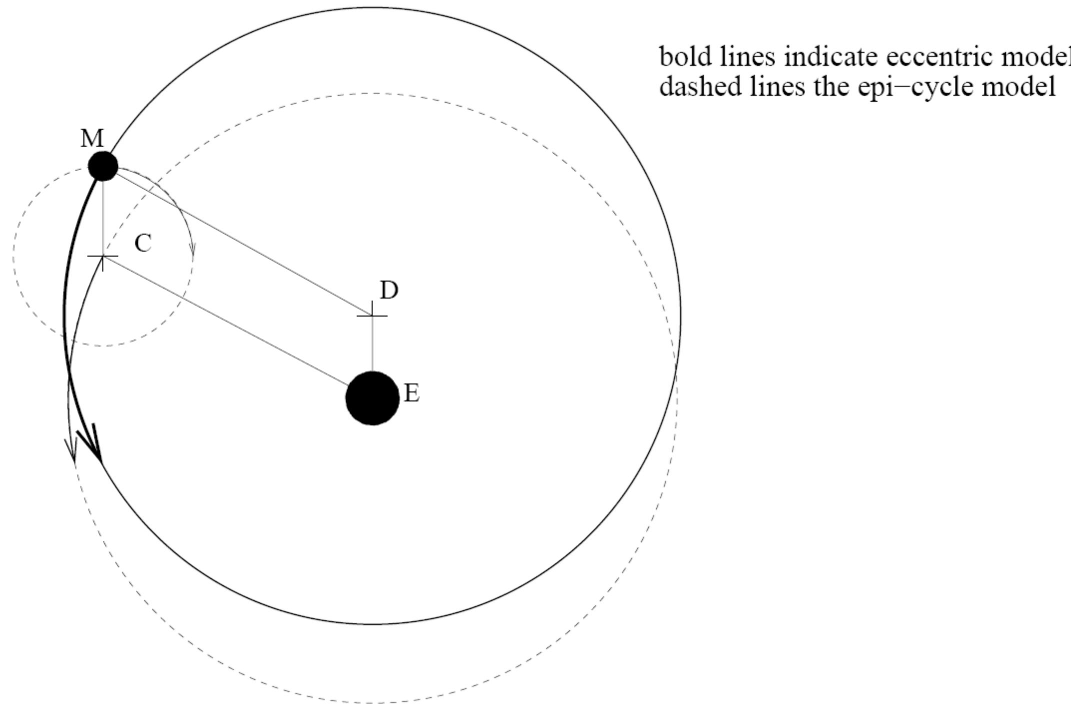
Ab Initio

– Before Newton came Kepler, before him Ptolemy

– Before Ptolemy came Hiparchus and Appolonius

Ptolemy’s epi-cycle model is well known, but it is equivalent to Appolonius’s Cycle and Deferent Model

Comparison of Ptolemy and Apollonius



Hipparchus's actual model?

In 1900 a group of sponge divers sheltering from a storm anchored off the island of Antikythera.

Diving from there they spotted an ancient shipwreck with bronze and marble statuary visible.

Further diving in 1902 revealed what appeared to be gearwheels embedded in rock. On recovery these were found to be parts of a complicated mechanism, initially assumed to be a clock. Starting in the 1950s and going on to the 1970s the work of Price established that it was not a clock but some form of calendrical computer.

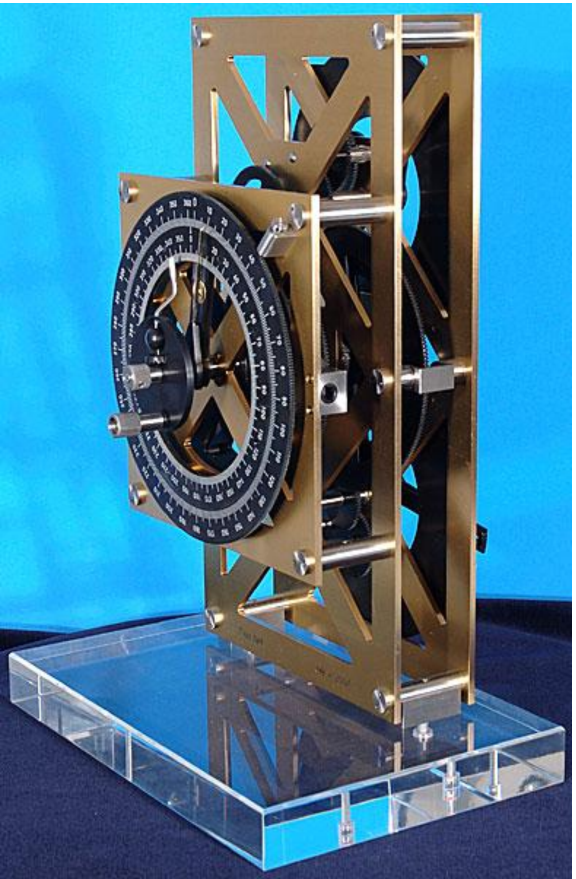
Using X-rays, modern reconstructions have been built showing that it physically implemented Appolonius's model of the lunar orbit.

The original machine dates from the 2nd century BC but modern reconstructions have been built.

I show a particularly beautiful one by Tania van Vark.

You turn the handle and get predictions of the position of the sun and moon in the sky and the dates of eclipses.

It emphasises how a scientific model is a *microcosm* emulating a *macrocosm*.



- **Since the invention of the Universal Computer in the 1940s, it was no longer necessary to build special purpose mechanical models of physical system.**
- **A universal computer is a physical device that can be configured to simulate any physical process.**
- **It is configured by the input of an appropriate mathematical function representing the model.**

Replica of the first Universal Computer, with two of its builders at the 50th anniversary in 1998

Deutsch's principle

'Every finitely realizable physical system can be perfectly simulated by a universal model computing machine operating by finite means'.

Proceedings of the Royal Society of London A **400**, pp. 97-117 (1985)

Key principles of modelling

- **Generation of testable predictions**

– A model which makes no testable predictions is useless

- **Elegance or simplicity**

– Occam’s Razor ‘Entities should not be multiplied without cause’.

- **Mathematically we can view any use of a model as**

$$(p,d) \leftarrow M(d)$$

- **Where *p* are the predictions, *M* is the function encoding the model, and *d* are the input data.**
- **After running the model we have both the predictions and the original data.**
- **For the model to be elegant we want to maximise its information yield**

$$Y = I(p)/I(M)$$

- **Where I(x) means the information content of x**

For the model to be useful we want to maximise the mutual information in the predictions *p*, and observed system *o*.

$$\textbf{Max}\ I(p;o) = H(p) - H(p|o)$$

Where $H(p)$ is the uncertainty or entropy in p and $H(p|o)$ is the uncertainty in p given o

Whilst minimising the information in the model *I(M)*

That is to say we should avoid models that contain a lot of internal information – in the worst case such a model simply tabulates the observations.

Why entropy?

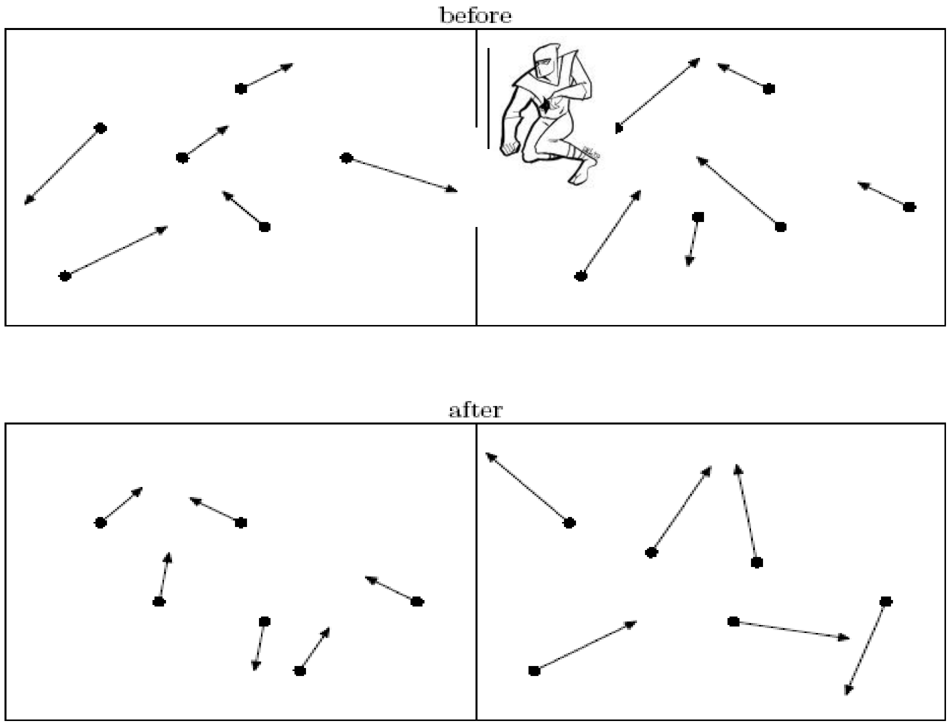
- **In the formula to find mutual information we used the H function for entropy. Why?**
- **Surely entropy has to do with thermodynamics which studies things like the efficiency steam engines?**
- **Yes, but a key discovery of the 20th century was how information and entropy are linked.**

The model Newcomen steam engine, now in the Hunterian Museum Glasgow University, on which James Watt worked in 1765, and from which his invention of the separate condenser came.

The Basic Problem of Information

- What is information?
- How does it relate to entropy?

- Clausius – no process possible that has the sole effect of transfering heat from a colder to a hotter body.



Gas initially in equilibrium. Daemon opens door only for fast molecules to go from A to B, or slow ones from B to A. ! Slow molecules in A, fast in B. B hotter than A, and can be used for power.

Information has produced power!

Boltzmann

- **Maxwell's proposed counter-example to the second 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*.**

$$S = -k \int f(v) \log f(v) dv$$

- **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.**

Shannon

The communications engineer Shannon introduces the concept of entropy as being relevant to sending messages by teletype.

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

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, this is essentially the same as Boltzmann’s formula.

Hence information = entropy.

Common measure

Information measured in bits provides a common means of measuring both the model M and the predictions p

- Suppose we have a vector of observations O which we have reason to believe are given to an accuracy of 3 digits. Then each observation contains
- $\log_2 (1000)$bits ≈ 10 bits
- Suppose we have a prediction vector P which we assume is to the same accuracy.

We can estimate H(P|O) by histograming the distribution of the ratio P/O and then applying Shannon's entropy formula

$$H(S) = -\sum p_i \log_2 p_i,$$

to the distribution

Information content of the model itself

If we want to compare models, we can decompose each of the models into two parts

1. A basic structure or formula
2. A set of auxiliary parameters or constants that has to be provided

Each of these can be given an information measure. The formula is measurable in terms of the number of bits needed to write it down as a string of digital characters.

The parameters are measurable in terms of the number of parameters and the accuracy in bits to which each has to be given.

Models and Laws

Sciences designate as laws those models that:

1. **Have a simple, elegant formulation with few parameters**
2. **Make excellent predictions in an apparently unlimited number of cases**

Applying this to economics

One may ask how much of what is taught in undergraduate economics consists of

a) Empirically testable and empirically tested propositions

b) Formulae that are elegant and simple

c) Simple formulae that are so universal and excellent in their predictive power as to deserve the name Laws.

- **There is obviously an immense wealth of empirical studies in the economics literature.**
- **But I am more concerned here with the basic theory that is taught to students starting economics. What these students would be learning when a physics student would be learning classical mechanics, or when an electrical engineer would be learning Maxwell’s equations.**
- **How well founded are the models that the economics students are taught?**

Testability – neoclassical value theory

- **Bear in mind that I am speaking as an outsider as I only studied neo-classical economics to undergraduate level.**
- **At that time I was told that the labour theory of value was now known to be inaccurate and superseded by the subjective utility theory.**
- **But it is hard to see how the subjective theory is even testable. What quantitatively testable predictions about the price structure of the whole economy can it make?**
- **Can one derive from it a predicted price vector to compare with the actual price vector?**
- **If not, one has to put it in a basket labelled '***not even wrong***'.**

Testability – classical value theory

The classical theory of value on the other hand does make testable predictions. Once can make concrete predictions about market prices, using either Sraffa's formula

$$p = (1 + r)Ap + w$$

Or the formula from Marx's Capital

$$v = Av + \lambda$$

And you can then see how good the predictions each one makes are.

- **Until the 1980s it has to be said that the proponents of classical theories were as axiomatic in their approach as the proponents of neo-classical theories.**
- **Since then however there has been a growing realisation that not only were classical theories testable, but that they should be rigorously tested if any progress was to be made in resolving theoretical disputes.**

- **It was found, somewhat to the surprise of everyone, that the two leading classical theories gave almost the same accuracy in their predictions, in which case parsimony may favour the simpler model.**
- **I hope that later speakers will touch on this showing in practice how to the labour theory of value can be tested using an information theoretic measure.**

Idea of a conservation law.

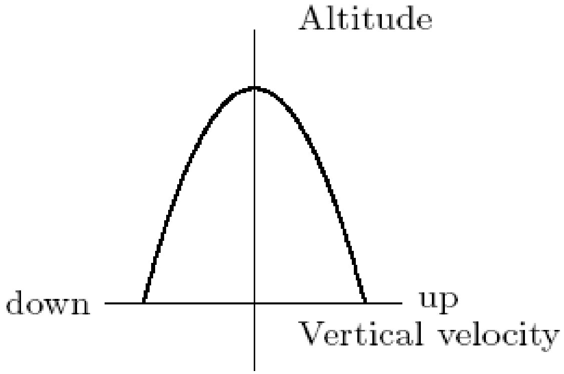
- I said that the most powerful models for the predicting of reality are called laws. Among the most paradigmatic here are conservation laws. In physics conservation of energy, charge, probability.
- Conservation laws reveal hidden symmetries in the structure of reality. Noether's theorem states that any symmetry of the action of a physical system has a corresponding conservation law.
- The symmetry and the conservation are not necessarily immediately evident as the following examples will show.

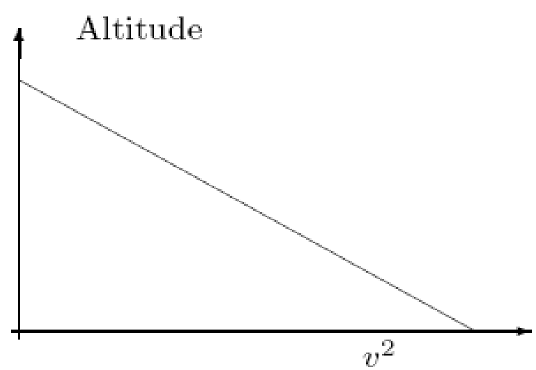
University of Glasgow

Put very loosely, Emmy Noether's principle says that in a physical system, a conserved quantity at one level of abstraction corresponds to a symmetry property of the system at another level of abstraction.

– Translational symmetry implies conservation of momentum
– Temporal symmetry implies conservation of energy

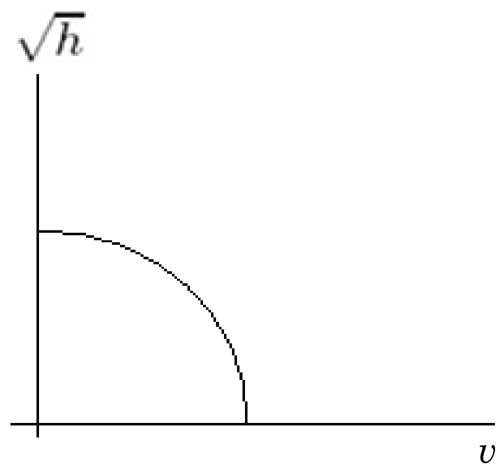
Example





- **Top shows points in phase space traversed by a projectile thrown upward in a gravitational field**
- **Bottom, points in the space of (altitude, velocity squared) traversed by the particle**
- **Bottom diagram shows conservation of energy**

Find the symmetry



- **If we transform to yet another space we find the path is a circle.**
- **Here we can see the symmetry associated with the conservation of energy.**
- **The new representation means that we can treat the path as the result of rotational symmetry in a vector space.**

Another example is quantum mechanics

Why do we use amplitudes whose square gives us probabilities?

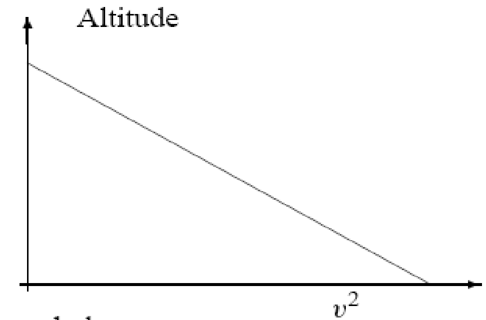
- **Because this vector space, which allows the application of unitary rotation operators, projects onto the space of probabilities which are a scalar conserved property.**
- **This is the same as the change of representation we had to illustrate conservation of energy.**

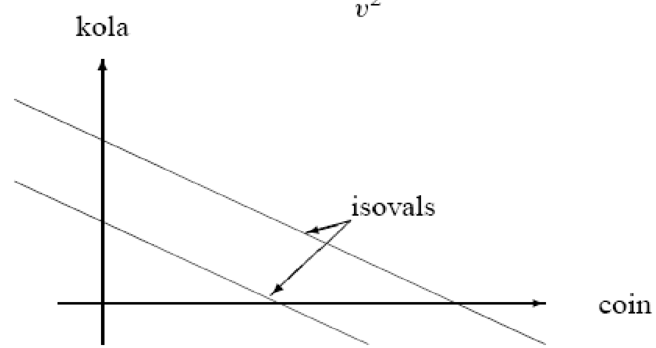
Similarity between worth/wert and

- **Commodity money space is not a vector space, since the metric it follows is**

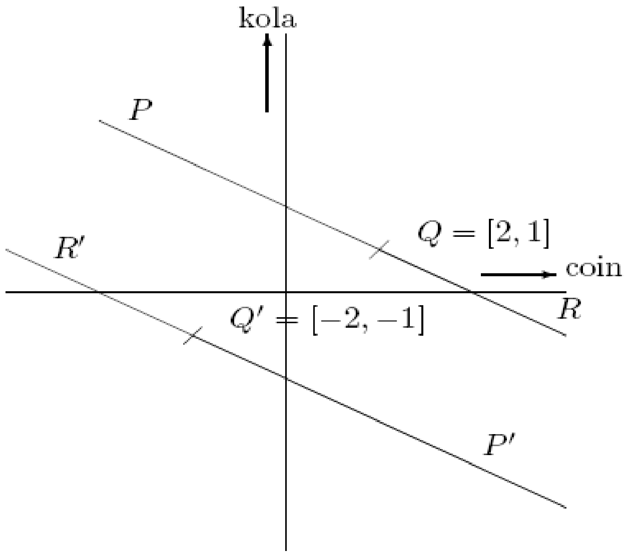
$$d = |\alpha \Delta_x + \beta \Delta_y|$$

- **This makes it analogous with a system with a hidden conserved quantity**





Debt



- The unit circle in commodity space is composed of disjoint hyperplanes.
- One is the set of positions reachable by a net creditor the other by a net debtor.
- Discontinuity indicates net debtor can never become net creditor by conservative operations.

- Suppose it was and that coin exchange for Kola at 1 to 1, then do the following

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.

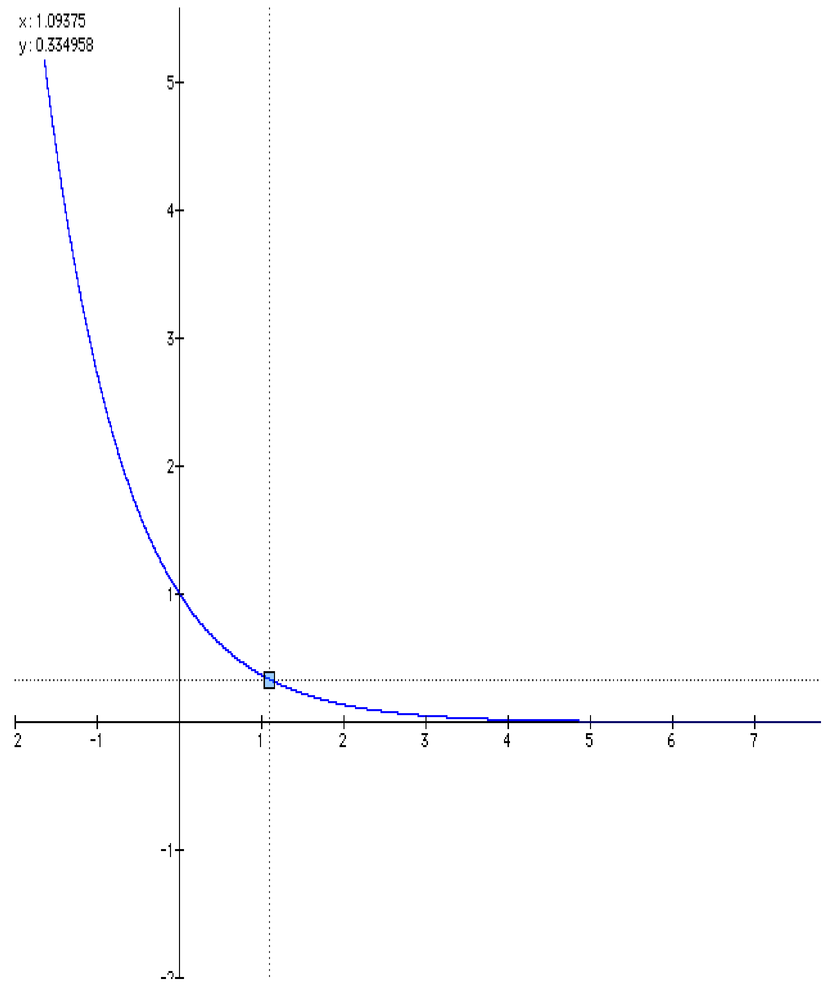
- I end up with more kola than I had at the start, so this cannot be a set of equivalent exchanges. The second step is illegal within the context of the Euclidean metric, since it involves operating upon one of the coordinates independently. But in the real world, commodities are physically separable, 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.

- We have developed 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.
- I would argue that the application of Noether's theorem gives us a more rigorous way of formulating what Marx was arguing in the chapter on the forms of value in volume I of capital. These should be seen as an attempt to define commodity exchange and exchange value as a conservative system.

Applying conservation laws and entropy to wealth distribution

- Thermodynamics predicts that systems tend to settle into a state of maximum entropy.
- The conservation laws specify that whilst this occurs energy must be conserved.
- Boltzmann showed that this implies that the probability distribution of energies E_i that meets these two criteria is

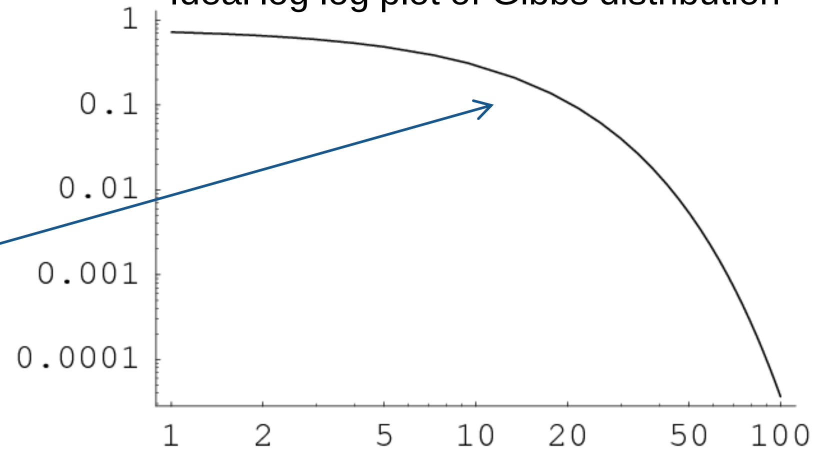
$$p_{E_i} = e^{\frac{E_i}{kT}}$$

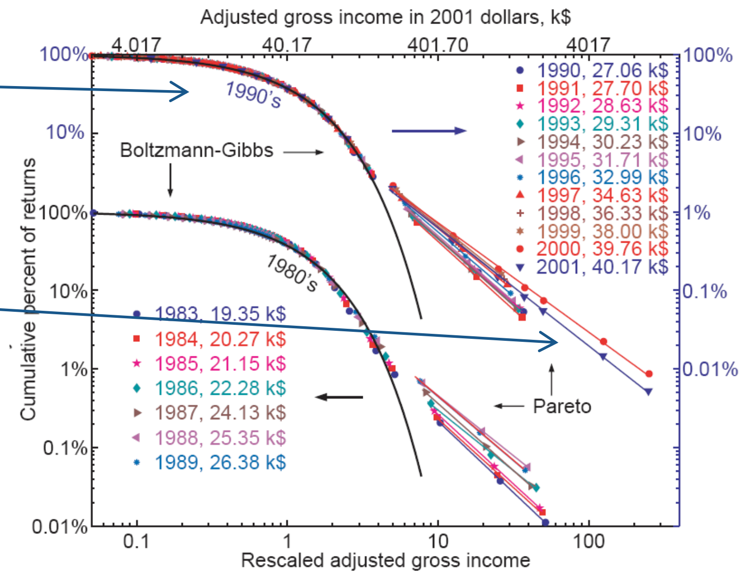


Yakovenko has argued that since money is conserved in the buying and selling of commodities it is analogous to energy.

- If the system settles into a maximum entropy state then monetary wealth will come to follow a Gibbs Boltzmann distribution.
- He is able to show that the observed income distribution for 96% of the US population is well explained by a negative exponential distribution of the Gibbs form.
- There remains a superthermal tail of income (the top 4%) whose income is not conformant with maximal entropy but follows a power law distribution.

Ideal log log plot of Gibbs distribution





Non thermal distribution

- Thermal distribution arises from the application of the conservation law plus randomness.
- Non thermal distribution from the violation of conservation. Tied to income from capital and the stock market.
- Consistent with Marx's analysis that profit in general can not arise within a conservative system, but from something outside of the conservative system – production of surplus value.
- Wright has shown that random exchange models generate combined Gibbs + power law distributions as soon as you allow the hiring of labour.
- This is again consistent with Marx's old analysis.

University of Glasgow

Summary

- A model must make testable predictions to be scientifically meaningful.
- Information theory gives us a uniform means to measure both models and the predictions of models in order to evaluate their adequacy.
- One should not be afraid to make use of it.
- The results of tests using modern methods are consitent with broadly marxian models.