

# Thermoeconomics: An axiomatic theory of aggregate economic phenomena

Nick Chater<sup>a</sup>, R.S.MacKay<sup>b\*</sup>

<sup>a</sup>Warwick Business School, <sup>b</sup>Mathematics Institute,  
University of Warwick, Coventry CV4 7AL, UK  
Nick.Chater@wbs.ac.uk, R.S.MacKay@warwick.ac.uk

October 6, 2023

## Abstract

An axiomatic approach to macroeconomics based on the mathematical structure of thermodynamics is presented. This approach deduces relations between aggregate properties of an economy, concerning quantities and flows of goods and money, prices, and the value of money, without any recourse to ‘microeconomic’ foundations about the preferences and actions of individual economic agents.

The approach has three important payoffs. Firstly, it provides a new, and more solid, foundation for some aspects of standard macroeconomic theory, such as the existence of market prices, the value of money, inflation, the symmetry and negative-definiteness of a macro-version of the Slutsky matrix, and the Le Chatelier-Samuelson principle, without relying on implausibly strong rationality assumptions over individual microeconomic agents. Secondly, the approach generates new results, including implications for money flow and trade when two or more economies are put in contact, in terms of new concepts such as economic entropy, economic temperature, goods values, and money capacity. We will see that some of these are related to standard economic concepts (e.g. marginal utility of money, market prices), yet our approach derives them at a purely macroeconomic level and gives them a meaning independent of usual restrictions, e.g. to rational agents. Others of the concepts, such as entropy and temperature, have no direct counterparts in standard economics, though they have important economic interpretations and implications. Thirdly, this analysis promises to open up new frontiers by building a bridge to ideas from non-equilibrium thermodynamics. In contrast to the strict economic notion of equilibrium, the present approach starts from a statistical physics notion of equilibrium and thereby has the potential to extend to the more realistic non-equilibrium settings, as in physics. More broadly, we hope that the economic analogue of entropy (governing the possible transitions between states of economic systems) may prove to be as fruitful for the social sciences as entropy has been in the natural sciences.

We restrict attention to “extensive” economic systems: those which can be subdivided into parts with the same properties apart from size (for example, we ignore possibly “lumpiness” that might be caused by large individual companies). We also restrict attention to exchange economies, leaving production, consumption, labour etc. for future treatment. Building on an axiomatic approach to thermodynamics by Lieb and Yngvason, we explain how extensivity plus plausible economic assumptions lead to the definition of an economic analogue of entropy, which we consider to be an aggregate utility, and an economic temperature, whose inverse is the marginal (aggregate) utility of money. We obtain a “second law” of economics, governing the allowed direction of changes of economic state. This approach captures the ‘value’ of nominal money in a way that does not rely on the usual rather arbitrarily chosen “baskets of goods” and their prices. It explains the direction of financial flow in terms of temperature differences. We illustrate the theory with an economic analogue of the Carnot cycle whereby a trader makes money out of temperature differences. We obtain quantitative conditions for trade. On making contact, nett trade flows occur if and only if they increase the total entropy.

We highlight the subset of “mutually beneficial” trades, which increase the entropy of each economy. We derive some macroeconomic relations between cross-derivatives. We introduce the concept of “pure money”. We discuss limitations and possible extensions of our approach, in particular to relaxing the extensivity assumption and incorporating important realistic aspects like production, consumption and manufacturing.

## 1 Introduction

Economies appear to exhibit many large-scale regularities, concerning the balance between supply and demand to clear markets, patterns of interdependence between prices and volumes of different types of goods, the value of

---

\*Order of authorship is alphabetical.

money and the movement of exchange rates, and many more. Yet economies also have many irregular properties—including crashes and booms, gluts and shortages, herding, fashion, and unpredictability due not just to technological or organizational innovation, but apparently due to endogenous, and little understood, fluctuations at many scales.

During the twentieth century, economists have primarily attempted to explain the mix of regularity and irregularity in aggregate properties of the economy through attempts to establish micro-foundations, i.e., explaining aggregate properties as arising from the interaction of economic agents with specified properties. Roughly speaking, the interactions of fully rational and informed consumers and producers each aiming to maximize their own utility are presumed to generate regularities, most notably through the concept of general equilibrium. In parallel, there have been attempts to explain irregular aspects of economic behaviour by exploring how partial information, and bounded rationality, can lead, for example, to booms and crashes. Thus, the pattern of explanation has been bottom-up: from the properties of economic agents and their interactions to the large-scale economic phenomena.

There is, however, an interesting complementary approach, which does not attempt to provide such micro-foundations, but instead makes assumptions about economic behaviour in the aggregate, and explores their consequences at the same aggregate level. This approach has the potential advantage that it may be possible to abstract away from detailed “micro” assumptions (e.g., about the psychological properties of consumers, or the behaviour of companies), which are likely to be very difficult to specify and highly variable between contexts. Interestingly, this strategy predates the modern emphasis on the micro-foundations of economics, and is implicit in arbitrage arguments from Adam Smith onwards (see discussion in [Su]), and which are the basis for much of financial economics [V].

In this paper, we explore this non-micro-foundational approach by applying conceptual and mathematical methods borrowed from “classical” thermodynamics, developed in the nineteenth century by Carnot, Joule, Clausius, Maxwell and Gibbs (“classical” means without involving statistical mechanics [LY99]). Classical thermodynamics begins with high-level principles, such as the conservation of energy (the first law of thermodynamics) and the non-decrease of entropy (the second law), and derives systems of relationships between large-scale observable phenomena including work, pressure, volume, temperature, and so on. It requires the existence of a priori abstract quantities, such as energy and entropy, which provide a concise and elegant framework for expressing these laws. These fundamental laws turn out to be extremely general and to apply to systems with very diverse components. Thus, while in specific contexts there are elegant connections to microfoundations (e.g., to statistical properties of interacting gas particles), the principles of classical thermodynamics are more general and more abstract (e.g., [LY02, Go]) and apply to diverse physical systems, chemical reactions, and biological processes.

Can a similar approach work for economics? Specifically, is there a thermodynamics of economic systems that can derive economic regularities without making detailed assumptions about consumers and companies? We formulate such a theory here, by reinterpreting thermodynamic quantities in economic terms. This provides an interesting new perspective on some well-known economic regularities, captured by equilibrium arguments in conventional mathematical economics. Most importantly, we establish an analogue of the second law of thermodynamics, for economic systems, which specifies possible transitions between states of an economy, ruling out those with non-decreasing entropy. Moreover, this approach provides a principled meaning for the value of nominal currency (e.g., dollars, euros), which does not depend on the prices of somewhat arbitrarily chosen “baskets of goods”: the marginal utility of money is the inverse “temperature” of the economic system (which can be called its “coolness”). It also opens up the possibility of using generalizations of thermodynamics, developed in the natural sciences, to understand more complex economic phenomena (including the analysis of economic systems which are out of equilibrium) without recourse to microfoundational assumptions about lack of information or bounded rationality at the level of individual economic agents. If such a high-level thermodynamic approach to the economy proves fruitful, it has the potential to complement and inform traditional microeconomic analysis, just as in physics and chemistry classical thermodynamics has complemented and informed statistical mechanics.

We begin in Section 2 by reviewing precursors of using thermodynamics in economics. We restrict attention in this paper to exchange economies, reviewed briefly in Section 3. In Sections 4-8 we state our main assumptions and their justifications. In Section 9 we deduce the existence of an entropy function for economies and the all-important second law of non-decrease of entropy. We compute the entropy for a simple example in Section 10 and Appendix A. After stating a few technical assumptions in Section 11, we deduce existence of a temperature function for economies and its relation to financial equilibrium in Section 12. A subsidiary concept of money capacity is introduced in Section 13 and its significance described. In Section 14 we deduce the existence of a market price for each good, via a concept of “value” of a good, and a key relation for reversible trade. In Section 15 we introduce the concept of “pure money” and discuss some of its consequences. Section 16 proposes a way to measure the temperature of an economy. Section 17 discusses the natural idea of how to make money from price differences between two economies. Section 18 describes the less intuitive way to make money out of the temperature ratio between two economies. Section 19 gives quantitative conditions for trade and highlights the significance of mutually beneficial trades, which increase the entropy of all parties. Section 20 derives relations between various partial derivatives, based on the existence and concavity of entropy. Section 21 raises some further issues for exchange economies. Section 22 discusses potential directions for development beyond exchange economies. The paper concludes with a summary and conclusions in Section 23.

## 2 Precursors

Many previous researchers have considered possible links between classical thermodynamics and economics. Indeed, the architect of modern mathematical economics, Paul Samuelson, stressed the importance of Gibbs' work as foundational to his own. But Samuelson (and precursors such as Slutsky) applied principles of maximization at the level of individual agents, rather than aggregate system properties. Samuelson despaired of numerous ill-thought-out proposals to attempt to find an economic analogue of entropy and doubted that such an analogue existed.

The present approach aims to vindicate those who have intuitively felt the existence of a deep connection between thermodynamics and economics, and to set Samuelson's concerns aside. We begin by reviewing some of the most notable precursors of the approach developed here.

The idea that thermodynamics is relevant to economics goes back more than a century. Irving Fisher's 1891 PhD (considered by many to be the first in Economics) was co-supervised (along with the sociologist Sumner) by Gibbs, an early champion of thermodynamics. Utility theory, as developed by Fisher, Slutsky, Hotelling, Houthakker, Samuelson and others, shares many features with thermodynamics. For example, the theory contains economic analogues of some of the Maxwell relations between partial derivatives (known in economics as the Hotelling conditions [S50]).

Connections between economics and thermodynamics are particularly prominent in Samuelson's work (e.g. see the review of Samuelson's legacy by Dixit [Di]). Specifically, Samuelson notes that [S72] "Pressure and volume, and for that matter absolute temperature and entropy, have to each other the same conjugate or dualistic relation that the wage rate has to labor or the land rent has to acres of land" and "My earlier formulation of the inequality  $[\sum_i \Delta p_i \Delta v_i < 0]$  owed much to Wilson's lectures on thermodynamics."<sup>1</sup> Yet Samuelson was also scathing about attempts to make connections between the subjects [S72]: "There is really nothing more pathetic than to have an economist or a retired engineer try to force analogies between the concepts of physics and the concepts of economics". In the light of Samuelson's own inspiration from thermodynamics, this statement is rather puzzling. But it has certainly deterred researchers from the economic and natural sciences from pursuing such potential analogies further.<sup>2</sup> In particular, while Samuelson found important links between thermodynamics and maximization problems in economics (typically defined for an individual agent, but including an individual company), the present link is more fundamental: it derives (rather than "forcing") an economic notion of entropy for whole economies, and shows that economies, or systems or economies, can only change in ways that do not decrease their overall entropy.

An influential book, published by Georgescu-Roegen in 1971 [GR], did appear to make a link between entropy and economics. But our reading of Georgescu-Roegen's claim that entropy governs economics, has, we believe, a rather different aim than the project Samuelson was critiquing (and that we are pursuing here). Georgescu-Roegen's aim was to explore how physical entropy might place constraints on economic activity, rather than attempting to create an analogue of entropy within economics itself. This direction has been pursued under various names including ecological economics and systems ecology (for one take, see [Ku]).

Subsequently, physicists like Jaynes [J] and Farjoun and Machover [FM], began again to address economics, but from the viewpoint of statistical mechanics, rather than classical thermodynamics. This developed into a vigorous field of research that was christened "Econophysics" by Gene Stanley in 1996. Particularly relevant articles from this school in our opinion are [Sas] and [RFP], and the book [C+] on econophysics related to the Marxist view of economics, which centres on labour as measure of value. Econophysics has applied a wide variety of ideas and methods developed for analysing complex physical systems to economic phenomena and especially to financial markets. Yet econophysics has not generally found favour with economists.

Others have attempted to establish links between classical thermodynamics and economics, but like Samuelson, they typically focus on mapping thermodynamics onto the utility function of individual agents, rather than considering economies from a purely macro-perspective as here. For example, [SDa, SDb] aim to build links founded on the observation that both maximizing entropy and maximizing utility are problems of constrained optimization, and draw out a number of detailed formal parallels.<sup>3</sup> A particularly sophisticated bridge between the two fields is outlined by [SF], who focus on both micro and macro-level analysis and the connections between them. A more recent development by Cooper and Russell [CR01, CR11] uses ideas from Samuelson to help

---

<sup>1</sup>We will derive an analogue of this inequality in Section 20.

<sup>2</sup>[SF] pick out a particularly clear instruction from Samuelson [S60] to economists and physicists to desist from seeking what he saw as spurious parallels: "The formal mathematical analogy between classical thermodynamics and mathematic economic systems has now been explored. This does not warrant the commonly met attempt to find more exact analogies of physical magnitudes – such as entropy or energy – in the economic realm. Why should there be laws like the first or second laws of thermodynamics holding in the economic realm? Why should 'utility' be literally identified with entropy, energy, or anything else? Why should a failure to make such a successful identification lead anyone to overlook or deny the mathematical isomorphism that does exist between minimum systems that arise in different disciplines?" The present paper shows that there is indeed a powerful economic notion of entropy and a corresponding second law of thermoeconomics, which go beyond the link between thermodynamics and maximization that was so productive in Samuelson's work.

<sup>3</sup>They also highlight early discussion hinting at analogies between the equations used in economics and thermodynamics including [CMIM, Da, Lis].

formulate physical thermodynamics, drawing on ideas from symplectic geometry: thus, their primary focus has the opposite explanatory direction from the account developed here, and a narrower scope. The present purely macroscopic approach, aiming to map the full axiomatic structure of thermodynamics into a macro-analysis of economics, seems to be new—although exploring potentially fruitful links with prior work is an interesting topic for future research.

The term we use for our approach, “thermoeconomics,” aims to express our objective of providing an analysis of economic theory paralleling the mathematical structure of thermodynamics. We distinguish this usage from the variety of loosely connected prior usages of the term, which typically point to attempts that, like [GR], aim to *combine* economic and thermodynamic principles in contexts as diverse as understanding plant growth, the emergence of order in biological evolution, designing commercial energy systems, or clarifying the limits to economic growth.<sup>4</sup>

The objective of thermoeconomics, in the sense used here, is both narrower in scope and has a very different focus: aiming to map the mathematical structure of thermodynamics into economics, in order to provide a purely ‘macro-level’ analysis of economic phenomena. Thus, we aim to construct an economic theory using the mathematical and conceptual machinery of thermodynamics. Despite various past attempts, and the intuition among many researchers (but not, as we’ve seen, Samuelson) that such a link might exist, a purely thermodynamic approach to economics seems not yet to have reached fruition.<sup>5</sup>

A key ingredient for our thinking was provided by the axiomatic formulation of thermodynamics by Lieb and Yngvason in a series of papers culminating with [LY02], which clarifies the underlying mathematical structure of thermodynamics, independent of its physical interpretation. Thus, Lieb and Yngvason’s formulation doesn’t require any prior notions of energy, entropy, temperature etc, which then need to be mapped directly onto economic phenomena. Rather, their approach provides axioms governing abstract relations  $\lesssim$  and  $\equiv$  between states, which we shall introduce in Sections 5 and 6 respectively. Thus, any analogies between the two domains can emerge naturally from the mathematics; equally, where there are disanalogies, these will emerge automatically too. The present work takes literally their invitation in [LY98]: “Maybe an ingenious reader will find an application of this same logical structure to another field of science.”

Reading [LY02] is not a prerequisite for our paper, however. We require only their axioms, Theorem 1 on the existence of entropy and their subsequent results on its concavity and differentiability (all of which we review here). While our exposition here is self-contained, we adopt their numbering system for the axioms (though inserting our own slight variants), and largely follow their notation.<sup>6</sup> There are many nuances in their treatment that we skip over; we leave a more extensive analysis for future work.

The approach developed here blurs the boundaries between macro and microeconomics. Many of the topics treated below, such as the formation of market prices, are typically studied in microeconomics starting from fundamental assumptions about the rationality of individual economic agents. We treat these questions from an aggregate point of view, by making assumptions about markets as a whole, rather than their participants. This may create a more natural bridge to macroeconomic questions concerning, for example, the value of money and flows of trade. We see the ability to generate strong theoretical results without specific micro-foundations as a major advantage of this approach, regarding both generality and tractability. Nonetheless, the functional form of the resulting entropy function depends on the micro-foundations, and it is of course interesting to consider how a macro analysis can connect with microfoundations. We show how this can be done in the special case of an exchange economy consisting of agents with a Cobb-Douglas utility function.

The economics community is, of course, divided on the value of axiomatic approaches. On the one hand, the axiomatic approach of Arrow and Debreu [AD] and McKenzie [McK] to general equilibrium is regarded by many theoretical economists as a pinnacle of intellectual achievement. On the other hand, practical economists tend to be sceptical of axiomatic arguments and of mathematical theories in general, as encapsulated by Krugman’s comment on the 2008 financial crisis [Kr]: “The economics profession went astray because economists, as a group, mistook beauty, clad in impressive-looking mathematics, for truth.” We take a pragmatic view, justifying the axioms from economic reality as far as we can but identifying their limitations and proposing directions for extension. In

---

<sup>4</sup>For example, in applying economic ideas to understanding living systems [CK], [Co] states “The term “thermoeconomics” denotes a paradigm shift in our understanding of the role of energy in living systems, and in evolution. It is based on the proposition that energy in biological evolution can best be defined and understood, not in terms of the Second Law of Thermodynamics but in terms of such economic criteria as productivity, efficiency, and especially the costs and benefits (or “profitability”) of various mechanisms for capturing and utilizing energy to build biomass and do work.” Others, apparently beginning with Myron Tribus in 1962 [MS], have used “thermoeconomics” for the exploration of how fundamental thermodynamic constraints in the physical world shape and constrain economic activity, whether at the level of the individual industrial processes [TE] or the entire economy in the tradition of [GR]. According to [DG], for example, “Thermoeconomics combines thermodynamic principles with economic analysis and brings some fundamental changes in the economic evaluation, design, and maintenance of processes.” In sum, no specific usage of the term has yet become well-established.

<sup>5</sup>Our own efforts in building a link between thermodynamics and economics has had a long gestation. We started discussions on formulating a link in 2010, including supervising MSc student Daniel Sprague’s simulations of simple economic systems in 2011. We arrived at our present position only after many iterations and false starts.

<sup>6</sup>We make slight adjustments to their notation for ‘adiabatic accessibility’ and ‘adiabatic equivalence’—and we also drop the term ‘adiabatic’ entirely, as it is unfamiliar in economics and is ambiguous even in physics where it is usually defined as involving no heat flow but often used instead for slow variation.

the present context, that of building a link between thermodynamics and economics, an axiomatic approach is particularly valuable because it provides a rigorous way of clarifying a common underlying mathematical structure, rather than relying on intuitive analogies between ideas in each domain.

Finally, the economics community is, as we have already indicated, divided on the value of “imitating” physics. We aim to set aside such general concerns, and focus on showing how a thermodynamic perspective on economic systems can yield valuable insights for economic analysis. We hope that, especially if the present approach can be extended beyond the exchange economies treated here, concepts of economic temperature, economic entropy, money capacity and so on, may become part of standard economic thinking. Indeed, we suspect that the problematic relationship between economics and physics arises not because economists have been excessively entranced by attempting to imitate physics, but because the connections between the disciplines have not been pushed far enough.

### 3 Exchange economies

In this paper we restrict attention to exchange economies. In an exchange economy, agents exchange various sorts of durable good with each other; no goods are produced or consumed. This is of course a very limited view of economics, ignoring production, consumption, manufacturing, recycling, labour, interest rates, etc, but we intend to extend the approach in later work. In any case, it is fairly standard in economic theory to reduce much to exchange, for example, one could exchange options to consume a good, provide labour, money, or supply a service, on some future date or conditional on a specific state of the world.

We suppose there is a particular durable good that we will use as a reference (a “numéraire” in economics) and will call *money*. We assume that at the aggregate level, more money is always “preferred” to less, no matter how much money and other goods are present already. A more precise meaning will become clear in our treatment below—but roughly this means that if a trader (a figure we will introduce in Section 5) were to offer freely to give or take money to or from members of an economy, they would in aggregate accept more money than they would give (indeed, it is natural to expect that this would be true for each individual, though this is not needed in our account).<sup>7</sup> Money should also be easy to store and exchange, though in our exchange economies all goods are durable and exchangeable; in particular, we will want it to be easily transferable to another economy. To avoid giving a national preference, we measure its amount in “aurum”, denoted by  $Au$  or  $\odot$ . The unit is named after gold, in the light of gold’s historical role as a numéraire for a large part of the world. Money may or may not have intrinsic value (contrast gold with paper); we explore the consequences of “pure money” in Section 15. We leave extending the approach to settings in which there is more than one currency (touched on briefly in Section 21.1) for future work.

As noted above, we take a deliberately macroscopic view. Hence we do not have to specify the process by which individual people choose which bundles of goods to exchange nor how frequently they do so. Nonetheless, we will study one microeconomic example (Section 10) to illustrate how the thermodynamic quantities of our macroeconomic theory emerge from a microeconomic model. The literature contains many other examples of microeconomic model of exchange, for which some references can be found in a recent paper [DGMS]. An important next step (on which we have already made a start and plan to report later) is testing the theory using computer simulations, exploring both existing models of micro-interactions and new variants of such models.

### 4 Equilibrium

Our formulation depends on an assumption we call A0, that any closed exchange economy with given amounts of goods and money in a class of systems settles down to a unique statistical equilibrium state.<sup>8</sup> This makes it a “simple system” in the language of [LY02]: the state is specified by a point in a subset of  $\mathbb{R}^n$ , for some dimension  $n$ , the coordinates being the amounts of each type of good and of money. One could also consider the number of agents as a coordinate for the state, but for present purposes we will take the number of agents in a system as fixed. An implicit assumption is that a simple system is “financially indecomposable”: however the specified amount of money is initially distributed it ends up distributed with the same probability distribution. So there is only one money coordinate. But we do not require the indecomposability assumption for types of good other than money. For example, we allow the possibility there might be no flow of a good from one part an economy to another, so that the amounts of that type of good in each part of the economy are independent coordinates for the state.

We emphasise that statistical equilibrium is not an equilibrium in the sense of Nash, where no player can benefit from a unilateral change [N50]. Nor is it an equilibrium in the sense of Arrow and Debreu, where prices are such that supply balances demand and then all exchanges stop [A51, D51]. Instead, it is a dynamic state of

<sup>7</sup>The corresponding aggregate property need not hold for other types of goods: there may be regimes where it is preferable not to have more, or even preferable to have less (e.g., if the good acts as a pollutant or is simply “surplus to requirements”).

<sup>8</sup>In their analysis of thermodynamic entropy, Lieb and Yngvason didn’t label the corresponding property as an axiom. Nonetheless, we feel it merits a label (A0).

continuing exchanges but with a steady probability distribution for the micro-state. For example, it is entirely possible that people’s preferences continually vary (e.g., they become bored with objects they currently have and would rather exchange them for different ones). Perhaps the closest analogue in economics is dynamic stochastic general equilibrium (DSGE) [Ga], the workhorse of treasury modelling. The idea of equilibrium in a strict sense has been criticised from early days in economics. For example, Irving Fisher wrote in 1933 [Fi] that “It is as absurd to assume that, for any long period of time, the variables in the economic organization, or any part of them, will ‘stay put,’ in perfect equilibrium, as to assume that the Atlantic Ocean can ever be without a wave.” A statistical equilibrium with steady (or slowly varying) macro-parameters is more plausible.<sup>9</sup>

The notion of statistical equilibrium is itself, of course, a simplification of real economic activity, in which there are, among other things, continual technological and business innovations. These considerations do not arise in our idealized exchange economies. Nonetheless, we believe that the present approach could usefully be extended to deal with such cases, expanding on the intuition that the only innovations that catch on are those that increase economic entropy. For now, we note that entropy is useful in physics even in a time-varying environment and helps explain the formation of complex structure in physics, chemistry and biology. In particular, it is worth making a comparison with biochemistry. Concepts from equilibrium thermodynamics like changes in Gibbs free energy are very useful in understanding which processes occur, even if a cell is a highly structured environment and subject to time-varying conditions. The assumption behind this analysis is local thermodynamic equilibrium.

A different concern with the equilibrium assumption is that an economic system might have multiple equilibria [Fa]. This might well be the case when there are fluxes of inputs and outputs; but without inputs and outputs, it seems to us a plausible working assumption that an economic system with specified amounts of goods and money and no obvious barriers goes to a unique statistical equilibrium. This type of indecomposability assumption is widely made in analysing physical systems, in particular in statistical mechanics. It is analogous to the property of ergodicity in stochastic dynamics, meaning that there is a unique stationary distribution and it attracts all initial distributions [Li]. Similarly, it is analogous to the property of mixing for measure-preserving dynamics  $\phi_t$ , meaning that for all measurable subsets  $A$  and  $B$ , the probability of  $A \cap \phi_t B$  converges to the product of the probabilities of  $A$  and  $B$  [Wa].

To gain an intuition for the significance of the assumption that our exchange economy has a unique statistical equilibrium, it is useful to consider how this assumption can be violated. Suppose that the economic system under study is decomposable into two or more non-interacting parts (e.g., corresponding to islands which are entirely cut off from each other). Then there will be a continuum of ways in which goods and money can be divided between the parts (and this division will be conserved as the system evolves, because there is by assumption no trade between the parts). Thus each of the continuum of possible divisions will correspond to a distinct region of the state space with its own statistical equilibrium. We assume, unless specified as here, that this does not occur.<sup>10</sup> An economic system with more than one money coordinate, corresponding to conserved amounts of money in financially disconnected parts of the systems, is called “compound”.

## 5 Accessibility

The next key notion is the *accessibility* of one state of an economy from another when subject to external influence (for example, through interaction with other economies, or wealthy traders). Indeed, the main aim of the present approach is to understand which transitions are, or are not, possible.

To build an intuition for why the question of accessibility is crucial, consider the following. One might suspect that a sufficiently rich trader could rearrange goods between agents in the economy in any way desired, by paying people to buy and sell to each other to order; but to do this, the people in the economy would thereby have to be given additional money by the trader to incentivize them to engage in the desired transactions. But money is itself part of the state of the economy. So the trader does not have complete control after all—some state-transitions are possible and others are not. Indeed, suppose the trader wanted rapidly (rather than arbitrarily slowly) to reverse the allocations of goods to its original state. Doing so would require further financial inducements; so that while goods (apart from money) could be returned to the initial state, the economy would now contain more money than before. The reader may have the intuition that the trader forcibly “moving” goods about the economy seems analogous to doing work on a physical system (e.g., compressing a gas by a piston). Just as moving a physical

---

<sup>9</sup>A possible criticism of our equilibrium assumption, especially from finance practioners, is that it cannot account for endogenous business cycles. We note that there is considerable debate in the academic literature concerning the nature and even existence of such cycles [SI37, FG94, Man89]. But we also note that the thermoeconomic framework can potentially generate such cycles through the system being held out of equilibrium by fluxes of inputs and outputs, just as a physical system may exhibit oscillations when held out of thermodynamic equilibrium (e.g., in Rayleigh-Bénard convection or the Belousov-Zhabotinskii reaction [NN]). This is an important direction for future work. Here, however, we allow only temporary fluxes of inputs and outputs (e.g. with an external trader).

<sup>10</sup>As already indicated, there is a parallel in classical statistical mechanics: for example, one assumes that there is no extra conserved quantity, beyond those already identified as labels for the state, that would prevent wide exploration of the space corresponding to the state. More generally, one assumes that there is no invariant surface separating the space into two parts, or even more generally, one assumes ergodicity of the conditional of Liouville measure on the specified values of conserved quantities. For a simple introductory reference, see [Set].

system from one state to another and back again will “heat” the system being operated upon (except when the movement is arbitrarily slow), so forcibly moving an economic system from one state to another will increase the economic temperature of the system (by increasing the amount of money in the economy, and thus reducing the value of money). We will make these intuitions rigorous below.

We consider a system of one or more exchange economies connected to an external trader who has unlimited goods and money, and ask what effects on the system the trader can achieve.

Let us say that a state  $Y$  of an economic system is *accessible* from state  $X$ , written  $X \lesssim Y$ , if the trader can move the system from state  $X$  to state  $Y$  with no nett change in the trader’s assets apart from money. Note that our notation is slightly different from that of [LY02], who use  $\prec$  instead.

For a simple exchange economy, specified just by total amounts of each type of good and money, because of the requirement that the trader end up with the same quantity of goods and our assumption that money is always desirable, we assume that the only accessible states are those with the same quantity of goods and at least as much money. But for a compound system, as the above discussion indicates, the set of accessible states is less obvious. A key contribution of the analysis below is to clarify this case, as we see shortly.

To continue introducing the framework, note that directly from the definition, we can see that  $\lesssim$  is a pre-order (it is reflexive:  $X \lesssim X$ , because the trader does not have to do anything to achieve no change; and it is transitive:  $X \lesssim Y$  &  $Y \lesssim Z$  implies  $X \lesssim Z$ , because the trader just has to make one change followed by the other). Transitivity of  $\lesssim$  is A2 of [LY02] but we get it for free here.

We say state  $Y$  is *reversibly accessible* from  $X$ , and write  $X \sim Y$ , if  $X \lesssim Y$  &  $Y \lesssim X$ . It is an equivalence relation: symmetric:  $X \sim Y$  iff  $Y \sim X$ , reflexive (which is A1 of [LY02] but again we get it for free) and transitive.

For later use, we will write  $X \prec Y$  if  $X \lesssim Y$  &  $X \not\sim Y$ .

Denote the state of a (compound) system consisting of two unconnected economies (or systems of economies)  $A$  and  $B$  in states  $X_A, Y_B$  respectively, by  $(X_A, Y_B)$ . In particular, if  $X_A \lesssim X'_A$  and  $Y_B \lesssim Y'_B$  then  $(X_A, Y_B) \lesssim (X'_A, Y'_B)$ , because the trader can make the changes to  $A$  and  $B$  in parallel. This is A3 of [LY02] but we obtain it for free again.

## 6 Financial equilibrium

A further key notion is financial equilibrium. This will be crucial to making sense of the notion of the “temperature” of an economy, which will represent the reciprocal of the marginal utility of money.

Given two simple economies  $A$  and  $B$ , we define their *financial join* to be the joint economy where money is allowed to flow freely between them but no other goods can (we say they are in financial contact). There are various reasons why this might occur and the present approach is neutral between them. For example, people working in one economy might wish to send money to relatives in the other. Even more simply, one person could own assets in both economies; when financial contact is allowed, they can choose to send money from their holdings in one economy to their holdings in the other. It will make sense to send money to the economy in which its value is higher, i.e., to the one with the lower temperature (to be formalised in Section 12).<sup>11</sup>

Denote by  $\theta(X_A, Y_B)$  the state of the financial join of  $A, B$  that is reached from initial states  $X_A, Y_B$  after  $A$  and  $B$  have come to equilibrium. The equilibrium is specified by the total amount of money and the individual amounts of the other goods in each economy. Because  $A$  and  $B$  are assumed to be indecomposable, it is reasonable to assume that the equilibrium state of the financial join is to a high approximation independent of the way the join is made.<sup>12</sup> The idea is that if say  $A$  were decomposable into  $A_1 + A_2$  then the effects of joining  $B$  to  $A_1$  or  $A_2$  or both would certainly be different, depending on the distribution of goods and money between  $A_1$  and  $A_2$ , whereas if  $A$  and  $B$  are indecomposable then any way of letting money flow between them should produce approximately the same statistical state.

The process of coming to equilibrium will in general lead to a nett transfer of some money from one economy to the other, but in equilibrium there is no nett flow of money between them (though there will be on-going, but zero-mean, exchanges of money between the economies). Note that we are hereby applying A0 to the financial join and thus assuming that allowing money to flow between the two economies is enough to achieve ergodicity of the joint system. We deduce that  $(X_A, Y_B) \lesssim \theta(X_A, Y_B)$ , because the trader does not have to do anything. This is A11 of [LY02] but we get it for free.

We also deduce that there exist states  $X'_A, Y'_B$  such that  $\theta(X_A, Y_B) \sim (X'_A, Y'_B)$ . To see why, let  $X'_A, Y'_B$  be the states of  $A$  and  $B$  after financial contact. Then the only difference between the left- and right-hand sides is that the left-side has financial contact between  $A$  and  $B$ ; but the trader can act as a financial contact between  $A$  and  $B$  on the right-hand side by simply detecting if someone in  $A$  wants to give money to someone in  $B$  and making the transfer for them. So in the presence of the trader, the two sides are the same. Thus they are reversibly accessible from each other. This is A12 of [LY02] and we get it again for free.

<sup>11</sup>A challenge for future work is to analyse some of the forces that drive real-world flows, such as interest rates and investment opportunities. We also leave treating surrogates for money, such as cheques or I-owe-yous (IOUs), and how to deal with credit, for later work.

<sup>12</sup>As far as we can see, Lieb and Yngvason did not address the corresponding question about thermal joins.

If there is no nett flow of money between  $A$  and  $B$  on bringing them into financial contact, we say their states are in *financial equilibrium* and write  $X_A \equiv Y_B$ . Thus, after nett money flow has occurred,  $X'_A \equiv Y'_B$ . From the definition,  $\equiv$  is symmetric.<sup>13</sup>

We assume also (A13') that  $\equiv$  is reflexive: i.e., that each state of an economy is in financial equilibrium with itself, or put differently, there are no internal barriers to money flow. Formally, this requires us to be able to make a copy of the system, to make the financial join of a system to itself, but let us suppose that this is possible. Then there is no money flow between these copies because of the assumed indecomposability of (each copy of) the system. Thus  $\equiv$  is reflexive.<sup>14</sup>

Finally, we assume that  $\equiv$  is transitive. This is A13 of [LY02].<sup>15</sup>

Thus,  $\equiv$  is symmetric, reflexive and transitive, and hence is an equivalence relation.

To give some intuition for the significance of  $\equiv$ , we will see that economies which are in financial equilibrium, i.e., such that there is no nett tendency for money to flow from one to the other, are at the same “temperature,” although the notion of temperature is only derived later in the exposition. We will see that this corresponds intuitively to money having the same “value” in each economy, so that there is no tendency to generate a nett flow of money from one economy to another.

## 7 Accessibility and financial equilibrium

A key component of Lieb and Yngvason’s approach to thermodynamics is to formulate connections between the two binary relations  $\preceq$  and  $\equiv$ . In particular, we want A14: for each state  $X$  of a simple system  $A$  there are states  $X_0, X_1$  with  $X_0 \equiv X_1$  such that  $X_0 \prec X \prec X_1$ . Intuitively the reason that this assumption is important is that it ensures that the equivalence classes of reversible accessibility and financial equilibrium “cut across” each other and avoid “degeneracy”.<sup>16</sup> This assumption is crucial below in deriving what we will call the second law of thermoeconomics. Here we give a plausibility argument for this assumption in the economic context.

First let  $X'_0, X'_1$  be  $X \mp 1\odot$ . Then  $X'_0 \prec X \prec X'_1$  because, as in Section 5, we have assumed that an economy accepts money from the trader but will not give money back to the trader for nothing. Thus, the trader can just give  $1\odot$  to move  $X'_0$  to  $X$  and  $1\odot$  to move  $X$  to  $X'_1$ , and each change is irreversible (so we get  $\prec$ ).<sup>17</sup>

Intuitively, under financial contact, money will tend to flow from an economy where it has less “value”, to an economy where it has greater value. Thus, consider two economies with the same goods, but different quantities of money. Within an economy, the value of any unit of money will be lower where the amount of money is greater (just as money is devalued when a central bank prints money). This means that the value of money will be lower in  $X'_1$  than in  $X'_0$ , and hence money will tend to flow from  $X'_1$  to  $X'_0$ .

So, clone the system into  $A_0, A_1$ , with initial states  $X'_0, X'_1$ . We wish to find a reversible path from  $X'_1$  to a new state and a reversible path from  $X'_0$  to a new state, such that the resulting states are in financial equilibrium. This can be done by the trader reversibly buying goods from  $X'_0$ . This increases the amount of money in comparison with the volumes of goods in that economy (and hence reduces the value of money there, to be formalised in Section 12). At the same time, the trader can reversibly sell those same goods to  $X'_0$  (increasing the value of money in that economy). The trader ends up in the same state except for amount of money. With sufficient trading, the two copies can be brought to financial equilibrium. Let  $X_0 \equiv X_1$  be the states achieved, then  $X_0 \sim X'_0 \prec X \prec X'_1 \sim X_1$ .

Note that, as pointed out by [LY02], A14 implies that for each state  $X$  there is a state  $Y$  such that  $X \prec Y$ , which is their A8. In economic terms, this just specifies that there is no “best” state—indeed, any state can be improved by the trader adding money.

## 8 Scaling symmetry: Extensivity

Next comes the crucial assumption A4 of scaling symmetry: that one can scale an economy  $A$  by any positive real number  $\lambda$  to obtain a scaled economy  $\lambda A$  and that if  $X \preceq Y$  for  $A$  then  $\lambda X \preceq \lambda Y$  for  $\lambda A$ . This assumption seems plausible if we think of an economy as consisting of lots of small units with a network of purely local interactions based on geographic proximity and we ignore the discreteness of the set of agents and possibly of some types of

<sup>13</sup>[LY02] deduce symmetry of  $\equiv$  in their Lemma 2(ii), from their axioms A4, A5, A7, A11 and A12 but we have not yet introduced A4, A5, A7, and prefer to reduce our dependence on them.

<sup>14</sup>A13' is not one of the axioms of [LY02]; instead they deduce it in their Lemma 2(i) from A4, A5, A7, A11 and A12. We prefer to list it as an axiom, to avoid relying on A4 and A5, and also reflexivity doesn't seem to us as clear a logical consequence as they claim: it might depend on the way in which the join is made. So we treat reflexivity as an axiom, A13'.

<sup>15</sup>One consequence to note is that this implies that for the equilibrium of a network of economies in financial contact, there are no nett cyclic flows of money. This is because the equilibrium for a spanning tree has no cycles so no nett money flows, and by transitivity when one adds the remaining links the economies are already in equilibrium, so there is no change to the nett flows.

<sup>16</sup>A thermometer is such a degenerate case—a system which varies along a single dimension, so that for all  $X, Y$  at different temperatures  $X \not\sim Y$ . For discussion of how to deal with this case, see [LY99]. We will introduce financial thermometers in Section 16.

<sup>17</sup>We do allow that money can flow from one economy to another, as in Section 6, but we preclude the possibility that an economy will voluntarily hand over money *to the trader*.

good. Then one can imagine a scaled version of the economy and it is plausible that the accessibility relation would inherit scaling symmetry. We call an economy satisfying A4 *extensive*.

Yet it may be a poor assumption for real economies. The concept of economies of scale is well-known and already argues against scaling symmetry. Indeed, an economy might need to be at least a certain size before it could include certain types of industry at all (say aircraft manufacture or investment banking).

In sum, the assumption of scalability is strong, and unlikely to hold in general in many aspects of real economies. Nonetheless, the analogy with thermodynamics in the natural sciences suggests that the assumption may be useful for many purposes. For example, classical entropy, which is based on the same scaling assumptions, has been of central importance in analyzing chemical reactions inside biological cells, although cells, like economies, are clearly enormously intricate mechanisms containing specific structures and processes operating at specific scales. In any case, we proceed with the scalability assumption, to explore whether it may provide a potentially useful idealization in some circumstances. Later we will briefly consider whether and how it may be relaxed.

A further assumption concerning scaling is A5: any system can be subdivided into two parts in an arbitrary ratio  $\lambda : 1-\lambda$  by cutting connections, and in particular the state  $X$  can be decomposed into a state  $(\lambda X, (1-\lambda)X)$  of the pair of systems, with  $X \sim (\lambda X, (1-\lambda)X)$ . Again, it is not clear that this holds in the economic context: cutting internal connections might make a substantial change to an economy, and the economy might not be homogeneous so the two parts might not be just scaled versions of the whole. We proceed nonetheless (and leave implications of weaker assumptions for later work).

Then by Lemma 4 of [LY], we obtain the “comparison hypothesis”: for all states  $Y = (\lambda_1 Y_1, \dots, \lambda_n Y_n)$  and  $Y' = (\lambda'_1 Y'_1, \dots, \lambda'_n Y'_n)$  of a system (with not necessarily the same subdivisions, but where necessarily  $\sum_i \lambda_i = \sum_j \lambda'_j$ ), then either  $Y \lesssim Y'$  or  $Y' \lesssim Y$  (including the possibility that  $Y \sim Y'$ ). That is, all states can be compared with each other by the accessibility relation. Note that this is a very strong claim: in economic terms, it means that for any two such states,  $Y$  and  $Y'$  containing the same stock of goods, the trader can either move the system from  $Y$  to  $Y'$ , or from  $Y'$  to  $Y$ , or both. The trader can always guide exchanges of goods either in one direction, or the other, or both, and give or take money; but cannot create or destroy goods. The economy in question may itself be a compound of any number of smaller component economies, so that the trades required may be arbitrarily complex.

Finally, we also assume Lieb and Yngvason’s A6: if  $(X, \varepsilon Z_0) \lesssim (Y, \varepsilon Z_1)$  for some  $Z_0, Z_1$  and a sequence of  $\varepsilon \rightarrow 0$  then  $X \lesssim Y$ . In economic terms, this means that if two economies are reversibly accessible when each is conjoined with an arbitrarily tiny scaled copy of some other economy, then they are also reversibly accessible from each other if that arbitrarily tiny scaled copy is entirely absent. This seems reasonable, if we set aside the discreteness of the set of agents and possibly of some types of good.

## 9 Economic Entropy

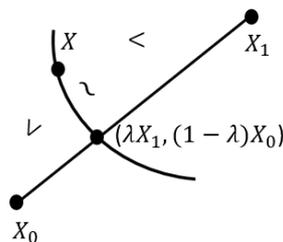
With the above assumptions we are now ready to begin to reap the rewards. The first reward is

**Theorem** (Theorem 1 of [LY02] applied to this context): There is a function  $S$  of state of all extensive economies such that for any two subdivisions  $Y, Y'$  of a system (using the notation of the previous section) then

$$Y \lesssim Y' \text{ iff } \sum_i \lambda_i S(Y_i) \leq \sum_j \lambda'_j S(Y'_j); \quad (1)$$

and the function  $S$  is unique up to orientation-preserving affine transformations:  $S \mapsto aS + b$  with  $a > 0$ .

We illustrate the proof from [LY02] in Figure 1. In keeping with thermodynamics, we call  $S$  *entropy*. The



**Figure 1:** Sketch proof of existence of entropy: given  $X_0, X_1$  as in A14, we draw the straight line of convex combinations and deduce that any  $X$  between them in the accessibility order is equivalent to one point on the line; the resulting  $\lambda$  is the desired entropy; changing the choice of  $X_0, X_1$  makes an affine transformation of  $S$ .

scale for entropy can be fixed by choosing a scale for temperature (to be defined in Section 12), leaving only its origin as arbitrary.

We deduce the all-important

**Second law of thermoeconomics:** On putting two or more economies into contact (in any way), the total entropy can not decrease.

To prove the Second Law, take systems  $A$  and  $B$  in states  $Y_A, Y_B$ , respectively and put them in contact. It could be just financial contact or permitting arbitrary flow of goods and money or permitting flows corresponding to a given price. The joint system goes to a state equivalent to some  $(Y'_A, Y'_B)$  (dependent on the type of contact). The trader does not do anything, so we have  $(Y_A, Y_B) \preceq (Y'_A, Y'_B)$ . By the property (1) we deduce that  $S(Y_A) + S(Y_B) \leq S(Y'_A) + S(Y'_B)$ .

Note that [LY02] consider the second law to be just the existence of entropy with the property (1), but we consider it worth stating this particular consequence. Note also that the terminology “second law” is historical: the first law of thermodynamics is conservation of energy. It took physicists as much thought and experiment to settle on the first law as on the second one. In our exchange economies, each type of good, including money, is assumed to be conserved, so that the “first law” of the conservation of money (and indeed, of goods) is trivial. In principle, there could be further conserved quantities needed to describe a macro-state, but at least in our context of exchange economies we do not see one.

For a simple system, with its state specified by total amounts of each type of good (including money), the accessibility relation is easy: a state  $Y$  is accessible from state  $X$  if and only if  $Y$  differs from  $X$  at most by having more money. But the consequences of entropy become significant for systems consisting of several parts with specifications of what transfers are possible.

For example, for a system with two parts, having a common money and a common type of good, with transfers only via a trader who has to end up with the same state except for money, the state space for the system is given by the amounts  $G_j, M_j$  of goods and money in parts  $j = 1, 2$ , but the allowed changes conserve  $G_1 + G_2$ . Within this 3D space, the 2D surface of constant total entropy separates the space into those states that are accessible from a point on this surface from those that are not accessible. Changes within the surface of constant total entropy are reversible. The others are not.

Alternatively, one could leave out the trader, but allow transfers of goods and money between the two parts: that is, the economies trade directly with each other, rather than each transaction being mediated by the trader. Then the possible changes conserve both  $G_1 + G_2$  and  $M_1 + M_2$ . In this 2D space there are curves of constant total entropy which separate accessible states from inaccessible ones. The picture is analogous to the indifference curves for individuals in standard economics, but now at the level of a whole multi-part economic system. These cases will be illustrated by simple examples in the next Section.

The second law suggests that we can think of entropy as an aggregate utility or value of an economy. But in contrast to standard utility theory in which the utilities of different agents are typically not viewed as being on a common scale [ER], one can meaningfully compare, and indeed add, the entropies of different parts of an economic system. Indeed, it is only the total entropy that can not decrease. Thus, in some interactions between two economies, the entropy of one economy can decrease, as long as this is compensated by an equally large or larger increase of the entropy of the other economy. For example, this happens when money flows out of an economy where it has less value to an economy where money has greater value.

It is worth reflecting on how extraordinarily strong is the existence of entropy, with the property (1). On the one hand, it provides a very restrictive constraint on possible transitions from one equilibrium state of a system or set of systems to another: it precludes any transition that leads to a decrease in (total) entropy, whatever the trader (or other exogenous agent or agents) does. On the other hand, it specifies that there is always an available path by which the trader or other external agency can move the economy in *any* entropy-increasing way in the state space. Entropy-conserving changes can also be achieved in a limiting sense (in practise, requiring transitions to be carried out arbitrarily slowly).

If we interpret entropy as corresponding to aggregate utility, then the first observation implies that trade, or financial flow, between any two (exchange) economies does not decrease aggregate utility. That is, there will always be gains of trade, except for the limiting case of zero gains. Indeed, if any number of economies are put into contact using any network of trade or financial connections, trade or financial flows will yield non-negative gains of trade for the compound system. Furthermore, any reduction in restrictions on trade (e.g., removing a barrier to trade for one or more goods, or removing a prohibition on moving money from one economy to another) will lead to non-negative gains of trade. Note, though, that the gains of trade may be unequally distributed—indeed, one or more economy may lose entropy, as long as the entropy of the whole system increases. We analyse trade in detail in Section 19.

One consequence of existence and extensiveness of entropy is that it provides a rationale for the standard economics notion of representative agent,<sup>18</sup> but with a twist. Given entropy function  $S(M, G)$  for an economy

<sup>18</sup>The notion of a representative agent, is widely and somewhat controversially used in providing a microfoundation for macroeconomic behaviour. [H96] traces the notion back to Alfred Marshall [M20], while [K92] finds similar ideas in Francis Edgeworth [E81]. The notion of a representative firm was critiqued by Lionel Robbins [R28]; and critiques of the broader coherence of the notion, in view

with  $N$  agents, we can define  $U_r(m, g) = S(Nm, Ng)/N$  and consider it as the utility for a representative agent to possess  $m$  money and  $g$  goods. Maximising the sum of the entropies of several economies then corresponds exactly to maximising the average utility of their representative agents, weighted by the numbers of agents in each economy. But there is no need to scale entropy to the level of individual agents. Our approach says that each economy has a utility function (its entropy) and a collection of economies in contact act to maximise the sum of their constituent utilities (i.e., maximizing the entropy of the collection of economies as a whole). Note that this maximization of summed utility over the system of economies can arise purely from the interactions of myopic or selfish agents, from their ability to trade and to move money from one economy to another, as long as there is convergence to a unique statistical equilibrium (Section 4). The “twist” is that there is no requirement that each economy attempts to “maximize its own utility.” For example, money can flow from one economy to another, lowering the entropy of one and increasing the entropy of the other, as long as the change in entropy of the whole system does not decrease. We will, though, also consider in Section 19.1 the important special case of what we will call “mutually beneficial” trade, which is constrained not to decrease the entropy of each economy (i.e., leading to an analogue of Pareto optimality at the level of economies).

To end this section, we discuss to what extent the entropy function depends on the choice of numeraire (money in our presentation so far).<sup>19</sup> We restrict attention to numeraires for which it is always desirable to have more (required for Sections 5 and 7). A priori, the accessibility concept depends on the choice, because the state space for a system has fixed amounts of all goods except the numeraire. But one can consider the restricted set of changes that fix the amount of numeraire too. Then accessibility of one state from another is independent of which good is considered to be the numeraire. Hence the partial orders determined by the two entropy functions must be the same. Then extensivity of the entropies implies that one must be an affine transformation of the other.

## 10 A toy illustration with interacting agents

One of the distinctive characteristics of the conception of thermoeconomics developed here is that it proceeds entirely at the level of aggregate behaviour, positing axioms about trade and financial flows and deducing conclusions such as the second law. Thus, just as classical thermodynamics makes no reference to the molecular basis of pressure, heat, entropy and so on, so thermoeconomics makes no reference to the micro-foundation of economic behaviour, in terms of the actions of individual agents. A great advantage of this strategy, both for physics and economics, is that the micro-foundations of real-world systems are often of great complexity and only partially known, so that providing a detailed micro-level analysis is not realistic—but nonetheless, laws at the aggregate level are still derived.

It is interesting, however, to illustrate the general theory by special cases in which micro-foundations are sufficiently simple that they can be analysed directly. In the physics context, a prime example is the molecular model of ideal gasses, leading to a foundational explanation of temperature and pressure and the entropy function  $S = N \log vu^{d/2}$  (though it is conventional to add  $N(d+2)/2$  and also to multiply by Boltzmann’s constant), where  $N$  is the number of molecules,  $v$  and  $u$  are the volume and energy per molecule and  $d$  is the number of degrees of freedom per molecule (e.g. for a diatomic molecule like  $O_2$  or  $N_2$ ,  $d$  is the sum of 3 translational plus two rotational degrees of freedom, at normal temperatures where vibration is suppressed quantum-mechanically).

In particular, let us compute the entropy for a toy example of an exchange economy with a single good in addition to money and  $N$  independent identical “Cobb-Douglas” agents. By Cobb-Douglas agents we mean firstly that they have a utility function

$$U(m, g) = g^{\alpha-1} m^{\eta-1}, \quad (2)$$

for amounts  $g$  of good and  $m$  of money, for some  $\alpha, \eta > 0$ . It is natural to think of both goods and money as being desirable, in which case  $\alpha, \eta > 1$ .<sup>20</sup> Secondly, they interact via stochastic dynamics that we’ll define precisely in the next paragraph. Importantly, in contrast to standard economic usage, our agents are not utility maximisers; instead, they use their utility functions to bias the probability rate for the outcome of random pairwise encounters towards increased utility for both, but with some “errors”.

Let us consider an encounter between one agent  $i$  and another agent  $j$ . Label the amounts of money and goods for each agent before the encounter by pairs  $m = (m_i, m_j), g = (g_i, g_j)$  of positive 2-component vectors, and after the encounter by  $m', g'$ . We suppose that transitions are restricted to those that conserve both goods and money:  $m'_i + m'_j = m_i + m_j$  and  $g'_i + g'_j = g_i + g_j$ . We denote the probability rate from  $m, g$  to  $m', g'$  by  $q(m, g, m', g')$ . Each agent biases the probability rate by making it proportional to their utility of the outcome.

---

of the heterogeneity of economic agents, have continued [K92]. Nonetheless, representative agent models continue to be influential in macroeconomics and finance [MP].

<sup>19</sup>We are not aware whether Lieb and Yngvason discussed the analogous question in the physical case, but a difference is that they allow volume to change, whereas we fix the amount of goods.

<sup>20</sup>We adopt the form (2) with minus ones in the exponents because it makes future formulae simpler and also allows the case of a good that is undesirable, such as a pollutant.

Thus we take

$$q(m, g, m', g') = k_{ij} (g'_i g'_j)^{\alpha-1} (m'_i m'_j)^{\eta-1} \delta(g'_i + g'_j - g_i - g_j) \delta(m'_i + m'_j - m_i - m_j),$$

where  $k_{ij} \geq 0$  is an encounter rate for  $i$  and  $j$  and the  $\delta$ -functions enforce conservation of goods and money. We suppose that the outcome is independent of all previous encounters of all pairs, conditional on the current state. The constraints  $\alpha, \eta > 0$  make the rate integrable and hence the stochastic process well-defined.

Assuming the graph formed by positive encounter rates between agents is connected, for given total goods  $G$  and total money  $M$  this stochastic process has a unique stationary probability that attracts all initial probability distributions (on a finite state space, this would follow from a standard result for Markov processes [Sen], but here the state space is continuous; nevertheless, the result can be obtained for continuous spaces by use of Dobrushin's dependency matrix, e.g. [M11]; or see [DGS] for treatment of the case with only money and  $\eta = 1$ ). In fact, the process is reversible (in the sense of Markov chains) with respect to the distribution with density  $\rho$  proportional to the product of the utilities, i.e.,

$$\rho(g, m) = Z^{-1} \prod_i g_i^{\alpha-1} m_i^{\eta-1} \delta\left(\sum_i g_i - G\right) \delta\left(\sum_i m_i - M\right),$$

where  $g, m$  now represent the vectors  $(g_1, \dots, g_N)$  and  $(m_1, \dots, m_N)$ . Hence we deduce that this distribution is the stationary one. It is a product of two Dirichlet distributions (one for the partition of goods, the other for money), and the normalisation constant

$$Z = \frac{G^{N\alpha-1} \Gamma(\alpha)^N M^{N\eta-1} \Gamma(\eta)^N}{\Gamma(N\alpha) \Gamma(N\eta)}.$$

The marginal distribution for one agent in this stationary distribution is a product of two independent Beta-distributions (one for goods, one for money). For example, the Beta-distribution for money has density

$$\rho(m) = \frac{\Gamma(N\eta)}{\Gamma(\eta)\Gamma((N-1)\eta)} \frac{m^{\eta-1} (M-m)^{(N-1)\eta-1}}{M^{N\eta-1}}.$$

It is an illustration of how inequality of wealth can arise, though note that that for  $\eta > 1$  and  $N$  large, the bulk of the distribution is peaked close to the mean  $M/N$ , and that each agent samples the whole distribution as time evolves.<sup>21</sup>

To determine the entropy of this toy economy we need to spell out how each agent interacts with an external trader. We specify this to be consistent with how agents interact with the other agents, i.e., biased in the same way by the utility function. We suppose the trader offers to buy or sell arbitrary amounts at a price  $\mu$ . Then the probability rate for each agent  $i$  to transition from  $m, g$  to  $m', g'$  is

$$q(m, g, m', g') = K_i g'^{\alpha-1} m'^{\eta-1} \delta(m' + \mu g' - (m + \mu g)),$$

for some  $K_i \geq 0$ , which is the trader's encounter rate with agent  $i$ . There is no contribution to the dependence on  $m', g'$  from the trader because it has unlimited goods and money. As the transition is along the "budget" line  $m' + \mu g' = m + \mu g$  we can eliminate  $m'$  in favour of  $g'$  and write the rate as a function of just  $g'$  (given  $m, g, \mu$ ):

$$q(g') = K_i g'^{\alpha-1} (m + \mu g - \mu g')^{\eta-1}.$$

This is a Beta-distribution again (though unnormalised), and its mean is

$$\langle g' \rangle = \frac{\alpha}{\alpha + \eta} (g + m/\mu).$$

The means over the initial stationary distribution were  $\langle g \rangle = \frac{G}{N}$ ,  $\langle m \rangle = \frac{M}{N}$ . Thus the change from the mean of  $g$  to the mean of  $g'$  is

$$\frac{\alpha M/\mu - \eta G}{N(\alpha + \eta)}.$$

Thus in a time  $\Delta t$  we obtain a change

$$\Delta G = K \left( \frac{\alpha M}{\mu} - \eta G \right)$$

in the quantity of goods, where  $K = \frac{\Delta t}{N(\alpha + \eta)} \sum_i K_i$ .  $\Delta G$  is positive, zero or negative according as  $\mu$  is less than, equal to, or greater than

$$\mu_c = \frac{\alpha M}{\eta G}. \quad (3)$$

The change in the amount of money is  $\Delta M = -\mu \Delta G$ . So we see that

$$\alpha \frac{\Delta G}{G} + \eta \frac{\Delta M}{M} = \left( \frac{\alpha}{G} - \mu \frac{\eta}{M} \right) \Delta G = \frac{K}{GM\mu} (\alpha M - \mu \eta G)^2 \geq 0,$$

<sup>21</sup>In particular, wealth does not concentrate in a fixed set of agents.

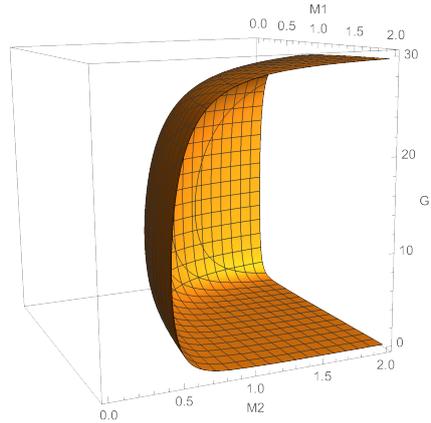
with equality iff  $\mu = \mu_c$ . This is precisely the condition  $\Delta S \geq 0$  for infinitesimal changes in

$$S = N \log \left( \left( \frac{G}{N} \right)^\alpha \left( \frac{M}{N} \right)^\eta \right). \quad (4)$$

Thus we deduce that (up to an affine transformation) the entropy of the toy economy is this function  $S$ . The factors of  $N$  are included to make the entropy scale correctly with  $N$ .

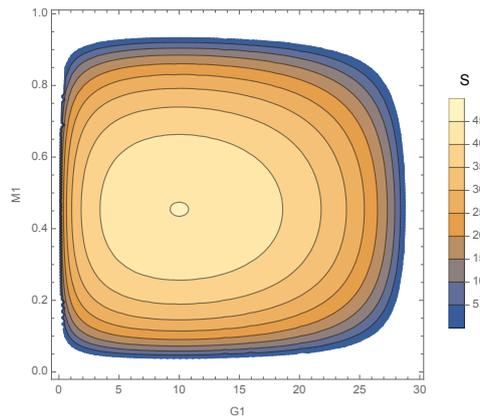
It is important to note that even though this example has homogeneous agents, the utility  $U_r(g, m) = \log g^\alpha m^\eta$  for the resulting representative agent is not the utility for a single agent of our Cobb-Douglas agent, which recall is  $U(m, g) = g^{\alpha-1} m^{\eta-1}$ . It is not even a function of the utility for a single agent. This is not too surprising, however, because our agents use their utility function to bias the probability rates, not to maximise it. Nonetheless, if the exponents  $\alpha$  and  $\eta$  of the Cobb-Douglas utility function are scaled up sufficiently, the agents will approach maximizing behaviour (as will be explained in Section 12).

To illustrate the consequences of entropy, Figure 2 shows the boundary of the region accessible from an initial



**Figure 2:** Boundary of accessible region for a two-part Cobb-Douglas economy from an initial endowment (anywhere on the displayed surface). The accessible region is to the right. The economies have numbers of agents in the ratio  $N_1 : N_2 = 1 : 2$ , and the same exponents  $\alpha = 1, \eta = \frac{5}{2}$ . The total goods  $G = 30$ .

endowment of a system of two Cobb-Douglas economies<sup>22</sup>. We see that adding money to either economy<sup>22</sup> is allowed, but also adding money to one while removing money from the other is allowed, within limits. More importantly, goods can be moved from one to the other but this might require adding money to one or both. Figure 3 shows contours of the entropy for division of goods and money between two Cobb-Douglas economies with no external trader. The allowed transitions are those that do not decrease the total entropy.



**Figure 3:** Contours of entropy  $S$  for a two-part Cobb-Douglas economy. The economies have numbers of agents  $N_1 = 10, N_2 = 20$ , the same exponent  $\alpha = 1$  for goods, but different exponents  $\eta_1 = \frac{5}{2}, \eta_2 = \frac{3}{2}$  for money. The total goods  $G = 30$  and total money  $M = 1$ .

<sup>22</sup>The choice  $\alpha = 1$  is perhaps unrealistic, because it implies the utility is independent of the amount of goods, but similar figures are obtained for  $\alpha > 1$ .

The reader may recognise (4) as the entropy of an ideal gas (discussed earlier), where  $N$  is the number of molecules,  $G$  represents volume of the container,  $M$  represents internal energy,  $\alpha = 1$ , and  $\eta = 1/(\gamma - 1)$  for ratio  $\gamma$  of specific heats (e.g.  $\gamma = \frac{7}{5}$  for a diatomic gas), modulo a conventional additional term  $N\gamma/(\gamma - 1)$  that has no effect if  $N$  is constant. Indeed, that analogy is why we chose this entropy function to try in (4). The analogy should not be pushed too far, however. In particular, in an accessible change for a gas driven by an external device whose only net change is in the height of a weight (as in [LY02]), the volume available to the gas can change, whereas for our exchange economies, the quantity of goods should return to its initial value.<sup>23</sup>

One could ask why the entropy of the trader was not considered in this analysis, but the assumption that the trader has unlimited goods and money can be regarded as saying that its entropy is independent of how much goods and money it has. More precisely, we are assuming that the trader has no preferences for amounts of goods and money, so all changes are allowed, which means all its states must have the same entropy. Strictly speaking, to talk about entropy of the trader we require it to be an extensive trading economy.

Another question is what rules we have perhaps implicitly placed on the trader? For example, if the trader were willing only to sell then  $\Delta G$  would be positive whatever price was offered, and thus for  $\mu > \mu_c$  the change in the above entropy would be negative. There is a risk of the trader acting as a Maxwell demon, but perhaps the effects are smaller than  $O(N)$ . We leave such questions for future thought.

Instead of guessing an entropy function as above, and showing that it fits, a more principled way to derive the entropy function is via the strategy of proof of Theorem 1 of [LY02]. We do that in Appendix A.

This toy example also has wider significance. We have stressed that thermoeconomics, like thermodynamics, is defined purely at the “macro” level, and does not depend on specific microfoundational assumptions. This makes macro analysis possible for systems for which a detailed microfoundational analysis is infeasible—which will typically be true for realistic economic systems, in the light of the heterogeneity and complexity of human behaviour.

It is nonetheless reassuring to find an explicit illustration of how there can be a bridge between the micro and macro accounts, where the micro-foundations are sufficiently simple to be tractable—just as statistical mechanics provides a micro-foundation for thermodynamics of sufficiently simple physical systems (e.g., ideal gases).

A purely macro-analysis is sufficient to derive the second law, and a variety of specific economic results in following sections; and macro quantities, such as economic entropy and temperature can in principle be measured directly—and these issues will be discussed further below. But a micro-foundation, where available, is nonetheless required to explain the specific form of the entropy function  $S$ , and hence the detailed behaviour of an economic system.

This raises the interesting question of the impact of different stylized microfoundational assumptions on aggregate behavior and how this behaviour aligns with the macro description in a thermoeconomic framework, and in particular, the question of how different microfoundational assumptions modify the entropy function. Many such assumptions are possible. Starting from our Cobb-Douglas model, for example, different agents  $i$  could have different  $\alpha_i$  and  $\eta_i$  (this leads to relatively straightforward extension of the above results, in particular the entropy has the averages of  $\alpha_i$  and  $\eta_i$  in place of  $\alpha$  and  $\eta$ ). Exchanges could be restricted to change of goods and money of opposite signs. Agents could make just a fraction  $f_i$  of their holdings available at each interaction. There can be more than one sort of good (in addition to money). Goods (and money) might come in discrete amounts. Moreover, agents might be subject to behavioural biases, such as a status quo bias in favour of whatever goods they hold at the time [SZ]. We anticipate that, unlike this toy example, the preferences of agents could result in an entropy function that is not just a sum of functions of the quantity of each type of good.

While the Cobb-Douglas set-up explored here yields analytically tractable results, the implications of more complex micro assumptions will not typically be analytically tractable and will require simulations. This is an important direction for our on-going and future work. In the rest of the present paper, we largely set microfoundations aside, and return to explore the implications of the purely “macro” thermoeconomic theory.

## 11 Convexity assumptions

To reap further rewards, we need a few more technical axioms from [LY02]. A reader who is happy to assume that  $S$  is differentiable with respect to relevant variables and that  $S$  is concave, will not need this section, but we include it for logical completeness.

Firstly, we assume A7 that for  $t \in (0, 1)$ ,  $(tX, (1 - t)Y) \preceq tX + (1 - t)Y$ . This seems straightforwardly true: the economy formed by putting two economies together is accessible from the pair of separate economies. A consequence of A7 is that the “forward sector”  $A_X = \{Y : X \preceq Y\}$  of any state  $X$  is convex: that is, for  $t' \in (0, 1)$ , if  $W \in A_X$ , and  $Z \in A_X$ , then  $t'W + (1 - t')Z \in A_X$ . Moreover, because the accessibility relation  $\preceq$  implies

<sup>23</sup>It is interesting to consider the possibility that, instead of amount of goods, the number of agents in an economy may be more closely analogous to volume. Thus, population size could be represented explicitly, and allowed to change, in the short term through migration, and in the longer term, through births and deaths. Questions of population change and flow are, of course, of considerable relevance to analysis in political economy and this is a direction deserving study. Note, though, that the current analysis allows the possibility that agents own goods and money in more than one economy, so it is not yet clear what it would mean to assign them to a particular economy.

non-decreasing entropy (by the second law), A7 immediately implies that  $S(tX + (1-t)Y) \geq tS(X) + (1-t)S(Y)$ . That is, the entropy function  $S$  is concave, a result that we shall use at various points below.

Secondly, we assume A9 that each forward sector  $A_X$  has a unique support plane at  $X$ . This tangent plane is furthermore assumed to be a Lipschitz-continuous function of  $X$  as described in [LY02]. This assumption perhaps requires economic justification.

Thirdly, we assume A10 that the boundary of each forward sector  $A_X$  is connected. This technical assumption looks innocuous in an economic context—at least, it is not clear under what plausible economic circumstances this assumption would be violated.

Lastly, we need A15 that for all pairs of systems  $A, B$  and states  $X$  of  $A$  there is a state  $Y$  of  $B$  with  $X \equiv Y$ . This is natural in an economic context: on putting the two systems into financial contact, nett money flow establishes financial equilibrium; if the total amount of money is varied to put  $A$  into state  $X$  then the resulting state of  $B$  will be the desired  $Y$ .

## 12 Economic Temperature

The second reward, at which we hinted in Section 6, is that there is a positive function  $T$  of state such that when two simple systems  $A, B$  are in financial contact there is nett money flow from  $A$  to  $B$  iff  $T(X_A) > T(Y_B)$ , and  $A, B$  are in financial equilibrium iff  $T(X_A) = T(Y_B)$ . In keeping with thermodynamics again, we call  $T$  *temperature*.

Under the axioms of the previous section,  $S$  can be proved to be differentiable with respect to  $M$  [LY02], and the function  $T$  is given explicitly by

$$\frac{1}{T} = \frac{\partial S}{\partial M},$$

holding amounts of all goods fixed except money  $M$ .

Note that  $T$  has unit  $\odot$  over entropy, which depends on both the scales for money and entropy. Nevertheless, ratios of temperatures are uniquely defined, independently of the scales. This implies that one can fix a scale for temperature by one reference system in a reference state.<sup>24</sup> Then, as promised, a fixed scale for entropy follows: it is the scale such that the change in entropy  $\Delta S$  is  $\Delta M/T$  for an infinitesimal change  $\Delta M$  in money with fixed goods. The unit of temperature is arbitrary (just as the unit of temperature, the Kelvin, is arbitrary in physics), but then entropy has units of money over temperature. We consider it simplest to measure temperature in units of money. For example, the reference system could be a Cobb-Douglas economy with  $1\odot$  per agent and then, as we will see shortly, it makes sense to define its temperature to be  $\eta^{-1}\odot$ . This has the consequence that entropy becomes unitless.

Following Skilling (personal communication) for physical thermodynamics, we call

$$\beta = \frac{1}{T}$$

the *coolness* of the economy. Its economic interpretation is the marginal aggregate utility of money (analogous to the marginal utility of money for an individual agent, a concept that goes back to at least [Hot]).

It is useful to refer back to the toy example of Section 10, with Cobb-Douglas agents where the probabilities of trades are proportional to their utilities  $U(m, g) = g^{\alpha-1}m^{\eta-1}$ , for amounts  $g$  of good and  $m$  of money. Then the resulting entropy function  $S = N \log(M/N)^\eta (G/N)^\alpha$  implies that

$$\beta = \frac{N\eta}{M}, \quad T = \frac{M}{N\eta}.$$

Note that the toy economy has the special feature that temperature is proportional to the amount of money per agent and independent of the amount of goods. To give an interpretation of temperature here, if one multiplies both exponents  $\alpha$  and  $\eta$  by a constant  $c > 1$  then the probability rates and the stationary distribution, proportional to the utilities, will be more strongly dominated by transitions corresponding to maximising utility. Thus, with increasing  $c$ , the degree of noise will decrease, and the transitions will be closer to pure utility-maximizing trades. If  $\alpha$  and  $\eta$  are multiplied by  $c$ , the temperature  $T = \frac{M}{N\eta}$ , is divided by  $c$ . Thus we see for this example that temperature quantifies deviations from utility maximisation by the agents. So standard economics with noiseless utility-maximizing agents might be considered to be the case of zero temperature.<sup>25</sup>

We will see that temperature being proportional to amount of money and independent of amounts of goods is a feature of any economy with “pure money”, to be discussed in Section 15.

When economies  $A$  and  $B$ , with  $T(X_A) > T(Y_B)$ , are put into financial contact there is nett money flow from  $A$  to  $B$  before they equilibrate. If the changes are made quasistatically (that is, via a sequence of equilibrium

<sup>24</sup>In physics, the standard reference state is the triple point of water, the temperature and pressure at which water can exist as a solid, liquid or gas).

<sup>25</sup>Note though that the context in which Hotelling [Hot] introduced marginal utility of money was utility maximisation subject to a given budget; the marginal utility of money entered as a Lagrange multiplier to relate changes in money to changes in utility and there is no reason for it to be zero.

states) so that the changes are reversible, then the money flow  $dM_{AB}$  from  $A$  to  $B$  in each step is given by  $T_B dS_B = dM_{AB} = -T_A dS_A$ . Then we see that  $A$  loses entropy,  $B$  gains entropy, and in fact  $B$  gains more than  $A$  loses, because  $B$  is at a lower temperature. The same is true however the changes are made. The cooler economy gains entropy at the partial expense of the hotter one. On the other hand, at the level of individual agents the benefit flows in the opposite direction: agents with money in the cool economy, where money initially has more value, find that the value of their money is eroded by an influx of money from the hot economy; conversely, agents who move money from the hot to the cool economy find the value of their money goes up, which is of course why they have the incentive to do so.<sup>26</sup> Note that the coolness of an economy should not be confused with the buying-power of money, because the latter depends also on what goods are available.

We can simulate financial contact of two of our toy economies, with exponents  $\eta_A, \eta_B$  for money not necessarily equal, by having agents  $i \in A$  interacting with agents  $j \in B$  at rate  $k_{ij}$  (many of these coefficients could be zero) and dividing their combined money according to distribution proportional to  $m_i^{\eta_A-1} m_j^{\eta_B-1}$  (a Beta-distribution again). On average, this results in a split of the money between  $A$  and  $B$  in the ratio  $\eta_A : \eta_B$ .

Note that because the ideal trader of our accessibility notion has constant entropy, their temperature is infinite. This might seem to imply infinite money flow from the trader on putting them in financial contact with an economy. But we do not use the trader in this way. Rather, the trader decides at what price to buy or sell.

To end this section we discuss how the concept of temperature depends on the choice of numeraire. Suppose one economy  $V$  has a scarcity of apples and surfeit of bananas, and the other  $W$  has the opposite. If apples are used as money, then financial contact will mean that apples flow from  $W$  to  $V$ , and hence that  $W$  is hotter than  $V$ . But if bananas are used as money, then financial contact will mean that bananas flow from  $V$  to  $W$ , and hence that  $V$  is hotter than  $W$ . Thus, even the ordinal notion of temperature is dependent on the choice of money. Note, though, that the choice of money does not affect which state transitions are possible in the economy (subject to the desirability assumption of Section 3), thus that *entropy*, and in particular the second law, is independent of the choice of numeraire (up to positive affine transformations).

### 13 Money capacity and inflation

Having obtained a concept of temperature for an economic system, we define a subsidiary notion that will be useful in analysing how changes in the amount of money in an economy affect its value. Following the terminology of heat capacity of a physical system, we define the *money capacity* of an economic system to be

$$C = \frac{dM}{dT}$$

for changes in temperature keeping amounts of all goods fixed except money.<sup>27</sup> Money capacity has a particularly simple form in our toy economy,

$$C = \eta N.$$

Note that money capacity is positive, by the following argument. From [LY02] and Section 11,  $S$  is concave, in particular with respect to  $M$  at constant goods. Now  $dS/dM = \beta$ , thus concavity of  $S$  gives  $d\beta/dM \leq 0$ . But  $\beta = 1/T$ , so  $dT/dM \geq 0$  and hence  $C > 0$ .

Money capacity is relevant to various economic questions. For example, money capacity is inversely related to the inflationary impact of adding money to an economy, while keeping the quantities of all other goods the same. The added money could flow in from a neighbouring economy or from the government printing money. Recall that the marginal utility of money,  $\beta$ , is the inverse of temperature, i.e.,  $1/T(M)$  (writing  $T$  as a function of  $M$ , as quantities of goods are held constant). Thus for one unit of money added to the economy (with no change in other goods), the change  $\Delta\beta$  in the marginal utility of money is  $\frac{d(T^{-1})}{dM} = -T^{-2} \frac{dT}{dM} = -\frac{1}{CT^2}$ , so that the relative change is

$$\frac{\Delta\beta}{\beta} = -\frac{1}{CT}.$$

Notice that this reduction in the value of money corresponds to inflation: indeed, the money capacity is a key ingredient deciding the inflationary effect of adding money to an economy, at least in our exchange economies. We can define the *inflation rate* to be

$$I = \frac{1}{T} \frac{dT}{dt}$$

(equivalently  $-\frac{1}{\beta} \frac{d\beta}{dt}$ ) with  $t$  representing time. Thus for rate of increase  $\dot{M}$  in money,  $I = \frac{\dot{M}}{CT}$ . Note that this definition of inflation is entirely abstract and not dependent of choices of specific price indices and their associated weighted baskets of goods.

<sup>26</sup>The government of an economy might attempt to prevent such flows by imposing capital controls.

<sup>27</sup>Note, though, that this quantity is a derivative, and hence might more aptly be called marginal money capacity (and the physical concept would analogously be marginal heat capacity).

Money capacity also provides a useful relation between temperature and entropy, namely

$$\frac{dS}{dT} = \frac{C}{T},$$

where again the derivative is taken with fixed amounts of all goods except money. The proof is that the change  $dM$  in money can be written as  $TdS$  from the formula for  $T = \frac{dM}{dS}$ , and also as  $CdT$  from the formula for  $C = \frac{dM}{dT}$ .

From this relation and positivity of  $C$  we deduce that  $S$  increases with  $T$ . In words, the increase in aggregate utility from unit increase in temperature is the money capacity divided by the temperature.

Money capacity also turns out to be important in quantifying fluctuations of the amount of money in a subsystem, as will be discussed in Section 16.

Finally, we comment that the above proof of positivity of money capacity  $C$  includes the possibility of  $C = +\infty$ . This is a situation analogous to physical systems at a phase transition. For example, on heating a solid at constant rate, its temperature rises until it reaches its melting temperature; the temperature remains constant there until it has all melted and then goes back to rising. The heat capacity is infinite when part melted and part not. The possibility of phase transition in economics merits exploration.

## 14 Values and market prices for goods

The third reward is that for each type of good there exists a market price,  $\mu = -dM/dG$  (i.e., the price is a rate of exchange between the good and money) at which the economy will make no nett exchange between the good and money with an external trader. Thus, we will arrive at an analogue of the celebrated general equilibrium result of Arrow and Debreu, albeit with a simple exchange economy, and with a statistical notion of equilibrium. Notice that this approach makes no assumptions about the rationality or otherwise of individual behaviour.

To derive this result, we first define a “value” for each type of good, actually a marginal value for one more unit of the good. For each type of good (other than money), with amount  $G$ ,  $S$  is differentiable with respect to  $G$  (under the axioms of Section 11). Let

$$\nu = \frac{\partial S}{\partial G},$$

keeping the amounts of all other goods fixed, including money. We call  $\nu$  the *value* for the good. This is an important concept in our theory. It is the analogue for goods of the coolness  $\beta$  for money (Section 12). Our quantifications of value of goods and of money provide a new theoretical basis for discussions of value that go back to Adam Smith’s “Wealth of Nations” (see, e.g. [Pag]).

Note that in contrast to money, however, the value of a good might be negative. This caters for goods that are undesirable or perhaps desirable if other types of good are scarce but not if they are plentiful. For example, a bulky good might be desirable if no alternative is available but undesirable if compact alternatives are cheap and plentiful; or a good might be desirable only when complementary goods are available (side-cars may be desirable if motorcycles are available but undesirable “junk” if not).

If we allow changes in the amounts of good and money by connection to an ideal trader then

$$dS = \beta dM + \nu dG. \tag{5}$$

There will be no spontaneous nett exchange of goods and money between the economy and the trader if and only if the exchange between money and goods is reversible, i.e., leads to no change in entropy ( $dS = 0$ ). By rearranging (5) with  $dS = 0$ , we can see that a reversible exchange between money and the good is possible iff the trader is offering a *market price*  $\mu = -\frac{dM}{dG} = \frac{\nu}{\beta}$ .<sup>28</sup>

For the example of Section 10 where we explored our simple Cobb-Douglas economy,

$$\nu = \alpha \frac{N}{G}, \quad \mu = \frac{\alpha}{\eta} \frac{M}{G}.$$

This  $\mu$  corresponds to the critical price  $\mu_c$  (3) we found in Section 10.

In view of the centrality of the second law in thermoeconomics, we have focused on changes in economic entropy  $S$  in (5). By contrast, it is common in thermodynamics to give priority to energy rather than entropy. In an economic context, the analogue of the first law of the conservation of energy in physics is the much less interesting conservation of money. Indeed, in an exchange economy, money and all goods are conserved by assumption. Nonetheless, if we choose to give priority to money over entropy then the analogous relation to (5) can be written

$$dM = T dS - \mu dG. \tag{6}$$

The relations (5) and (6) extend naturally to a vector  $G$  of amounts of goods and corresponding covectors  $\nu$  of values and  $\mu$  of prices. Furthermore, the relation (5) extends to exchange between any pair of goods, not necessarily including money. The coolness  $\beta$  is simply the special case of the value for money. For exchange of two types of good  $G_1, G_2$ , then  $dS = \nu_1 dG_1 + \nu_2 dG_2$ , so reversible exchange happens iff the trader offers the exchange ratio  $\nu_2/\nu_1$  for amount of  $G_1$  per amount of  $G_2$ .

---

<sup>28</sup>Note that as mentioned in Section 10, the trader has constant entropy so its state does not affect the allowed changes.

## 15 Pure money

What differences does it make if money has no intrinsic value (e.g. paper or digital money), as opposed to some good like gold? We argue in this case that for fixed vector  $G$  of amounts of goods, (i) the price  $\mu$  of any type of good, and (ii) the temperature  $T$ , should be proportional to the amount  $M$  of money in the economy, because the value of a unit of money will be determined by how large a fraction it is of the total quantity of money in the economy. Thus, declaring each aurum to be  $\lambda$  new-aurum should make no difference other than to multiply all prices and the temperature by  $\lambda$ —indeed, it is scarcely more than a trivial notational change, like deciding to use the word “dollar” to refer to cents.

The above two conditions (i) and (ii) imply that the value  $\nu = \mu/T$  for each type of good is independent of  $M$ , and that the coolness  $\beta$  is  $K(G)/M$  for some function  $K$  of  $G$ . Using  $\nu_j = \frac{\partial S}{\partial G_j}$  and  $\beta = \frac{\partial S}{\partial M}$ , we deduce from the first condition that  $\frac{\partial \beta}{\partial G_j} = 0$ . Thus, from the second condition,  $K$  is independent of the amounts of goods. As a result,

$$S = K \log M + F(G)$$

for some constant  $K > 0$  and function  $F$  of the amounts of goods. An example is the toy economy of Section 10 (where also the other good is pure).

The above holds for an economy with fixed number  $N$  of agents. The dependence on  $N$  should be extensive, that is

$$S = N(k \log M/N + f(G/N))$$

for some constant  $k$  and function  $f$ .

Money is likely to be “pure” to a good approximation in modern economies, and we discuss the implications at various points below. One implication that we can make already is that for pure money, the money capacity  $C = 1/K$  is constant. In particular it is never infinite. But historically, some currencies have certainly not been pure, for example silver. The requirement to use silver as currency for trade with China clashed with its increasing use in the UK as silverware for serving food and drink, a problem that notably Adam Smith and Isaac Newton addressed [Fay, Pag].

## 16 Financial thermometers

Having deduced the existence of temperature in an equilibrium economy (and hence a concept of the value of money, and the meaning of inflation, as in Sections 12 and 13), a natural question is whether and how temperature can be measured. A difficulty is that in practice an economy is in general not in equilibrium because it has inputs and outputs. Moreover, rather than the pure exchange economies we are treating here, real economies have production of raw materials, manufacturing, consumption, services, and so on. Nonetheless we address the question here for equilibrium exchange economies, with the hope of later generalisations.

Here, we focus on the question of how financial temperature can be measured in principle—i.e., how to construct an “ideal” financial thermometer. We leave the question of how well such an ideal can be approximated in practice for later work. To measure the temperature of an economy  $A$ , one could imagine floating a relatively small economy  $s$  on a ship, putting it in financial contact with  $A$  (with no transfer of goods) and studying the changes to  $s$  to infer the temperature of  $A$ .<sup>29</sup> The smallness of the ship thermometer  $s$  is needed (just as for a physical thermometer) so that the contact with  $s$  does not have a significant effect on the temperature of  $A$ , the economy whose temperature is being measured.

The intuition is that money will flow into, or out of, the ship thermometer from the mainland, bringing the temperature of  $s$  into balance with the temperature of  $A$ . If  $T_s < T_A$ , people in  $A$  will tend to move their money to  $s$  because goods are cheaper there (that is, the aggregate marginal utility of money is greater); this influx of money will in general raise the temperature (by positivity of money capacity, Section 13), until the temperature reaches that of the mainland. If  $T_s > T_A$ , the opposite will happen. Then the amount of money on the ship after equilibration is an increasing function of the temperature, so with calibration it can be used to measure the temperature.

Perhaps the amount of money on the ship is not readily visible, however. In this case, one could instead measure the price of some good on the ship, which is more likely to be visible (after averaging), and convert it to temperature. One can make a plausible story that if  $T_s < T_A$  then the influx of money can be used to buy goods in  $s$ , thereby driving up their price until  $T_s$  reaches  $T_A$ , though we have not proved that the price has to increase. If  $T_s > T_A$ , the opposite can be expected to happen. Thus the market price of each good on the ship will settle down to a function of the temperature of the mainland  $A$ , so one can read off the temperature from the market price after calibration.

This calibration is very simple in the special case where the currency is “pure”, with no intrinsic value except as a medium of exchange (as in the previous Section). The amount of pure money in an economy is proportional

---

<sup>29</sup>Calling  $s$  a ship thermometer is a natural metaphor when measuring the temperature of mainland economies with sea frontage, but of course applies quite generally.

to the temperature of that economy. Thus, if the amount of money in the ship thermometer is easily measurable, it provides a direct measure of the temperature of  $A$ . Indeed, from the previous section on pure money,  $T = M/K$ , so the temperature on the ship is proportional to the amount of money. Furthermore, for an economy with pure money, the price of a good is  $\mu = T \frac{\partial F}{\partial G}$ . The amounts of goods on the ship do not change, only money, and  $F$  is independent of  $M$ . So we deduce that the price of any good is proportional to the temperature.<sup>30 31</sup>

Note that this measure of temperature agrees with a conventional price index for comparing value of money in two economies, say  $A_1$  and  $A_2$ , if the goods in  $A_1$  and  $A_2$  are the same but all prices differ by a fixed factor. But our approach is much more general. Indeed, there need not be any overlap whatever between the goods in  $A_1$ ,  $A_2$ , or indeed  $s$ .

As an example of a ship thermometer, we can use the toy economy of Section 10. Combining previous results,  $\mu = \frac{\alpha N}{G} T$ , and noting that  $N$  and  $G$  remain fixed, we see that  $\mu$  gives a result proportional to  $T$ . So the market price of the good on the ship can be used to measure the temperature of the mainland economy. Alternatively, if one can measure the amount  $M$  of money on the ship then one can use the more direct formula  $T = \frac{1}{N\eta} M$ , proportional to  $M$ . But almost any quantifier on any type of ship thermometer would respond in some nontrivial way to the temperature of  $A$ .

As mentioned above, the scale for temperature is arbitrary (at least from our construction). So what can really be measured is ratios of temperatures. Our ship thermometer allows us to do this.

If we choose to use the amount  $M$  of money on the ship to infer the temperature, it is worth noting that if the ship thermometer is too small then fluctuations in  $M$  will make the temperature difficult to measure (and indeed, prices on the ship will be difficult to measure also). Indeed, the analogue of Einstein's formula for fluctuations in energy of a subsystem is that the variance

$$\text{Var} M = CT^2,$$

where  $M$  is the amount of money on the ship,  $C$  is the money capacity of the ship, and  $T$  is the (common) temperature of the ship and  $A$ . Using the toy economy of Section 10 for the ship, the measured temperature  $T_m = \frac{M}{\eta N}$  and  $C = \eta N$ , so

$$\text{Var} T_m = \frac{\text{Var} M}{\eta^2 N^2} = \frac{T^2}{\eta N}.$$

It is natural to compare  $\text{Var} T_m$  to  $T^2$ , hence to divide both sides of the above equation by  $T^2$ . Thus the conclusion is that the temperature will be measured more precisely if the number  $N$  of agents on the ship is larger and if their exponent  $\eta$  for money preference is larger. As with the design of physical thermometers, however, there are other considerations, such as how much the financial contact with the ship disturbs  $A$  and how fast the ship temperature equilibrates with that of  $A$ . We have begun to explore these issues in simulated exchange economies and plan to report on it in future work.

Note that the above analysis of fluctuations in money for a subsystem depends on an identification of entropy as representing the logarithm of available volume. That is true for the toy economies, where the entropy comes out to be the logarithm of the normalisation factor for the stationary distribution, but it is not clear in general. It is not a consequence of standard thermodynamics either. But assuming this, the discussion of fluctuations extends to other quantities, such as price:  $\text{Var} \mu = -T \frac{\partial \mu}{\partial G}_{|s}$ , where the partial derivative will be calculated in Section 20 (and shown to be non-positive and to contain another factor of  $T$ ). There are even fluctuations in the entropy of a subsystem, and there are computable covariances between fluctuations (but the relative sizes of all of these scale like  $O(1/N)$  in the thermodynamic limit, as was made explicit for  $T_m$  above).

Note also that the above formulae for variances both give results proportional to  $T^2$ . This shows that, under the assumption that entropy represents logarithm of available volume, temperature is not just the inverse marginal utility of money but is also connected fundamentally with the level of activity at equilibrium.

A natural question is whether one obtains the same temperature wherever one measures it in an economy. One might imagine that if the agents in one port have different preferences from those in another port then one would find different temperatures in the two ports. But under our assumptions that the economy is in equilibrium and there are no internal barriers to money flow in an economy, the equilibration under financial contact with the ship thermometer produces the same state on the ship in both ports.

We reiterate the contrast between our approach to measuring economic temperature (remember this is the inverse of the marginal utility of money) and conventional measures of value of money within countries (e.g., the Consumer Price Index) or cost of living ratios between countries (e.g. Penn World Tables, World Bank's International Comparison Program), which depend on rather arbitrary choices of baskets of goods. By contrast, the thermo-economic concept of temperature is independent of any of basket of goods—indeed there is no need for

<sup>30</sup>The ship thermometer with pure money is therefore analogous to an ideal gas thermometer in physics, for which volume is proportional to temperature at constant pressure

<sup>31</sup>Note that if the currency is not pure, but a good with some intrinsic value such as gold, then the picture is not so simple. Then an influx of gold into the ship may change the relative prices of other goods on the ship, which might be complements to, or substitutes for, gold. For example, plentiful gold might drag down the value of silver (if silver and gold are substitutes) as compared with, say, butter. These same interactions with other goods mean that the value of gold cannot be read off merely by measuring the ratio change in the quantity of gold on the ship.

two economies to have any good in common (except money) in order for us to take their temperature ratio. For realistic application, we need to extend to the case where different economies use different currencies. We make some remarks on this in Section 21.1.

To close this section, we discuss another measurement question: how one could measure the market price of a good in the ship thermometer? This may be a more practical method for measuring temperature of a real ship thermometer than measuring the amount of money on the ship, and also has independent interest. One way is to have an external trader propose a price and to monitor the nett flow of goods and money. If there is nett flow of money to the trader, the trader increases the price; if in the other direction, the trader decreases the price. By this process the trader should settle on the market price.<sup>32</sup> Again, the price should not depend on the particular agents to which the trader is connected.

## 17 Making money out of price differences

If two economies have some good in common, as well as money, and the price of the good is different in the two economies, then it is clear that an external trader can make money from the price difference. The trader buys goods in economy  $A$  at price  $\mu_A$  aurum where they are cheaper and sells them in the economy  $B$  at price  $\mu_B$  aurum where they are more expensive (i.e.,  $\mu_A < \mu_B$ ). For infinitesimal changes (so that the prices do not change), one aurum purchases  $1/\mu_A$  goods in economy  $A$ , which are then sold for  $\mu_B/\mu_A$  aurum in economy  $B$ . Thus the trader makes  $\mu_B/\mu_A - 1$  aurum.

In general, however, as trading continues, the prices will change because the amounts of goods and money in the two economies change. How much money can the trader make in this way?

In the simplest context with just one good besides money, the state of the joint system follows the path of increasing  $G_B$  with  $dG_A = -dG_B$ ,  $dM_A = \mu_A dG_B$ ,  $dM_B = -\mu_B dG_B$ ,  $dM_T = (\mu_B - \mu_A)dG_B$  (using subscript  $T$  for the trader) until  $\mu_B = \mu_A$ . The path is contained in the intersection of the surfaces of constant entropy for the two economies, because trade at market prices is reversible and so does not change entropy. Thus one way to describe the final point is that it minimises  $M_A + M_B$  subject to conservation of  $S_A$ ,  $S_B$  and  $G_A + G_B$ . And the amount of money made by the trader is the difference between the initial and final amounts of money in the economies.

This description extends to the case of more than one tradeable good, where the total amounts of each good are conserved. The first-order conditions for a minimum enforce equality of prices in the two economies for each good at the final point. The existence of the final point and its connectedness to any initial condition with the given entropies requires some more analysis, of course. In particular, it is proved in Section 20 that the price of a good decreases as the quantity of that good increases by trade at market price (i.e. taking into account compensating change in money), and vice versa.

We have seen that the maximum amount of money,  $M_T$ , the trader can extract through exchanging goods and money is achieved by minimizing  $M_A + M_B$  subject to conservation of  $S_A$ ,  $S_B$  and  $G_A + G_B$ . But can the trader do better? From the second law of thermoeconomics, one might ask whether the trader can find a way of minimising the total amount of money left in the economies  $A$  and  $B$  subject to conserving goods and *total* entropy,  $S_A + S_B$ , rather than the more restrictive condition of conserving the individual entropies  $S_A$  and  $S_B$ . This is indeed achievable by the trader, but not in general by the above strategy. It requires a more general approach, to be explained in Section 19. To get there we first explain how one can make money out of temperature differences with no transfer of goods.

## 18 Making money out of temperature differences

The industrial revolution was powered by the observation that one can extract useful work from physical temperature differences. Moreover, in physics, one important consequence of the thermodynamic concept of temperature<sup>33</sup> is that it gives a formula for how much work can be extracted from temperature differences (more precisely from temperature ratios). Indeed, the efficiency of an ideal heat engine (called a Carnot cycle) is how Carnot introduced the thermodynamic temperature scale. So the process of extracting work from a temperature difference turns out to be both practically and theoretically very significant.

It is therefore natural to ask whether there is an interesting analogy in economics. As spelt out in the previous section, a trader can make money out of price differences by buying goods in an economy where they are cheaper and selling where they are more expensive. If price differences correspond to temperature differences then one could view this as making money out of temperature differences.

But the question here is narrower: how much money can an external trader extract from a temperature ratio between two economies, without any nett change in amounts of goods in either economy. That is, how can the

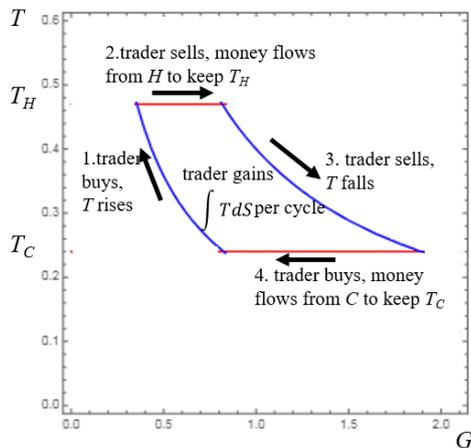
<sup>32</sup>This is similar to the use of a Wheatstone bridge in physics to measure resistance.

<sup>33</sup>as opposed to purely empirical measures of temperature.

trader profit purely from the flow of *money* from the hot economy to the cold economy (just as a heat engine extracts work purely from the flow of heat between hot and cold bodies)?

To do this, we propose an economic Carnot cycle. The trader will use a “boat economy,”  $B$  (essentially our ship thermometer of the previous section, but independent of its use to measure temperature), which can shuttle costlessly between two mainland economies  $H$  (hot) and  $C$  (cold) (perhaps across a narrow strait), with temperatures  $T_H > T_C$ , and put into financial contact with either. The trader will make reversible trades with  $B$ , but not with either  $H$  or  $C$ ; there will, though, be flows of money from  $H$  to  $B$  and from  $B$  to  $C$ .

The trader can extract money from the economic temperature difference between  $H$  and  $C$  by applying the following four steps. First, the trader moves  $B$  near-reversibly to the temperature of  $H$ , by continually buying goods slightly above the current market price in  $B$ , thus gradually, and hence reversibly, pushing up prices on the boat until its temperature is  $T_H$ . Second, now that  $B$  and  $H$  are at the same economic temperature, the trader puts them into financial contact and begins selling goods to  $B$  slightly under the market price. Now there is less money and more goods in  $B$ , marginally increasing the buying-power of money, and hence drawing in money from  $H$ , while keeping the temperature of  $B$  at  $T_H$ . The third step is for the trader to disconnect the financial contact between  $B$  and  $H$ , and to continue reversibly selling goods to  $B$ . Now the buying power of money on  $B$  increases, as the boat contains many goods and little currency with which to buy them—i.e., the temperature of  $B$  reduces to  $T_C$ . The fourth step is for the trader to put  $B$  and  $C$  in financial contact and to begin buying goods from  $B$ , thus marginally lowering the buying power of money in  $B$  and thus leading to money to flow from  $B$  to  $L$ . The trader can then disconnect  $B$  to  $L$  and begin the cycle again. This process is illustrated in Figure 4.



**Figure 4:** An economic Carnot cycle in the plane of amount  $G$  of goods and temperature  $T$  for an island economy with  $\alpha = 2$ ,  $\eta = 5/2$  operating between two mainland economies at temperatures  $T_H = 0.47$  and  $T_C = 0.24$ .

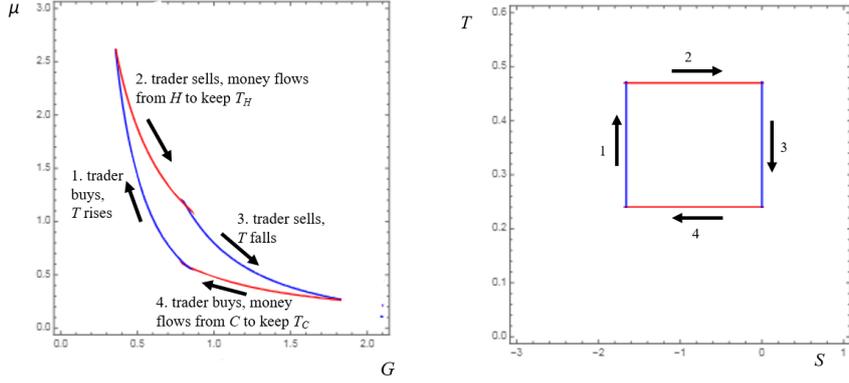
The nett effect of each cycle is to take some money from  $H$  and distribute it between  $C$  and the trader; the trader obtains money by feeding off the temperature difference. If carried out reversibly, the trader takes a fraction  $(1 - T_C/T_H)$  of the money leaving  $H$ . Namely, the money leaving  $H$  is  $-T_H dS_H$  and the money entering  $C$  is  $T_C dS_C$ , and for reversible changes the total change in entropy is zero, so  $dS_H = -dS_C$  ( $B$  returns to the same state after a cycle so there is no nett change in  $S_B$ ); hence the ratio of the money leaving  $H$  to the money entering  $C$  is  $T_H/T_C$ . This is the same formula as for the efficiency of a physical Carnot cycle, because it relies on the same mathematics. In a future paper, we intend to report on demonstration of a Carnot cycle in computer simulations of exchange economies, and exploration of its efficiency.

In physics, it is most common to plot a Carnot cycle in the plane of pressure and volume. That would correspond here to price and amount of goods. But for some purposes the best is the plane of temperature and entropy. So we replot the figure in these two ways in Figure 5.

More generally, in a cyclic process, the money a trader can extract per cycle with reversible moves is the area of the loop in the  $(S, T)$ -plane, as in Figure 5(b). To see this, denote the money and goods on the boat by subscript  $B$  and for the trader by subscript  $T$ . For the boat economy,  $dM_B = T dS - \mu dG_B$ , and reversible moves imply that the change in the trader’s money  $dM_T = \mu dG_B$ . So we deduce that  $dM_T = T dS - dM_B$ . But in a cycle,  $M_B$  returns to its initial value, so  $\int dM_T = \int T dS$ .

Note that the above discussion in terms of a cycle assumes that the hot and cold economies are sufficiently large that their temperatures are not affected by the flow of money in one cycle. In reality, the temperatures of the two economies would change slightly and the cycle would have decreasing amplitude in the temperature.

A Carnot cycle can also be used in reverse, just as in physics (e.g. refrigerator or heat pump), to create economic temperature differences by putting in money, or more generally to drive money flow against a temperature difference. The trader puts in money  $M_T$ , the cold economy puts in  $M_C$  and the hot economy receives  $M_H =$



**Figure 5:** Carnot cycle replotted in the planes of (a) amount  $G$  of goods and price  $\mu$ , (b) entropy  $S$  and temperature  $T$ .

$M_T + M_C$ , with the ratio  $M_H/M_C$  anything less than  $T_H/T_C$ . In particular, the hot economy can receive anything up to  $M_T/(1 - T_C/T_H)$ , a considerable enhancement on just receiving  $M_T$  directly from the trader, especially if  $T_C$  is only a little less than  $T_H$ . We call the ratio  $M_H/M_T$  the *coefficient of performance*, by analogy with heat pumps.<sup>34</sup>

These observations have potentially interesting practical implications. Clearly a large and rich trader can intervene directly to change the buying power of money in a small state (say an isolated island). For example, it can make a small state “hotter” (i.e. its money having less value) simply by sending money to the smaller state. But it can generate the same effect more efficiently if it has access to another island, and a “boat” which shuttles between the islands (or some equivalent mechanism). The trader can now more cost-effectively cause inflation in one state by simultaneously causing deflation in the other (even if the second small island is initially cooler, so that this flow of money is from cooler to hotter). The efficiency gain will be a factor up to a theoretical maximum of  $1/(1 - T_C/T_H) = T_H/(T_H - T_C)$ . If the value of money is similar in each economy (i.e.,  $T_H - T_C$  is small), the maximum efficiency gain can be arbitrarily large. It is interesting to ask whether approximations to such phenomena can occur in real financial markets through government or private speculator interventions. A fuller analysis of such cases is likely to require an extension of the present account to allow that different states have different currencies, a point we touch on briefly in Section 21.1.

## 19 Trade

A key question that has motivated us from the outset is how to determine which trades are viable between two economies. In the previous two sections, we have considered special cases through which a trader can extract money from a pair of economies by exploiting differences in prices (Section 17) and differences in temperature (Section 18). Now we are in a position to develop a general account of trade, whether mediated by a trader or involving direct exchange between economies.

According to the present analysis, nett exchange of goods and money between two economies can occur if and only if it does not decrease total entropy. Thus we obtain an analogue of the traditional notion of ‘gains of trade’. Yet the gains might not benefit each economy individually. We distinguish trade that increases the entropy of all parties by the term ‘mutually beneficial trade’, or *MB trade* for short. Note that the mutual benefit here is at the level of economies, not individuals. So, for example, when economies at different temperatures are put into financial contact, money will flow from the hotter to the cooler. This may be beneficial at an individual level (because money has more value when transferred from the hotter to the cooler economy). But the outflow of money will transfer entropy (aggregate utility) from the hotter economy to the cooler. Of course, if an economy has a government, that government might try to maximize the aggregate utility of its own economy, and hence might attempt to put restrictions on such “capital flight”.<sup>35</sup> A goal of our theory is to provide a framework for quantitative analysis of the benefits of trade, including finding which trade is mutually beneficial.

<sup>34</sup>Note that for refrigerators the coefficient of performance is conventionally measured instead as  $M_C/M_T$ , which is  $M_H/M_T - 1$ .

<sup>35</sup>There are also many other possibilities. For, example, an external trader could operate a Carnot cycle to exploit a temperature difference and then use the proceeds to make unprofitable trades. Or it could operate a Carnot cycle to make or enhance temperature differences by putting in money made from profitable trades. We amplify on this below, but first talk about MB trade.

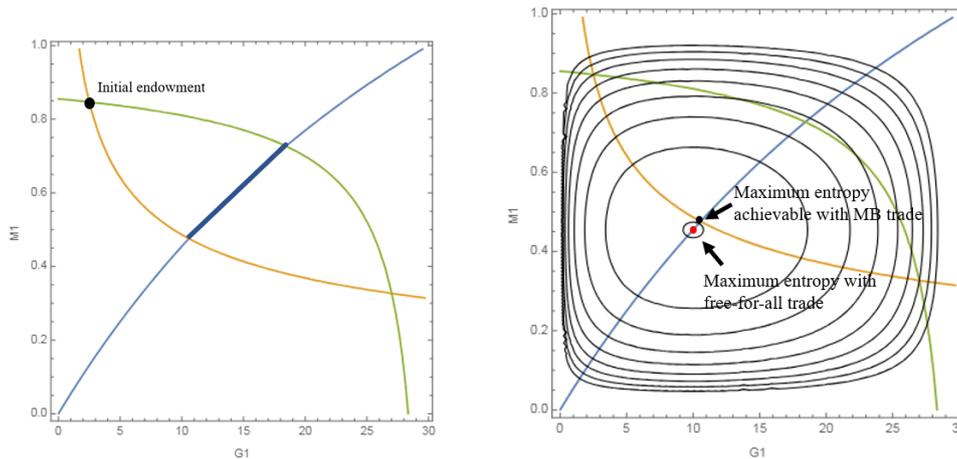
## 19.1 Mutually beneficial trade

For a single good and money, and infinitesimal trade maintaining equilibrium, there is a simple (and very natural) criterion for the entropy of both economies to increase: the goods move from the economy with the lower market price, say  $\mu_A$ , to that with the higher market price  $\mu_B$ , and the price for the trade is anything in between the two. This results from positivity of temperature and equation (6), which we rewrite here as  $T dS = dM + \mu dG$  for reversible changes. Note that in a sequence of trades the market prices might well change, so this condition has to be satisfied at all stages of the trade.

Trades might involve more than one good and not necessarily include money. Our MB condition is again that the entropy of both economies increase. Writing  $dS = \sum_i \nu_i dG_i$  for reversible changes (where one of the goods could be money, in which case its value  $\nu_i$  is  $\beta$ ), the condition for the package of transfers  $dG_i$  (which can have either sign, in particular that for money is likely to have opposite sign to most of those for goods) from  $A$  to  $B$  to be MB is that  $\sum_i \nu_i^A dG_i < 0$  and  $\sum_i \nu_i^B dG_i > 0$ . This does not have as simple an interpretation as in the case of one good and money, but allows combinations that might not be MB pairwise. Trade can (and often does) involve goods that do not exist originally in the receiving economy. It is natural to assign value  $\nu = +\infty$  for an unavailable good, but the value will become finite as soon as some of the good is imported.

Reversible trades must take place slowly—in particular, they are quasi-static, meaning that they occur slowly enough for both economies to be at equilibrium throughout the trading process. But trades need not be quasistatic. All that is required for MB is that the entropy of neither economy decreases.

MB trade is very similar to exchange between rational agents with preferences given by total preorders, often illustrated by the “Edgeworth box” [St]. Here we obtain the same picture at the macroeconomic level, with the preferences being given by the entropies of the two economies. It is illustrated in Figure 6(a) for the same pair of economies as in Figure 3. The families of curves of constant entropy (isentropes) for the two economies (of which



**Figure 6:** (a) Edgeworth box plot for the two Cobb-Douglas economies of Figure 3. The orange and green curves are *isentropes* for the first and second economy, respectively (i.e., entropy is constant along the curves). The initial endowment is at their upper left intersection. The blue curve is the Pareto frontier of equal prices in the two economies, i.e., the set of points where the isentropes for the two economies touch; it is thickened for the part achievable by MB trade from the initial endowment. (b) The same plot with contours of total entropy superimposed, from Figure 3; the maximum entropy point is inside the smallest contour of total entropy.

only one is shown for each here) are tangent to each other along a curve called the Pareto set. The tangency condition of the isentropes is that the market prices are the same in both economies. This is the condition for nett trade to stop (though equal and opposite trades may continue, at the common market price). It is important to note that there is not a unique final state from given initial state: any point on the part of the Pareto curve of tangencies of constant entropy curves between the initial constant entropy curves is a possible final state. Total entropy maximisation (subject to the total goods and money) would select a particular point of the Pareto curve, but it might not be in the allowed interval. Indeed for the chosen initial endowment of Figure 6, the maximum entropy point is not in the allowed interval, as can be seen in Figure 6(b). So a purely macro analysis gives no determinate answer to how prices will settle—presumably, factors such as the relative patience and skill of the negotiators on each side will decide the outcome.<sup>36</sup> Notice, in particular, that MB trade (without financial contact and hence flow of money) does not require that the temperatures of the two economies become equal

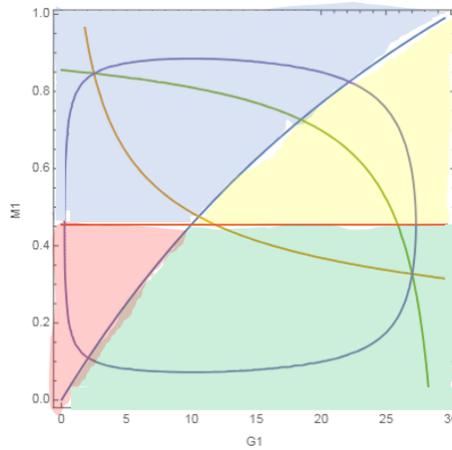
<sup>36</sup>Or, according to the theory of the Edgeworth box, the final point ought to be the one on the Pareto curve from which the common tangent to the isentropes goes through the initial endowment point, meaning that the whole sequence of trades is conducted at the price corresponding to the slope of that straight line.

(recall from Section 12 that for Cobb-Douglas economies,  $T = \frac{M}{N\eta}$ , so equal temperature corresponds to the ratio  $M_1 : M_2 = N_1\eta_1 : N_2\eta_2$ , hence for  $N_1 = 10, N_2 = 20, \eta_1 = \frac{5}{2}, \eta_2 = \frac{3}{2}$  and  $M_1 + M_2 = 1$ , we get  $M_1 = \frac{5}{11}, M_2 = \frac{6}{11}$ , which includes the maximum entropy state but nothing else on the Pareto curve).

## 19.2 General trade

Let us continue some more analysis of Figure 6(b), but dropping the requirement that trade must benefit both economies. Now our only restriction is that trade must not decrease the entropy of the system overall. Thus, in the “contour map” of overall entropy, trades can move only “up hill” or, as a limiting case, around the contours. In particular, this allows trades to lead to final states in the region below the orange curve but inside the constant total entropy curve through the initial state, where economy 1 loses entropy but economy 2 gains more. Notably, in this example, that includes the maximum entropy state itself.

It is useful to classify the possible moves of the pair of economies in this simple case with money and one good. We divide the state space into “quadrants” according to the signs of the temperature difference  $T_2 - T_1$  and price difference  $\mu_2 - \mu_1$ . See Figure 7. Recall from the end of the previous subsection that in our toy economies,  $T = \frac{M}{N\eta}$ , so the curve of equal temperatures is a line of constant  $M_1$ , namely  $M_1 = 5/11$  for the chosen parameter values. Recall also that the curve of equal prices is the Pareto curve. From a starting point other than that of



**Figure 7:** Figure 6(a) with the addition of the line of equal temperatures (red) and just one total isentrope (mauve). The (red) line of equal temperatures and the (blue) curve of equal prices divide the space into four quadrants. The top left quadrant (shaded in blue) is where  $T_1 > T_2$  and  $\mu_1 > \mu_2$ . The acute-angled quadrant (yellow) to the upper right is where  $T_1 > T_2$  and  $\mu_1 < \mu_2$ . The lower right quadrant (green) is where  $T_1 < T_2$  and  $\mu_1 < \mu_2$ . The lower left quadrant (red) is where  $T_1 < T_2$  and  $\mu_1 > \mu_2$ .

maximum total entropy, the previously considered case of MB trade takes place in the cone of directions where  $\mu_2 dG_1 < -dM_1 < \mu_1 dG_1$ . Following the second law, this always points to higher entropy. Investment takes place in the direction with  $(T_2 - T_1)dM_1 > 0$  and  $dG_1 = 0$ . This, too, always points to higher entropy. Taking convex combinations of the two we obtain a larger cone of possible directions, depending on the quadrant. But the space of all directions that don’t decrease total entropy is a half-space, which raises the question of how to achieve the remaining entropy-increasing directions. One way to achieve them is by allowing connection to a third economy 3 under the control of an external trader T. T uses 3 to create a Carnot cycle to extract money from the temperature difference between 1 and 2, or in reverse to move money against a temperature difference.

To see how this works in more detail, let us pick a particular set-up, for example, where  $T_1 > T_2$  and  $\mu_1 > \mu_2$  (this is the top left quadrant in Figure 7; the other quadrants can be dealt with similarly). Economy 1 can invest money  $m$  in 3, which gets divided into  $em$  for 2 and an initial profit  $(1 - e)m$  for our trader, T. The value  $e \in (T_2/T_1, 1)$ , which is the complement of the efficiency (i.e. the efficiency is  $1 - e$ ), can be arbitrarily close to  $T_2/T_1$  as we approach an ideally efficient Carnot cycle, but cannot exceed this value. T then uses the profit,  $(1 - e)m$ , to buy  $g$  goods from 1 at a price  $\tilde{\mu}_1 > \mu_1$  and sell them to 2 at a price  $\tilde{\mu}_2 < \mu_2$ . Of course, the trader, T, will lose money in doing so, specifically  $(\tilde{\mu}_1 - \tilde{\mu}_2)g$ . The trader may wish to break even, thus balancing the initial profit  $(1 - e)m$  against this loss  $(\tilde{\mu}_1 - \tilde{\mu}_2)g$ , i.e., by choosing  $m$  and  $g$  so that  $\frac{m}{g} = \frac{\tilde{\mu}_1 - \mu_2}{1 - e}$ . In bilateral trade, of course, money and goods flow in opposite directions. But by the trader using a third economy, it has managed to arrange a nett flow of money and goods in the same direction. Here there are both a nett money flow  $\tilde{m} = \frac{e\tilde{\mu}_1 - \mu_2}{1 - e}g$  from 1 to 2 and also a nett goods flow  $g$  from 1 to 2. By choice of  $e, \tilde{\mu}_1, \tilde{\mu}_2$  in the indicated ranges, the ratio  $\tilde{m}/g$  can be made anything greater than  $\frac{T_2\mu_1 - T_1\mu_2}{T_1 - T_2}$ . The latter can be written as  $\frac{\nu_1 - \nu_2}{\beta_2 - \beta_1}$ , which is the

slope of the curve of constant total entropy. So the trader can achieve  $dG_1 < 0$ , i.e., moving goods from 1 to 2, with any slope steeper than the isentrope.

Similarly, still in the case  $T_1 > T_2$ ,  $\mu_1 > \mu_2$ , changes with  $dG_1 > 0$  and  $dM_1 > -\mu_2 dG_1$  can be made, up to the slope of the isentrope. T buys goods  $g$  from 2 at price  $\tilde{\mu}_2$  above  $\mu_2$  and sells them to 1 at price  $\tilde{\mu}_1$  below  $\mu_1$ , but with  $\tilde{\mu}_1 > \tilde{\mu}_2$ . The profit  $m = (\tilde{\mu}_1 - \tilde{\mu}_2)g$  is used to drive money flow from 2 to 1 against the temperature difference, causing money to flow into 1 of up to  $m/(1 - T_2/T_1)$ . The nett money flow into 1 is then given by subtracting  $\tilde{\mu}_1 g$  that 1 paid the trader, T, for the goods. The result of this process is to increase the temperature difference between 1 and 2—analogue to refrigeration in the context of physics, where the temperature difference between the inside and outside of the fridge is increased by expending energy. Here, the expenditure of money is required to increase the economic temperature difference between two economies. Choice of the efficiency of this process of “refrigeration” and the prices  $\tilde{\mu}_j$  yields  $dM_1/dG_1$  anywhere between  $-\mu_2$  and the slope of the isentrope.

Let us consider entropy as a measure of the aggregate welfare of an economy, and consider a “benevolent social planner” deciding whether trade should be allowed between two economies. Of course, if the trade were to reduce the summed welfare between two economies, the social planner would disallow it—but we know that, in the present set-up, this cannot happen: trade cannot decrease overall entropy, by the second law. The social planner might, instead, allow only trades which benefit both economies (where the benefit for one may be 0)—this is, of course, MB trade. But this may seem too restrictive. There may be trades that lead to large welfare benefits for one economy, B, and only small losses, for the other economy, L. How can we make this precise? The thermoeconomic framework has a simple answer: the social planner may allow trades for which the curve of constant total entropy through the final state intersects the region where both economies gain entropy. This is because such a final state could subsequently be moved along the isentrope (arbitrarily closely) to end up in the MB region. Such trading outcomes are what we might call potentially-MB trades.<sup>37</sup> An equivalent way to put the condition is that a trade is potentially-MB if and only if the final entropy is not larger than the maximum entropy over the MB region.

Note that such compensation is not always possible: i.e., not all trades which increase overall entropy are potentially-MB. Figure 6(b) provides an illustration. Suppose that the initial state of the two economies (before trading is allowed) at the top left intersection of the orange and green curves moves to the maximum entropy point or close to it, then no entropy-preserving change can move it to the other side of the orange curve. Thus the entropy of economy 1 remains below its initial value as we move along the isentrope, and hence this trade is not potentially-MB.

Does MB trade have an analogue in physics? It is usual in thermodynamics to consider increase of total entropy as the sole principle. Yet consider a cylinder with an insulated movable partition between two gases (the famous “piston problem” [Wr]). If the pressures  $P_j$  on the two sides are different then the the partition will move, thereby transferring volume from one side to the other, and also the pressure difference will do work  $((P_1 - P_2) dV$  in the quasistatic case). But there is a continuum of choices for where the work goes. We could suppose the partition is attached to a dry friction device (that produces resistive force independent of the speed of movement) located in the left-hand side, whose friction force is controlled to be just less than the force on the piston resulting from the pressure difference; then the work goes to heat the left-hand side. Or a dry friction device on the right-hand side, or a combination of both. These situations correspond to MB trade.

In contrast, one could just let the partition fly and after some oscillations it would settle down to some division of the volume and the energy between the two parts. In this case there is no reason that the individual entropies should both increase. There is, for example, the maximum entropy state where in addition to the pressures equalising, the temperatures also equalise. Even with a perfectly insulated piston, this will be achieved over a long timescale via fluctuations in the position of the partition due to finite numbers of molecules on each side (this effect is outside the domain of classical thermodynamics, but can be expected to apply to economics too). In the case of a small or zero pressure difference between the two sides, the entropy of the initially hotter side decreases.

The economic analogue of the latter case is a free-for-all flow of goods and money with no restriction to exchange at prices between the market prices and allowing money, or goods, to flow simply because they are more valuable in another economy. The only constraint is that the total entropy can not decrease. For infinitesimal transfers  $dM_{AB}, dG_{AB}$  this constraint reads as

$$(\beta_B - \beta_A) dM_{AB} + (\nu_B - \nu_A) dG_{AB} \geq 0,$$

from (5), but it is possible that the entropy of one economy would decrease in the process.

In particular, we note that financial contact between economies does not lead to gains in entropy for both economies, only for the one receiving money, as we saw in Section 12. To take a concrete example, suppose two closed economies are identical, except regarding the amount of money available in each (say, one has ten times the quantity of  $\odot$  than the other). This means that prices for all other goods will be higher in the economy with more  $\odot$ . Indeed, if we suppose that  $\odot$  have no intrinsic use, aside from as a means of trade (unlike gold, which can be crafted in jewelry, for example), then it is reasonable to assume that the prices in one economy will be ten times higher than in the other, following Section 15). Then, if unlimited financial contact (but no trade) is

<sup>37</sup>Potentially-MB trades may remind the reader of the Kaldor-Hicks criterion in welfare economics. Here, though, compensation is assessed in terms of transfers of utility—corresponding to reversible trades along an isentrope—rather than transfers of money.

allowed between the economies, money will rapidly (i.e., not reversibly) flow from the economy with high prices (the ‘low temperature’ economy) to the economy with low prices (the ‘high temperature’ economy), and the prices of goods will fall in first economy and rise in the second. This corresponds to an increase in entropy in the first economy, and a decrease in entropy in the second.

### 19.3 Gains of trade

One of the most fundamental questions in economics concerns the maximum potential gains from trade that can be achieved by allowing trade between economies that were not previously in contact.

The present approach allows maximum gains from trade to be studied in a particularly simple way. We propose that rather than trading directly, all trade is carried out via an intermediary—the now-familiar trader—who can exchange goods for money with either economy. Trade of a bundle of goods  $(dG_i)_{i \in I}$  from  $A$  to  $B$  and suitable amounts of money can be facilitated by the trader if  $\sum_{i \in I} (\mu_i^B - \mu_i^A) dG_i > 0$ , this being the maximum amount of money that the trader siphons off, because the minimum money that  $A$  is willing to receive is  $\sum_i \mu_i^A dG_i$  and the maximum that  $B$  is willing to pay is  $\sum_i \mu_i^B dG_i$ .

The trader can then simply donate the money that it has siphoned off to the two economies (in any proportions), thus increasing the economic entropy of both to some extent. In this way, the gains of trade are split between the two economies. This will also allow us to analyze the maximum increase in economic entropy that can be achieved - which will be reached by donating the gains of trade to whichever economy is economically cooler, because by definition the marginal aggregate utility of money will be greater in that economy.<sup>38</sup> This will continue until the two economies are the same temperature, after which the gains should be distributed between them to keep them at equal temperatures.

We can simplify the problem of finding the maximum profit that the trader can extract by noting that this must be via purely reversible trades, i.e., which will not change the economic entropy of the compound system comprising the pair of economies (although it may shift economic entropy reversibly between one economy and the other, as we note below). This follows because reversible trades will minimize the economic entropy of the compound system. Thus any irreversible trades will lead to the compound system having a higher economic entropy. Thus the trader can guide the system to any such state by first engaging in irreversible trades and then raising the economic entropy of the compound system by donating some or all of its gains of trade. But in doing so, the profit retained by the trader is thereby reduced. So we can conclude that, in order to retain the gains of trade, the trader should use only reversible trades.

Note that this analysis extends to the case of many economies. If we have three economies,  $A$ ,  $B$  and  $C$ , for example, the maximum amount of money that can be extracted by the trader will be determined in part by the maximum amount that can be extracted by trading reversibly between  $A$  and  $B$ . This will yield a compound system in which the prices and economic temperature are the same for  $A$  and  $B$ —if this were not so, then further trading (or voluntary flows of money from the higher temperature economy to the lower temperature economy) could yield more money for the trader—but by assumption the trader has extracted all money possible from the pairwise trade. This compound system can then be conjoined with  $C$  and the trader can engage in further reversible trades until  $A$ ,  $B$  and  $C$  all have the same prices and temperature.

Interestingly, the maximum amount of money that the trader can extract by bringing  $A$ ,  $B$  and  $C$  into equilibrium is independent of the order or manner in which the trades are carried out. For examples, the trader might directly make reversible trades between  $A$ ,  $B$  and  $C$ . Or the trader might initially allow trade only between  $B$  and  $C$ , and then trade with  $A$ ; or the trader might interleave trades between  $A$  and  $B$ , with trades between  $A$  and  $C$ ; or trade in any other pattern, so long as the trader ultimately fully exploits any trading opportunities. That is, while the trader can bring the economies into equilibrium with different equilibrium prices and temperature, only one set of equilibrium prices and temperature will allow the trader to extract the maximum amount of money. For suppose the contrary, that there are two distinct equilibria that the systems could fall into, which would both yield this maximum. Then we could split the entire economy of  $A$ ,  $B$  and  $C$  in half; extract this maximum amount of money by shifting to the distinct equilibria in each half (thus extracting, supposedly, the maximum that the trader can obtain). But then the trader could make reversible trades exploiting the price and/or temperature differences between the two equilibria, thus extracting more money—contradicting the hypothesis that the equilibria had allowed the trader to extract the greatest possible amount of money. Thus, the maximum amount of money that can be extracted from a system of economies (once all reversible trades have been exploited) corresponds to a single equilibrium of prices and temperatures for the entire system.

Moreover, the path of trades by which the prior states of  $A$ ,  $B$  and  $C$  are shifted to that equilibrium by reversible trades does not affect the amount of money extracted by the trader. Again we can prove this by contradiction. Suppose that the trader could obtain the maximum  $M$  money by following one path of reversible trades, but only  $M' < M$  by some other path of reversible trades. Then there would be a “loop” of reversible trades which would take the initial state of the system to the equilibrium, and back to its original state, as follows: follow the path yielding  $M$  to be trader and then follow the other path in reverse, yielding  $-M'$  to the trader.

---

<sup>38</sup>Donating money to an economy of finite size will raise its economic temperature, because we proved that money capacity is positive.

Thus, by following this loop, the system ends where it started, and the trader has gained  $M - M' > 0$ . But this is impossible, because money has been created and given to the trader “from thin air.”<sup>39</sup>

Thus, we can define “gains of trade” between any set of economies as the maximum amount of money that can be extracted by the trader by effecting trade between these economies. This gain of trade also measures the “benefit” that can be obtained through direct trading between economies, without the trader siphoning off any of the surplus (noting, as above, that any entropy-increasing set of trades can be achieved by the trader by first extracting a profit by making reversible trades, and then distributing that profit to the various economies). Where, for example, there is free-for-all direct trading between economies, the gain of trade generates changes in entropy in the constituent economies (which, in sum, will be non-negative, although one or more economies may lose entropy).

Summarising the above discussion, various natural economic results follow, albeit only in the context of exchange economies. On the “positive” side for the benefits of trade:

1. First, gains of trade, as defined above, are always non-negative for the union of two or more countries, and will almost always be positive. These gains are independent of the order in which trades are carried out.
2. Second, the gains of trade for many countries trading together is at least as large as the gains of trade for any partition of those countries into “trading blocks.”

But on the negative side, none of these benefits need arise from spontaneously free-for-all trade between economies.

1. Free-for-all trade may lead to disbenefits (i.e., loss of entropy) for some economies (even though the set of economies as a whole will gain).
2. Such free-for-all trade can sometimes, but not always, be compensated by a benevolent social planner who can engage in reversible trades (which therefore don’t affect the entropy of the system as a whole), such that trade becomes mutually beneficial.

The notion of gains of trade can be viewed as analogous to a generalization of the notion of “available energy” in physics or “exergy” in engineering: i.e., the amount of work that a physical system can do by bringing it into equilibrium with a large reference environment (see [M20] for a pedagogical introduction). By extracting money from one or more economies, the trader is, analogously, extracting value in a usable form, i.e., which could potentially be used to buy goods from some other party. A difference is that we do not assume interaction with a large environment, just interaction of two or more economies; the same can be done in physics and engineering but one tends there to think of the environment as large.

Another concept in physics is “free energy”. This applies to systems that are maintained at a given temperature (Helmholtz free energy) or temperature and pressure (Gibbs free energy). It quantifies the amount of work that can be done by the system, this time because some other aspects, like chemical reactions, are not yet in equilibrium. This is likely to be relevant in economics too, when we reach the stage of considering manufacturing. We briefly discuss this further below in Section 21.4.

In engineering, exergy is often useful for making calculations (e.g. of the capacity of an energy store), and it is likely that this will be true also in the economic contexts. Note, though, that no new principle is being invoked here, over and above the operation of the second law.

A key issue in this idealized world of exchange economies (and one relevant to real-world economic policy) is the impact of trade barriers on gains of trade. Suppose that the barriers are sufficiently sparse that, for all goods,  $G_i$  and all pairs of economies,  $E_j, E_k$ , there is possible sequence of allowable bilateral trades (possibly via intermediate economies) by which a  $G_i$  in  $E_j$  can be exchanged for money in the  $E_k$  and vice versa. We suppose also that money can flow from any economy to any other (possibly via intermediate economies). This will be sufficient to ensure that the system of economies reaches equilibrium prices and temperature. This will be the same equilibrium as if there were no barriers in place, and the gains of trade are unaffected. We consider a weaker form of trade barrier, in the form of tariffs, below.

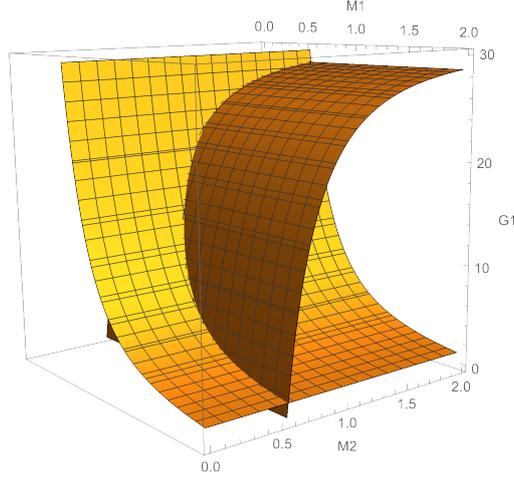
Let us compute the maximum profit that a trader can make from the two Cobb-Douglas economies of Figure 2 without decreasing the entropy of either economy. Maximum profit is made by restricting to no change in either entropy  $S_1, S_2$ . Then this is a curve on the boundary of the accessible region. See Figure 8. The maximum profit is given by the tangency of this curve to a plane  $M_1 + M_2 = \text{cst}$ . A computation with Lagrange multipliers (allowing different exponents  $\alpha_i$  and  $\eta_i$ ) yields the tangency at the state with

$$G_1 = \frac{\alpha_1 M_1 / \eta_1}{\alpha_1 M_1 / \eta_1 + \alpha_2 M_2 / \eta_2} G, \quad G_1^{\alpha_1} M_1^{\eta_1} = e^{S_1 / N_1}, \quad \text{and} \quad (G - G_1)^{\alpha_1} M_2^{\eta_2} = e^{S_2 / N_2}.$$

This does not have an explicit solution  $(M_1, M_2)$ , but the trader’s profit is given by the difference between the initial and final  $M_1 + M_2$ .

---

<sup>39</sup>Comparing with physics, this is analogous to finding a way of generating energy “for free,” thus violating the conservation of energy.



**Figure 8:** The intersection of surfaces of constant entropy for the two economies. It lies in the surface of constant total entropy shown in Figure 2.

In the case that economy  $A$  is large compared to  $B$ , the maximum profit (not respecting non-decrease of entropy in either economy) that can be made is, as noted above, given by the ‘exergy’ of  $B$  relative to the temperature and value for  $A$ :

$$W = M_B - M_B^e - T_A(S_B - S_B^e) + \mu_A(G_B - G_B^e),$$

where superscript  $e$  denote the values if  $T_B = T_A$  and  $\mu_B = \mu_A$ . But again, the entropy of one of  $A$  and  $B$  will decrease, because the total entropy remains constant for maximum extraction of money.

For two economies of general sizes, the trader makes maximum profit by conserving total entropy while extracting money to bring the pair to a state of equal temperatures and prices, where they can be merged. As an example, for two Cobb-Douglas economies with the same exponents  $\alpha$  and  $\eta$ , the maximum money that can be extracted is given by minimising  $M_1 + M_2$  over the surface  $S(M_1, M_2, G) = S^0$  shown in Figure 2, where  $S^0$  is the initial entropy. This is given by the point of contact of the tangent plane  $M_1 + M_2 = \text{cst}$  with the surface, equivalently by the conditions  $\beta_1 = \beta_2$  and  $\nu_1 = \nu_2$ . We can solve this even if the two economies have different  $\alpha_i, \eta_i$ , as follows.  $G_i = p_i G$ , where  $p_i = \frac{\alpha_i N_i}{\alpha_1 N_1 + \alpha_2 N_2}$ .  $M_i = \eta_i N_i T$ , where  $T$  denotes the final temperature, to be determined by making  $S = S^0$ . Denoting  $q_i = \eta_i N_i$  and initial values by superscript  $0$ , we obtain

$$(q_1 + q_2) \log T = \sum_{i=1,2} \alpha_i N_i \log \frac{G_i^0}{p_i G} + q_i \log \frac{M_i^0}{q_i}.$$

Thus

$$T = \left[ \prod_{i=1,2} \left( \frac{G_i^0}{p_i G} \right)^{\alpha_i N_i} \left( \frac{M_i^0}{q_i} \right)^{q_i} \right]^{\frac{1}{q_1 + q_2}}.$$

The final amount of money in the system is

$$M_f = (q_1 + q_2) T = \left[ \prod_{i=1,2} \left( \frac{G_i^0}{p_i G} \right)^{\alpha_i N_i} \left( \frac{M_i^0}{q_i \bar{M}} \right)^{q_i} \right]^{\frac{1}{q_1 + q_2}} M^0,$$

where  $M^0 = M_1^0 + M_2^0$  and  $\bar{M} = M^0 / (q_1 + q_2)$ . This is at most  $M^0$  and the difference is the trader’s profit.

## 19.4 Tariffs

The possibility that trade may not be mutually beneficial, and can reduce the economic entropy of a participating economy, raises the question of how governments might seek to control such trade, e.g., using tariffs or subsidies.<sup>40</sup> We introduce an unusual form of tariff to make the analysis more straightforward, which we call the reversible tariff. If an economy  $A$  imposes a tariff  $\tau_i^A$  per unit of good  $G_i$  arriving in  $A$ , it also gives a rebate of  $\tau_i^A$  per unit of good  $i$  leaving  $A$ . Similarly, economy  $B$  will apply a reversible tariff of  $\tau_i^B$  per unit of good  $G_i$ . This means that the amount of revenue collected depends only on the change of distribution of goods before and after trade occurs. The tariffs may be negative, i.e. subsidies for import.

<sup>40</sup>Of course, the government might also impose tariffs or quotas in response to special interests or simply to raise revenue

If trade goes to maximum entropy (i.e. with money flows not necessarily related to purchases), then the equilibrium after trade is that the value of one  $\ominus$  will be the same in both economies and prices *including the tariff* will be equal in both economies, so that for each good  $i$ :  $\mu_i^A + \tau_i^A = \mu_i^B + \tau_i^B$ .<sup>41</sup>

The revenue raised (or subsidy paid, if negative) by each government depends on the volumes  $\Delta G_i$  of goods of type  $i$  that have moved from economy  $A$  to  $B$ , and the imposed tariffs. Explicitly, the revenues  $R_A, R_B$  raised by the governments of  $A, B$ , respectively, are

$$R_A = - \sum_i \Delta G_i \tau_i^A, \quad R_B = \sum_i \Delta G_i \tau_i^B.$$

One could ask what is the optimal set of tariffs for a government. One objective is for the government to maximise the entropy of its own economy, assuming that the tariffs imposed by other governments are either absent or have some fixed value. Such tariffs may, of course, be to the detriment of other economies with which trade is being conducted. It is also interesting to consider the strategic setting in which multiple governments each set tariffs to maximize the entropy of their own economies, in the light of the tariffs of the others—which is likely to require a game-theoretic analysis. A further interesting variation is where the government aims to maximize the entropy of its economy, while raising some required amount of revenue. The latter question relates, of course, to the problem of optimal taxation, where a government wants to raise some amount tax while minimising the reduction in entropy of its economy.<sup>42</sup> Such questions have, of course, been subject to extensive analysis using standard economics (e.g., [Ram]), and it will be interesting to explore them in the present framework. We leave these topics for future study.

## 19.5 Trade frictions

In reality, trade involves irreversible costs, for example, for transport of goods or money-transfer fees. The corresponding money goes to agents in one or both economies or to an external trader. This implies that MB trade requires a sufficient gap between the prices to cover the costs. It also implies a reduction in the amount of money that an external trader can make from facilitating trade.

Irreversible effects also occur when large trades are made. For example, the prices an external trader proposes to two economies for a large quantity of goods have to be different from the market prices, which are for infinitesimal amounts of good. By convexity, the trader makes at most as much money as by slow continuous trading, and typically strictly less money.

## 20 Relations between partial derivatives: analogues of principles of Le Chatelier, Hotelling, Slutsky and more

A central theme of Samuelson’s Nobel Lecture [S72] is the power of explanation based on maximization in economics. Contrasting his viewpoint with what he considers vague appeals to optimization or rationality in other social sciences, he highlights that it has been possible to find testable, quantifiable relationships from the assumption that economic agents are maximizers. His examples focus on individual economic agents as maximizers.

In the light of the second law of thermoeconomics, the present analysis allows all of those relations (and more) to be derived at the level of whole economies, where what is maximized is entropy/aggregate utility. We outline these results here, with a variety of natural extensions. As we noted in Section 2, Samuelson was sceptical of any analogue of entropy to be maximized by the economy as a whole; but it turns out that the maximization analysis that Samuelson provided at the level of individual agents can be extended in this way. Specifically, here we explore the consequences of existence and concavity of aggregate utility  $S$  on partial derivatives. In economics the partial derivatives most commonly studied are “elasticities”: change in demand for a change in price, but we consider it more appropriate to treat the inverse matrix, change in price for a change in amounts, known as “flexibilities” (and are not alone in this [Hou]).

First, we show that the concavity of the entropy (aggregate utility) function leads to non-positivity of many partial derivatives, and signs for other suitable combinations of them. These are analogous to Le Chatelier’s principle in physics, that the response to an external change is to oppose the change. They are also analogous to results obtained by Samuelson for the effects of prices on demands, but our results are for an economy in aggregate, rather than a single agent.

Second, the symmetry of second partial derivatives leads to relations between various first partial derivatives. In thermodynamics, these are known as Maxwell relations. In economics, similar symmetries of cross-elasticities

<sup>41</sup>To prove this, write the changes in money in  $A$  and  $B$  as  $dM_A = -\tau_A dG_{AB} - dM_{AB}$  and  $dM_B = \tau_B dG_{AB} + dM_{AB}$ , where  $dG_{AB}$  and  $dM_{AB}$  are the amounts of goods and money moved from  $A$  to  $B$  (with sign), and for notational simplicity we suppose a single type of good. Then  $dS = \beta_A(-\tau_A dG_{AB} - dM_{AB}) - \nu_A dG_{AB} + \beta_B(\tau_B dG_{AB} + dM_{AB} + \nu_B dG_{AB})$ . Taking the variation with respect to  $M_{AB}$  yields  $\beta_A = \beta_B$ . Taking the variation with respect to  $G_{AB}$  yields  $\beta_A \tau_A - \beta_B \tau_B + \nu_A - \nu_B = 0$ . Using  $\beta_A = \beta_B$  leads to the result.

<sup>42</sup>In our exchange economy, the most natural tax is applied to the stock of goods, and we can ask how to determine the optimal tax rates across the different types of good.

follow from existence of a utility function, which Samuelson called the Hotelling conditions [Hot] (though they appear to go back to Slutsky [S115]). Again, however, the Hotelling conditions apply to individual agents (or firms), whereas our relations apply at the level of the macro-economy.<sup>43</sup>

Third, we derive relations between partial derivatives holding different variables constant. These are analogous to Slutsky's equation relating compensated to uncompensated demand, and to Samuelson's extension of Le Chatelier's principle.

In the following subsections, we treat these three aspects, and make some additional remarks about symplectic structure and areas.

## 20.1 Non-positivity results

We begin with the dependence of values of goods on their quantities. Recall that the values  $\nu_i$  for goods of type  $i$  are defined by  $\nu_i = \frac{\partial S}{\partial G_i}$  (including that for money, which we denote by  $\beta$ ). The value for a good quantifies the marginal utility for one unit of the good. It follows from concavity of  $S$  that

$$\frac{\partial \nu_i}{\partial G_i} \leq 0. \quad (7)$$

It says that the value for a type of good decreases (strictly speaking, does not increase, but we will not always repeat this for simplicity) if the amount of the good is increased. For money, this gives  $\frac{\partial \beta}{\partial M} \leq 0$ , which says the marginal utility of money decreases if the amount of money increases. We have already seen in Section 13 that this can be rewritten as non-negativity of money capacity,  $C = \frac{\partial T}{\partial M} \geq 0$ , because the economic temperature  $T$  is the reciprocal of  $\beta$ .

These results extend to any principal minor of the second derivative of the entropy (namely, minors of odd order are less than or equal to zero, those of even order are greater than or equal to zero). Thus, for example, for a single type of good,

$$\frac{\partial \beta}{\partial M} \frac{\partial \nu}{\partial G} \geq \left( \frac{\partial \beta}{\partial G} \right)^2. \quad (8)$$

The results of the last two paragraphs can be classed as examples of Le Chatelier's principle. Originally formulated for physical systems, it says that in response to an external change, a system adjusts to counter-act the change. Samuelson derived such relations for a profit-making monopolist [S72], considering prices as external and the amount of goods bought as response, and then went on to derive refinements, which we'll mention below. Our relations are for macro-economies and we consider amounts of goods as the external variables and the values as responses (which seems reasonable given that at a macro level prices can respond rapidly but volumes of goods changes only slowly), but if the entropy is strongly concave (i.e. its second derivative is negative-definite) then one can invert the second derivative and the inverse is also negative-definite.

We next extend this result to prices instead of values. To interpret (7) for a good other than money, recall that the values  $\nu_i$  determine prices  $\mu_i$  by  $\mu_i = \nu_i/\beta$ . So for a good other than money the inequality can be written as

$$\mu_i \frac{\partial \beta}{\partial G_i} + \beta \frac{\partial \mu_i}{\partial G_i} \leq 0.$$

It is conventional in economics to think about "demand" as a function of prices, i.e., to ask how the amount of goods possessed by an agent depends on prices. As we are considering a whole economy, where prices are determined endogenously, here it is more natural to consider how prices depend on amounts of goods. Nevertheless, if invertible, one can invert matrices of partial derivatives to deduce how amounts of goods depend on prices.

Extending the consequences of concavity a further, we consider infinitesimal variation of any combination of goods and money. Since  $\nu_i = \frac{\partial S}{\partial G_i}$  ( $i$  including money here), the infinitesimal change  $\delta \nu_i$  is given by  $\delta \nu_i = \sum_j \frac{\partial^2 S}{\partial G_i \partial G_j} \delta G_j$ . From concavity of  $S$  we deduce that

$$\sum_i \delta \nu_i \delta G_i \leq 0. \quad (9)$$

This equation has the units of entropy and it includes a term  $\delta \beta \delta M$  (because, for these purposes, money is just one more good). This is analogous to the formula Samuelson obtained thanks to the lectures by Wilson, mentioned in Section 2 [S72], but again for a whole economy rather than a single agent, and for prices rather than values (he was treating a revenue function rather than our entropy). It has the advantage that it expresses the relation in a way that treats values and amounts of goods symmetrically.

Converting (9) from values to prices for goods other than money by  $\delta \nu_i = \beta \delta \mu_i + \mu_i \delta \beta$ , we obtain

$$\delta \beta (\delta M + \sum_{i \neq M} \mu_i \delta G_i) + \beta \sum_{i \neq M} \delta \mu_i \delta G_i \leq 0.$$

<sup>43</sup>This parallel between Maxwell and Slutsky's results has been noted in the context of individual agents, e.g., [S72].

This can be rewritten as

$$\sum_{i \neq M} \delta \mu_i \delta G_i \leq \frac{\delta T}{T} (\delta M + \sum_{i \neq M} \mu_i \delta G_i).$$

The expression in parentheses is the monetary value of the infinitesimal change. If the changes are made at market prices then it is zero, so in this case we deduce that

$$\sum_{i \neq M} \delta \mu_i \delta G_i \leq 0, \quad (10)$$

just as Samuelson's formula.

We discuss possible significance of equation (10). It is relevant to the question of motion on isentropes of a system, at market prices. In this case, the inequality becomes an equality, because the lefthand side is the second variation of the entropy. In the language of symplectic geometry, we deduce that the isentrope in the space of  $(\mu_i, G_i)$  is a Lagrangian submanifold:  $\sum_i d\mu_i \wedge dG_i = 0$  on tangents to the isentrope. One way to say what this means is that the amount of money one can extract by moving along it is independent of path (which we touched on earlier in our discussion of gains of trade in Section 19.3). This is because  $dM = -\sum \mu_i dG_i$ . This view takes ideas of Samuelson from micro-economics that have been developed by Russell and Cooper [Ru, CR01, CR11], to macro-economics. It may be particularly relevant for a multi-part economy, where moving along an isentrope could correspond to reversible trade or the (infinitesimally) slow flow of money from one economy to another.

## 20.2 Symmetry results

Let us now consider the consequences of the symmetry of second derivatives. For example,  $\beta = \frac{\partial S}{\partial M}|_G$  and (considering just a single other good, for notational simplicity)  $\nu = \frac{\partial S}{\partial G}|_M$ , so

$$\frac{\partial \beta}{\partial G}|_M = \frac{\partial \nu}{\partial M}|_G. \quad (11)$$

Using  $\beta = \frac{1}{T}$ ,  $\mu = \frac{\nu}{\beta}$  and  $C = \frac{\partial M}{\partial T}|_G$ , this can be written as

$$\frac{\partial T}{\partial G}|_M = \frac{\mu}{C} - T \frac{\partial \mu}{\partial M}|_G,$$

which is non-trivial and testable. It relates inflation (the increase in the temperature  $T$ ), on introducing more of a good, to the market price, the money capacity and the change in the market price on introducing more money. An interesting question is whether one can say something about its sign. If a good is desirable ( $\mu > 0$ ) one might expect its price to go up if more money is added to the economy; but this would still not settle the sign of  $\frac{\partial T}{\partial G}|_M$ . It is plausible that if a good is desirable, then when more is added to an economy, its price will go down—which will drag down all prices to stay in equilibrium, so money becomes more valuable, i.e.,  $T$  goes down:  $\frac{\partial T}{\partial G}|_M \leq 0$ ; but it is not clear that this is always right. We think one could make examples with either sign. We already have examples where it is zero (our toy economy or indeed any economy with pure money). The only constraint we see is (8).

Similarly, one can consider the effect on the temperature of changing the amount of a good by purchase at the market price. We denote such a derivative by

$$\frac{d}{dG} = \frac{\partial}{\partial G} - \mu \frac{\partial}{\partial M} \quad (12)$$

(called a ‘‘compensated’’ derivative in economics, and derivative at constant entropy in physics). From  $\frac{d\beta}{dG} = \frac{\partial \beta}{\partial G} - \mu \frac{\partial \beta}{\partial M}$ ,  $\mu = \nu/\beta$ ,  $\beta = 1/T$  and (11), we obtain

$$\frac{dT}{dG} = -T \frac{\partial \mu}{\partial M}.$$

Intuitively again for a desirable good, the righthand side looks negative (if we add more money to an economy but keep the amounts of goods fixed, we might expect prices to increase). We can prove this if money is pure: using  $\mu = \nu/\beta$ , we deduce  $\frac{\partial \mu}{\partial M} = T(\frac{\partial^2 S}{\partial M \partial G} - \mu \frac{\partial \beta}{\partial M})$ . For pure money the mixed second derivative is zero. Concavity of entropy implies that  $\frac{\partial \beta}{\partial M} \leq 0$ . So we would deduce that purchase of new goods is deflationary. On the other hand, it might be that the good is substitutable for money but people prefer money, in which case the price of the good might go down if money increases. Thus we can not in general deduce a sign for  $\frac{dT}{dG}$ . Rather fancifully, we can imagine a prison economy in which the supply of currency,  $Au$ , is drastically restricted, and people trade primarily with, say, matchsticks. If aurums are allowed to flow into the prison, then matchsticks may no longer be used for trading at all, and their value may collapse to be close to zero. In such circumstances, it may be possible that the *price* of matchsticks, measured in  $Au$ , may actually fall when more money is added to the economy.

As another example of consequences of symmetry of second derivatives, we obtain the identity

$$\frac{\partial \mu}{\partial T}|_G = \frac{\partial S}{\partial G}|_T,$$

where the amount of money is allowed to change to accommodate the indicated variations. The proof is to write  $d(TS - M) = S dT + \mu dG$ , and use symmetry of the second derivatives of  $TS - M$ .<sup>44</sup> This identity tells us that the marginal change in price of a good, given a change in financial “temperature” (fixing the amount of the good), is the same as the marginal change in the aggregate utility, given a change in the amount of that good (fixing the value of money). One can extend this to multiple types of good, obtaining symmetry of the derivatives  $\frac{\partial \mu_i}{\partial G_j}|_T$ .

Next we consider the changes in prices resulting from purchases of goods, possibly of a different type. Suppose an additional quantity of goods of type  $j$  is bought by the economy from elsewhere at market price. How much does the market price of type  $i$  of good change? This means we want to know the “flexibilities”

$$\mathcal{M}_{ij} = \frac{d\mu_i}{dG_j} = \left( \frac{\partial}{\partial G_j} - \mu_j \frac{\partial}{\partial M} \right) \mu_i, \quad (13)$$

using the above compensated derivatives. This is a variant of the Slutsky equation, which normally describes the effect on demand of a change in prices, taking into account the concomitant change in money, but here we are considering the effect on prices of a change in amounts of goods. Using  $\mu_i = \nu_i/\beta$ ,  $\nu_i = \frac{\partial S}{\partial G_i}$ ,  $\beta = \frac{\partial S}{\partial M}$ , after some manipulations we obtain

$$\beta \mathcal{M}_{ij} = \frac{\partial^2 S}{\partial G_j \partial G_i} - \frac{\partial}{\partial M} \left( \frac{\nu_i \nu_j}{\beta} \right). \quad (14)$$

We see that this is symmetric in  $i$  and  $j$ .

One may also ask about the sign of the diagonal components of  $\mathcal{M}$ . They are non-positive, because for  $i = j$  the right hand side of (14) can be written (suppressing the indices) as

$$\left[ \begin{array}{cc} 1 & -\mu \end{array} \right] \left[ \begin{array}{cc} \frac{\partial^2 S}{\partial G^2} & \frac{\partial^2 S}{\partial G \partial M} \\ \frac{\partial^2 S}{\partial M \partial G} & \frac{\partial^2 S}{\partial M^2} \end{array} \right] \left[ \begin{array}{c} 1 \\ -\mu \end{array} \right] \leq 0, \quad (15)$$

by concavity of  $S$ . Thus  $\frac{\partial \mu}{\partial G}|_S$  is  $T$  times this and so is non-positive.

Furthermore, we deduce that the flexibility matrix  $\mathcal{M}$  is negative semi-definite, because for any vector  $x$  with coordinates corresponding to types of goods,

$$\beta x^T \mathcal{M} x = \left[ \begin{array}{cc} x^T & -x \cdot \mu \end{array} \right] [D^2 S] \left[ \begin{array}{c} x \\ -x \cdot \mu \end{array} \right] \leq 0, \quad (16)$$

where  $D^2 S$  is the matrix of second derivatives including with respect to money as the last coordinate.

These observations are reminiscent of some standard results in economics. Firstly, for a profit-maximiser who chooses prices as a function of demands for goods, the flexibilities are shown to form a symmetric negative semi-definite matrix, just as we have obtained here for how prices respond to amounts of goods in a macro-economy. Secondly, for a utility-maximising agent, where prices are considered as independent variables and demands as responses, the elasticities (derivatives of demands with respect to prices) are shown to form a symmetric negative semi-definite matrix. We will derive the analogue of the second result for a macro-economy next under a non-degeneracy condition.

If  $S$  is strongly concave ( $D^2 S$  invertible) then for  $x \neq 0$  we obtain strict inequality in (16). Furthermore, we obtain that  $\mathcal{M}$  is invertible and its inverse  $\mathcal{M}^{-1}$  is negative-definite.  $\mathcal{M}^{-1}$  describes how amounts of goods in an economy depend on prices. In particular, the cross-elasticities are symmetric and the diagonal components of  $\mathcal{M}^{-1}$  are negative.

The use of the above assumption of strong concavity raises the question of whether there is an economic analogue of the phase transitions that are so important in physics, where a continuous change in parameters of a system leads to an abrupt change in its properties (e.g., the cooling of water leading it to turn into ice). Phase transitions will occur where  $S$  is not strictly concave. If the quantity of goods is in an interval of non-strict concavity then the population may partition into one fraction with average goods per agent at one end of the interval and another fraction with average goods per agent at the other end (by analogy with liquid/vapour coexistence in chemistry). Exploring this type of phenomenon in an economic context is an interesting direction for future work. Note that the other apparent way in which phase transitions might arise, through a jump in derivative of the entropy, is ruled out in the current model by the proof of differentiability of entropy.

<sup>44</sup>Spelling this out in more detail, we know that in general  $d(TS - M) = S dT + T dS + \mu dG$ . But when trading at market prices, entropy will be unchanged, so that  $dS = 0$ . Trading at market prices also implies that the change in money  $dM$  is precisely balanced by the change in the quantity of the good,  $dG$  multiplied by its price  $\mu$ . This gives  $d(TS - M) = S dT + \mu dG$ . The first partial derivatives are therefore:  $\frac{\partial(TS - M)}{\partial G}|_T = \mu$  and  $\frac{\partial(TS - M)}{\partial T}|_G = S$ . Using the symmetry of second derivatives ( $\frac{\partial^2(TS - M)}{\partial T \partial G} = \frac{\partial^2(TS - M)}{\partial G \partial T}$ ), gives the required result that:  $\frac{\partial \mu}{\partial T}|_G = \frac{\partial S}{\partial G}|_T$ .

### 20.3 Partial derivatives under different constraints

Returning to the discussion of results on partial derivatives, we derive two more results analogous to a result of Samuelson's that he called the Le Chatelier-Samuelson principle.

The first result compares the dependence of value for a good on the amount of the good under two conditions: fixed money or fixed temperature. The result is that if  $\frac{\partial \beta}{\partial M} \neq 0$  (else it might not be possible to hold temperature fixed) then

$$\frac{\partial \nu}{\partial G}|_M \leq \frac{\partial \nu}{\partial G}|_T \leq 0.$$

The proof is that fixed temperature requires  $0 = d\beta = \frac{\partial \beta}{\partial G} dG + \frac{\partial \beta}{\partial M} dM$  (where unless indicated, partial derivatives are at constant  $M$  or  $G$  respectively), so  $dM = -\frac{\partial \beta}{\partial G} / \frac{\partial \beta}{\partial M} dG$ . Then

$$d\nu = \frac{\partial \nu}{\partial G} dG + \frac{\partial \nu}{\partial M} dM = \left( \frac{\partial \nu}{\partial G} - \frac{\partial \nu}{\partial M} \frac{\partial \beta}{\partial G} / \frac{\partial \beta}{\partial M} \right) dG.$$

So

$$\frac{\partial \nu}{\partial G}|_T = \frac{\partial \nu}{\partial G} - \frac{\partial \nu}{\partial M} \frac{\partial \beta}{\partial G} / \frac{\partial \beta}{\partial M}.$$

The factors in the numerator of the second term on the right are equal by symmetry and the denominator is negative. Hence  $\frac{\partial \nu}{\partial G}|_T \geq \frac{\partial \nu}{\partial G}|_M$ . But also the righthand side can be recognised as the quotient of a 2-dimensional minor of  $D^2S$  (non-negative) by  $\frac{\partial \beta}{\partial M}$  (negative), hence  $\frac{\partial \nu}{\partial G}|_T \leq 0$ .

For the second result, we consider two or more types of good and take into account changes in money required to purchase or sell goods. As before, we consider the amounts of goods as the independent variables and the prices as dependent. The result is that

$$\frac{d\mu_1}{dG_1|_{G_2}} \leq \frac{d\mu_1}{dG_1|_{\mu_2}} \leq 0. \quad (17)$$

The partial derivatives here are those that take into account compensation by the change in money, as in (13), but holding  $\mu_2$  fixed in the second instead of  $G_2$ . For notational simplicity, we will indicate which variables are held fixed only when they are not just the other goods. We assume that  $\frac{d\mu_2}{dG_2} \neq 0$  (hence negative, by (15)), else it might not be possible to hold  $\mu_2$  fixed. The proof is similar to the first result. We use  $d\mu_i = \sum_j \frac{d\mu_i}{dG_j} dG_j$ , to show that if  $d\mu_2 = 0$  then  $dG_2 = -\frac{d\mu_2}{dG_1} / \frac{d\mu_2}{dG_2} dG_1$  and then

$$\frac{d\mu_1}{dG_1|_{\mu_2}} = \frac{d\mu_1}{dG_1} - \frac{d\mu_1}{dG_2} \frac{d\mu_2}{dG_1} / \frac{d\mu_2}{dG_2}.$$

By symmetry of the partial derivatives, the numerator in the second term is non-negative, and by negativity of the denominator we deduce the first inequality in (17). The second inequality comes by recognising the righthand side as the quotient of a 2-dimensional minor of  $\mathcal{M}$  by  $\frac{d\mu_2}{dG_2}$ .

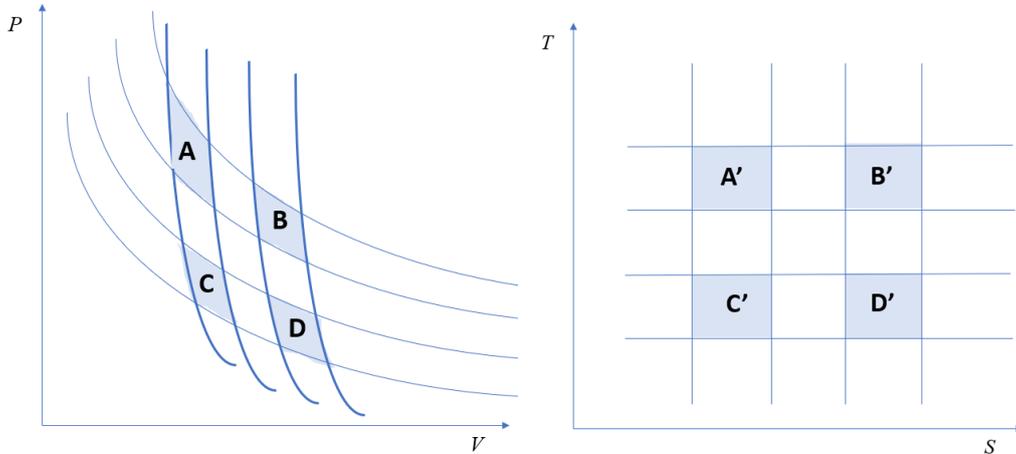
Samuelson (and commentators such as [MR]) frame discussion of his relations in terms of short- and long-term, but they seem to us to depend simply on whether goods or prices are held fixed.

Note that many of these cross-relations simplify if money is pure (in the sense of Section 15). In particular, all the cross-derivatives between money and goods are zero. In more detail, if money is pure, one observation is that  $T = M/K$  and so the money capacity  $C = K$ . Another is that the compensated derivative  $\frac{dT}{dG} = -C\mu$ . A third is that in the Le Chatelier-Samuelson principle,  $\frac{\partial \nu}{\partial G}|_T = \frac{\partial \nu}{\partial G}|_M$ .

There is a further economic consequence that follows nicely in the present thermoeconomic account—perhaps not surprisingly, as Samuelson derived it by analogy with thermodynamic considerations. In his Nobel lecture, he notes “While reading Clerk Maxwell's charming introduction to thermodynamics..., I found that his explanation of the existence of the same absolute temperature scale in every body could be true only if on the  $p - v$  diagram that I earlier referred to in connection with LeChatelier's Principle, the two families of curves—steep and light or less-steep and heavy—formed parallelograms like  $a$ ,  $b$ ,  $c$ ,  $d$  in Figure 2 which everywhere have the property that Area  $a$ /Area  $b$  = Area  $c$ /Area  $d$ . And so it is with the two different economic curves. It is a consequence of the Hotelling integrability conditions...” Samuelson then draws out the economic parallel: that a price change for some good has less influence on the quantities of good held when the quantities of other goods can change (but with fixed prices), than when the quantity of those other goods is fixed.

Samuelson applies his analysis to an individual agent: specifically a “profit maximizing monopolist” who adjusts the inputs required in production. In our account, as usual, we instead focus on the economy as a whole. In particular, we consider Carnot cycles running around the periphery of the areas  $A$ ,  $B$ ,  $C$  and  $D$  which Samuelson identifies. The amount of money that can be extracted by the trader in one turn of the Carnot cycle is represented by the area  $\int PdV$  of the region whose boundary is enclosed by the path. Replotting in  $TS$  coordinates, the Carnot cycle now corresponds to running round the sides of rectangles of a grid. The areas are preserved in the re-plotting. That is, in  $TS$  coordinates the amount of money extracted by the trader by going round the Carnot cycle is represented by the area  $\int TdS$ : i.e.,  $Area_A = Area'_A$ ,  $Area_B = Area'_B$ , etc. In the

$TS$  plot, it is immediately clear that  $Area'_A/Area'_C = Area'_B/Area'_D$ , hence giving Samuelson's result for  $PV$  coordinates:  $Area_A/Area_C = Area_B/Area_D$ . See Figure 9. Note, again, that while Samuelson's result applies only for an individual maximizing agent, the present result holds for the economy as a whole.



**Figure 9:** Relationship between volume (quantity of goods) and price, with changes at a constant temperature (light lines) and entropy (heavy lines). The same graph replotted in temperature-entropy coordinates, so that the changes move the system along the lines of a grid. The amount of money that the trader can extract by reversibly running a Carnot cycle along these areas is  $\int PdV = \int TdS$ —thus the areas  $A, B, C$  and  $D$  are the same as the corresponding areas  $A', B', C'$  and  $D'$ . The equal area rule clearly applies to the areas on the grid in  $TS$  coordinates, and hence in the original  $PV$  coordinates in Samuelson's discussion.

## 21 Further considerations for exchange economies

Here we raise some further aspects of exchange economies that seem natural topics for future work extending the thermoeconomic framework.

### 21.1 More than one currency

We have so far restricted ourselves to considering a single currency, the Aurum  $Au$ . How can the thermoeconomic approach be extended to deal with multiple currencies, either within a single economy or between economies?

There are various cases in which multiple currencies are used in a single economy, e.g. in the long and controversial history of “bimetallism” in Europe, the U.S. and elsewhere, where two forms of money, typically silver and gold, are used in parallel (see [Fay]), and in the recent proliferation of cryptocurrencies, though these are only rarely used for trading.<sup>45</sup> In the thermoeconomic framework, we can simply view the currencies as different forms of good. Then as with any other goods there will be a market exchange rate, generalising the notion of market price, as in Section 14. This is the exchange rate that if posted by an external trader would lead to no nett flow of either currency to or from the trader.<sup>46</sup>

The market exchange rate between two currencies within an economy will change depending on the quantities available—for example, the discovery of large new silver deposits would tend to lower the value of silver currency in relation to gold. The case in which an economy has two forms of pure money is particularly easy to analyze. Then  $S = k_1 \log M_1 + k_2 \log M_2 + F(G)$ , for some constants  $k_1, k_2$  and function  $F$  of amounts  $G$  of all other goods. It follows that the price  $\mu_i$  of a good using currency  $i$  is  $\frac{\nu M_i}{k_i}$ . So if the amounts of the two currencies are fixed, the ratio  $\mu_1/\mu_2$  of prices for a given good is constant  $(\frac{M_1 k_2}{M_2 k_1})$ .

In the second context, where two economies use different currencies but currency flow between economies is allowed, then currency exchange is a special case of trade, in which economy  $A$  can only buy goods in economy  $B$  in two steps: first exchanging currency  $A$  currency for currency  $B$ , and then using currency  $B$  to purchase the good. Thus, the currencies play a key intermediary role in trade. When two or more economies with different currencies are in equilibrium, the prices of goods will be the same in all economies (taking account of any tariffs, as above) up to a scale-factor depending on the exchange rates between those currencies. The market exchange rate between economies will such that rate there is no nett flow of currency between economies. These conclusions

<sup>45</sup>Although bitcoin has been made legal tender in El Salvador: <https://www.bbc.co.uk/news/world-latin-america-57398274>

<sup>46</sup>In practice, of course, bimetallism typically involved the government setting a fixed exchange rate between gold and silver.

follow directly from the analysis of trade outlined above: while we discussed trade above in the context of a single common currency, but that was not really necessary (indeed, the analysis of trade still applies where one allows exchanges of bundles of goods that do not involve any currency). We leave the extension of this type of analysis to more complex cases, including the impact of fixed exchange, currency controls, and so on for future work.

A further direction to explore systematically is the impact of the choice of the goods or goods that are treated as currencies. We noted in Section 9 that entropy does not depend on the choice of numeraire, but other economically relevant quantities, such as temperature and of course prices, will do so. Moreover, if we take “frictions” in exchange seriously (which will make perfectly reversible trades impossible) then it is interesting to ask how (and to what extent) the existence of money reduces these frictions, which may in turn throw light on which numeraires are “good” choices to serve as currency. Another direction for future work is to consider the economic significance of different forms of money, from commodity money such as gold or silver, to paper money and purely digital currencies; and the impact of IOUs, cheques, the existence of credit, and so on. These and related questions appear tractable within a thermoeconomic framework, but the details remain to be worked out.

## 21.2 Spatial inhomogeneity

An important feature of real economic systems is that they are geographically and sociologically inhomogeneous. How does this affect temperature and prices? According to our theory, if in equilibrium then temperature and prices should not depend on position in the economy. But let’s compare with physics. For an atmosphere in a gravitational field, if it is in equilibrium (in particular, with no input or output fluxes of radiation) then indeed the temperature becomes constant. But the pressure  $p$  does not. Instead it is  $p + g\rho h$  that becomes constant (where  $\rho$  is the density,  $g$  the gravitational field and  $h$  the height). So do prices require some analogous correction?

## 21.3 Heat and work

In thermodynamics the distinction between heat and work is usually important. Is there an analogue in economics? It seems that money transfer for goods at market prices should qualify as work and any other money transfer as heat. Concretely,  $dM = T dS + \mu dG$  for quasi-static changes. So sale of goods at market price can be regarded as work  $\mu dG$  done on the system and all other money flow  $T dS$  as heat flow into the system.

Heat flow and work have different consequences. Work is that component of the trade or other interaction that is reversible: it moves the system on an isentrope, and thus leaves the aggregate utility of the system unchanged. Moving the system isentropically will, though, typically involve the system losing or gaining money, which may be captured, or provided, by some external system, such as our trader. By contrast, “heat flow” into a system is not reversible: the system moves to a state with higher entropy, i.e., moving to a new isentrope, with higher aggregate utility.

Thus, in the context of trade between two economies, we can see the gains of trade being split, in some proportions, between the trader (or, more generally, any intermediary or system of intermediaries as a real economy with importers, wholesalers, retailers and so on) (work) and the agents in the economies (heat flow). At one extreme, the trader trades slowly at market prices and captures all the gains of trade; at the other extreme, there is a free-for-all of direct exchange, leading to new equilibrium prices, and the entire gains of trade are assimilated as “heat flow” by the agents in the economy itself (where the amount gained depends purely on the final equilibrium, and not the path by which it is reached).

## 21.4 Free energy

As touched on in Section 19.3, it is common in physical thermodynamics to consider systems in thermal equilibrium at a temperature  $T$ , and then to replace the principle of non-decrease of entropy  $S$  by that of non-increase of free energy  $F = E - TS$ . If a network of economies is in financial equilibrium, as one could presume is the case for those with no currency controls, then our theory would give the same temperature (fluxes of inputs and outputs are neglected here). It then becomes natural to consider the global economy to be at a given temperature and individual economies are then more simply described by  $F = M - TS$ , considered as a function of  $T$  and  $G$ . The principle of maximising entropy is replaced by minimising  $F$ . Similarly, if goods were freely traded, one could expect a common price, and it becomes simpler to minimise  $M - TS + \mu G$  as a function of  $T$  and  $\mu$ . From this perspective, an economy whose free energy can be reduced is one that can effect useful changes.

While notions of free energy, exergy and their generalizations may be useful for analyzing special cases in economics, it is likely to be more useful in general just to look at potential increases in entropy. For a single economy at equilibrium, of course, it is already at maximum entropy and no increases are possible. But for a composite economy, one can look at the possible increases, as discussed in Section 19.

## 21.5 Entropy as liberty

John Stuart Mill’s great works of political philosophy, *On Liberty* [MillL] and *Utilitarianism* [MillU], attempt to promote the idea that a free society, where liberty is maximized, will also maximize welfare.<sup>47</sup> But the link between maximizing liberty and welfare has been controversial ever since.

The thermoeconomic framework provides a potentially interesting perspective on the relation between ideas of utility and liberty, by drawing a parallel with the “macro” interpretation of entropy of classical thermodynamics and the “micro” interpretation of entropy provided by statistical mechanics. In statistical mechanics, entropy gains a concrete interpretation as the logarithm of the available volume corresponding to given macro-state. Volume is measured using Liouville measure in phase space [Set] or its quantum analogue (commonly done by counting eigenstates, but more generally by using Fubini-Study metric [BH]). Is there an interpretation of economic entropy as the logarithm of the available volume of a macro-state? This would connect to ideas of maximising liberty: the volume at the “micro-level” associated with a macro-state corresponds, intuitively, to the number of different micro-states are compatible with that macro-state. Thus, higher entropy macro-states impose fewer constraints at the micro-level, leaving economic agents greater “freedom of choice.”

For the toy economy of Section 10, the natural notion of volume is the normalisation constant  $Z$ . To leading order in  $N$ , the logarithm of  $Z$  depends on  $G$  and  $M$  as  $N \log G^\alpha M^n$ , so indeed the entropy corresponds to logarithm of available volume. If a scale for entropy can be chosen so that it is logarithm of available volume then the fluctuation formulae from statistical mechanics carry over (mentioned in Section 16). In the simple context of an exchange economy, maximizing liberty is maximizing the number of ways in which goods and money can be distributed across agents. For example, if we have two economies,  $A$  with only apples and money, and  $B$  the other with bananas and money, and then allow them to trade, then the volume of possible stocks of goods increases—each agent in both  $A$  and  $B$  is now “free” to have non-zero stocks of apples, bananas and money. Thus, allowing trade can be viewed as increasing the liberty of agents, and this remains true independent of the details of the preference or decision-making mechanisms of those agents. Indeed, increasing liberty might be viewed as a virtue of trade, and indeed markets more generally, irrespective of whether agents have coherent preferences (and hence well-defined utility functions at the level of individual agents).

There is a potentially interesting parallel with Sugden’s [Su04] opportunity criterion defence of the market. He argues that the virtue of markets in an exchange economy such as we have described here is that they maximize the possible alternative bundles of goods that an agent can hold in exchange for their current bundle. This is very close to the present perspective. Sugden is considering maximizing the number of potential exchanges for each agent, using deterministic agents in a competitive equilibrium. By contrast, the statistical mechanics viewpoint sketched here considers the volume of states generated by actual exchanges in statistical equilibrium considering the economy as a whole. We leave exploring potential links for future work, but note that the present approach potentially yields an “opportunity criterion” justification not just for individual markets but for any entropy-increasing transition. That is, certain types of economic processes (e.g., trade, financial contact) may be justified on the grounds that they lead to increased “opportunity sets” for consumers, in the terminology of [Su04]. This analysis may also help clarify the alignments, and divergences, between the utilitarian and liberal political objectives that trace back to Mill.<sup>48</sup>

## 21.6 Migration

We have considered each economy as consisting of a fixed number of agents. But suppose we allow them to move between economies. An immediate question is how to define a reversible change, which would thus leave entropy unchanged. A natural approach is for our trader to mediate, as before, but now charging people entry when moving from one country to another (the charge could be negative, incentivising people to move to a country). The trader would extract money from a system of several countries by facilitating “migratory contact” between economies, and extracting the maximum possible amount of money by slowly reducing the charge until there is no nett flow of people. The process could then be reversed by the trader slowly levying the same charges, but with opposite sign. The “gain of trade” extracted by the trader will behave as gains of trade when goods are exchanged, i.e., depending purely on the initial and final state. That is, that the amount of profit the trader can extract will be independent of the order in which migrations occur. Moreover, when there is free movement of people, and thus no charge from the trader, the entropy of the system will increase (not necessarily to the mutual benefit of each economy). There will, of course, be interesting interactions between migrations of people and flows of money and exchanges of goods. So, for example, suppose that the distribution of preferences might differ across economies, such that in economy 1 people mostly prefer apples to bananas but the opposite in the otherwise identical economy 2. Then we should expect a mutually beneficial exchange of apples and bananas between the economies (which could be used to harness a gain of trade, by our trader). But alternatively, if migration is allowed, some people

---

<sup>47</sup>The welfare theorems of conventional economics take a step in this direction, if we assume that competitive equilibria are the result of maximally free choice in trading of goods and a Pareto optimum is at least a local optimum in utility for individual agents (though not necessarily for their summed utilities).

<sup>48</sup>Connections can also be made with “future state maximisation” models of control process underlying animal collective behaviour, as outlined in [CT], according to which each animal acts to maximize the variety of its possible future sensory states.

might leave economy 1 for economy 2 (attracted by the availability of bananas), while people might move from economy 2 to economy 1 (attracted by the availability of apples). Here trade and migration are, to an extent, substitutes. With sufficient migration before trading begins, the difference in tastes between the economies could be eliminated entirely (perhaps generated a large profit for the trader in the process); but if so there will be no further net trade (and no gains of trade for the trader to exploit). Conversely, sufficient trade will equalize the prevalence and prices of apples and bananas in each economy, reducing or eliminating the incentive to migrate. Clearly any mix of trade and migration is possible—and the gains of trade (and, relatedly, the entropy gain for the pair of economies as a whole) will depend only on the final state (the stock of people, money and goods in the economies), and will be independent of the path by which this state is achieved (e.g., migration first, then trade; the opposite; or any other variation).

We leave further exploration of migration to future work, including the crucial question of taking account of the “frictions” involved. Note, though, that it is possible to make progress (e.g., establishing the aggregate utility increase arising from the movement of people, in this idealized setting) without needing to make microfoundational assumptions about the distribution of preferences, or decision-making procedures, of individual agents.

## 21.7 Non-extensivity

We made explicit in Section 8 that the present analysis assumes scaling symmetry—i.e., that economies can be divided into identical copies with the same properties apart from scale. It seems unlikely that this is a good approximation. For example, as we noted earlier, economies will contain organizations at many scales, and some industries will not be viable at a small scale. Nonetheless, there are robust power-law scaling laws throughout many aspects of the economy [Gab], so it is possible that a more general notion of entropy, such as Tsallis entropies, can be applied [Ts]. Recent work in non-extensive statistical mechanics (e.g. the special issue [B+]) may provide a useful starting point.

## 21.8 Externalities

A particularly important application of the framework will be to consider externalities, where (in our simple exchange economy) the goods held by an agent affect not only that agent’s welfare, but indirectly affect the welfare of others (either positively or negatively). So, for example, we might imagine an economy of leisure activities around a small lake, where people can buy and sell fishing rods among other things. But buying (and, presumably using) a fishing rod will reduce the utility of fishing for others, by reducing the number of fish for them to catch. The trader can slowly and hence reversibly sell fishing rods to the lakeside economy until the price of a fishing rod is near zero<sup>49</sup>. At this point there are almost no fish and almost no utility to be gained from owning a fishing rod.

Thus, when fishing rods are few, a subset of people enjoy fishing (so that the summed utility for fishers may be large); but when the number of fishing rods is high, the summed utility for fishers may be much lower. Thus, the existence of externalities breaks the relationship between entropy and “utility” in any sense relevant to the welfare of individual agents. This is analogous to the breakdown of the welfare theorems in conventional microeconomics. The connection between the equilibrium outcomes of a competitive market and Pareto optimality requires that there are no externalities, and does not hold otherwise.

Following conventional economic analysis, we can address the problem of externalities in our lakeside economy by, for example, having a market for a fixed number of fishing permits (a “cap and trade” scheme), or simply fixing the number of fishing rods, so that voluntary exchanges involving fishing rods no longer harms other agents. Notice that the present approach still provides a potential route to analyse whether one set of rules in an economy is “better” than another, without having to measure and sum the utilities of individual agents. Specifically, we can make clones of our lakeside economy, each governed by different rules; and then put these in contact and ask about the direction of the flow money when the clones are put in financial contact<sup>50</sup>. A natural assumption is that money will flow to the clone with the “best” set of rules—which imply that an economy with such a set of rules would have a lower temperature (and thus that the marginal utility of money will be greater). We leave this and related cases (e.g., dealing with positive externalities) for future work.

We see the fact that the thermoeconomic framework applies straightforwardly in the context of externalities as a significant strength: economies will be governed by the second law whether there are externalities or not. All that changes is the *interpretation* of entropy as corresponding to aggregate welfare. Indeed, by operating at the macro-level, the thermoeconomic approach, and the second law in particular, applies irrespective of the micro-foundations concerning agents and their interactions, so long as the macro-level axioms hold. So, for example, there are no requirements that agents make decisions based on consulting a utility function of any form, that their preferences are transitive or complete, and so on. Moreover, an agent’s utility (if it has one) and behaviour

<sup>49</sup>Reversibility assumes, perhaps implausibly, that fish stocks would fully recover as the number of fishing rods reduces—though the assumption is perhaps less implausible if we recall that reversible trades are required to be arbitrarily slow

<sup>50</sup>or indeed migration of people, if we include this in our analysis.

may depend in complex ways on the utilities and behaviours of other agents [FF]<sup>51</sup> or on comparisons with the agent’s own prior or current experience [VI]. Different microfoundations will, of course, change the properties of the entropy function, just as in physics where different microscopic properties of a gas (for example, whether it approximates an ideal gas in which molecules do not interact or whether the gas molecules are subject to van der Waals forces) determine the entropy function. Crucially, though, the macro-level analysis can proceed and macroscopic quantities such as prices, temperature and entropy can be measured, even when the microfoundations are not known or are too complex to model.

## 22 Beyond exchange economies

We have developed our theory in the context of exchange economies. But this leaves out many features of real economic systems. Here we sketch some possible future developments. We are optimistic that prior applications of thermodynamics, which has been extended to deal with many complex processes ranging across physics, chemistry, atmospheric sciences, physiology, and ecology, may provide powerful tools and analogues in providing such extensions.

A key feature of real economies is that some goods are produced and some are consumed. So, for example, crops grow, and food is consumed. Also, money is printed by central banks. Thus it will be important to extend thermoeconomics allow fluxes of goods and money into and out of a system. This appears to pose no in-principle difficulties—indeed, it is a standard situation in physical thermodynamics, called non-equilibrium thermodynamics. For small deviations from equilibrium it still produces a steady state, which to leading order minimises the rate of entropy production, but for larger deviations it can produce time-dependent states, such as cycles.

Another key feature of real economies is that goods are processed and combined to make other goods. To include manufacture is a relatively simple extension from exchange economies. A combination of goods is put together in some way to produce a new sort of good. Manufacturing bears strong resemblance to chemical reactions, so we might expect to be able to carry over many results from chemical thermodynamics, which studies, among other things, which reactions will spontaneously occur, and which cannot, because they would lead to decreasing entropy. Chemical thermodynamics can treat cascade reactions, where the products of one reaction then drive another; and competition for inputs between different reactions. These all seem to have natural analogues when considering networks of manufacturing processes which may compete with, and feed into, one another.

Chemical reactions are typically assumed to be reversible, at least in principle. For example, the energy released when chemical bonds are formed is assumed precisely to balance the amount of energy required to break those bonds—so that both matter and energy are perfectly conserved. This might appear to contrast with the economic case, where it “costs” money both to assemble a car from its component parts *and* to disassemble it back into its component parts. But the contrast may be more real than apparent: energy is required (in the form of heat) to bake a cake from flour, eggs, butter and sugar, but surely a great deal more energy is needed to go through the extremely complex process of attempting to reconstitute the flour, eggs, butter and sugar from the cake. Un-baking a cake no more releases energy than dismantling a car releases money.

We see a close analogy between businesses and the technologies they deploy in economics and enzymes in biochemistry. By introducing innovative processes for converting (less valuable) inputs to (more valuable) outputs, businesses lower the cost of surmounting the barrier to making this transformation; this seems analogous to the operation of enzymes that lower the energetic barrier (“activation energy”) required for a chemical reaction. Biological systems involve cascades of enzyme-driven reactions of staggering complexity; and the enzymes are, of course, themselves constructed through similarly complex processes. The emergence of complexity in biochemical and economic systems may have interesting parallel aspects, which may potentially shed light on both.

Another direction to develop is theory for fluxes as a result of value differences. As we have seen, on putting economies into contact, differences in values (including coolness) lead in general to flows between them. The flows will depend on the nature of the contact. For small differences in value one can postulate a linear relation between the value differences  $\delta\nu$  and the flows  $\delta F$ . In general this is given by a matrix  $L$ :  $\delta F = L \delta\nu$ . The matrix can have off-diagonal terms, e.g. a difference in coolness could drive a flow of a good different from money. This produces entropy at rate  $\sigma = \sum_j \delta F_j \delta\nu_j = \delta\nu^T L \delta\nu$ . In the physical context, an important observation by Onsager is that because the total entropy can not decrease under the flows, the matrix  $L$  is symmetric and non-negative. The same applies in economics. The significance of this result for economics needs exploring. For example, suppose two types of good are more useful together than not, and one type can move between economies, the other not. We suppose money can move between them also. Then if all moves that increase total entropy are permitted, the mobile good will move to make the values for it equal, and money will move to make the temperatures equal. Thus the prices for the mobile good will equalise. Do we get an economic analogue of osmotic pressure? One can

---

<sup>51</sup>Indeed, we can view the wide variety of social preferences, whether positive or negative, as particular forms of externality, and as generating interdependencies between agents, which will not typically impact the “macro” assumptions that underpin the present approach.

also consider a continuum of economies, each in local thermodynamic equilibrium. Then small gradients in values will produce fluxes.

Two key aspects of real economies which will require building on the thinking in terms of chemical reactions and fluxes are labour and finance. Labour is, of course, fundamentally irreversible: it takes labour to make changes and also to un-make them. Thus, a self-sustaining economic system with labour requires low entropy influxes and high entropy outfluxes of materials, analogous to how the earth-system feeds off turning low entropy sunlight into high entropy infrared radiation. Labour is, of course, also closely connected with the size of the population; a more complete model will need to capture births, deaths and flows of population, as well as the process by which people to enter or exit the labour market.

Finally, we note that questions of the capital, investment, interest rates, central banks and indeed the entire financial system are absent from the exchange economy outlined here. Here again a perspective based on fluxes is likely to be required.

## 23 Conclusion

The fundamental result of our paper is that extensive exchange economies have an entropy function that governs the allowed changes in state on interaction between economies: they are the changes for which the total entropy does not decrease. This is in distinction to the Pareto criterion in standard microeconomics, where the allowed changes are those in which the utilities of neither agent decrease. Utility functions specify preferences but can not in general be added meaningfully. In contrast, entropy can be added, thus allowing meaning to our principle that total entropy can not decrease. One might ask how the entropy of one economy could decrease in an interaction, but an answer is that agents can move money to the other economy where it is of more value to them, which is, of course, a feature of the real world.

One perspective on our work is that it provides a justification for the “representative agent” of traditional economics. In conventional economic analysis, while the objective is to understand aggregate behaviour of the economy, the concept of utility applies primarily to individual agents. This creates a gap between macro and micro analysis that is often fixed by postulating the existence of representative agents, with the assumption that the behaviour of such agents should match the behaviour of the economy as a whole. In the thermoeconomic framework, this rather convoluted line of reasoning is not needed: the whole economy behaves as if it were maximizing a utility function and hence can be viewed as a rational utility-maximizing agent. On the other hand, in interactions with other economies, the set of economies maximises their total utility, rather than each attempting to maximise just its own, as mentioned above. The benefits of moving money to another economy are not represented by the entropy of the originating economy.

Entropy leads to a variety of consequent quantities for an economy, like temperature, money capacity, market prices, and relations between them. In particular, we introduce a notion of temperature for an economy. Traditional economics can be viewed as temperature zero. Thus, if real economies have positive temperature then we have opened up an important new direction in economics. Some consequences are the nett direction of money flow between economies is from hotter to cooler, the possibility of money pumps that can exploit temperature differences or enhance them, and a definition of inflation as increase in temperature.

For a real economy, even an exchange economy, the entropy function will in general be unknown. But just as in physics, one could try to measure it. For example, steam turbines operate on steam, which is far from an ideal gas, but engineers use empirically-compiled steam tables to compute optimal operation. By analogy, we might expect that a new branch of econometrics can be developed to measure temperature, money capacity, entropy and other relevant quantities.

If permitted to exchange some types of good and/or money, nett trade can occur in any way that increases the total entropy. This includes mutually beneficial trade, in which the entropies of all partners increase, but also other trades where one gains and another loses entropy. Entropy allows one to compute the maximum amount of money that a trader could make.

This paper makes only a start on the thermodynamics of economic systems. Much more can be done. In particular, one should include production and consumption, manufacturing and recycling, enterprise, labour, migration, births and deaths, interest rates, multiple currencies, non-extensivity, and consider the possibilities of phase transitions and dynamical behaviour. Many of these directions reflect the fact that real economic systems are not in equilibrium. Even if they might still be steady-state at the macro-level, they are subject to fluxes of inputs and outputs. In the physical context, thermodynamics places limits on the rates at which processes can occur, given the input and output fluxes (e.g. see [Kl] for the Earth system). The same can be expected to apply to economics. Connections to recent developments in non-equilibrium thermodynamics of physical systems (e.g., [FE, Ma, MN]) are also likely to be relevant.

The theoretical and practical significance of thermodynamic ideas, originally developed for understanding the efficiency of steam engines, has had a lasting impact across physics, chemistry, biology and engineering, the ramifications of which are still being explored. This raises the possibility that establishing the relevance of a parallel mathematical structure in an economic domain may, when more substantially worked out, have similarly

far-reaching implications.

We hope that once fully developed, this theory may provide powerful methods for thinking about and quantifying economic systems at the macroscopic level, and may have implications for practical matters such as trade negotiations, management of an economy, and investment.

## Acknowledgements

We are grateful to Daniel Sprague for taking on an initial exploration of the ideas for his MSc project in 2011, to Elliott Lieb and Jakob Yngvason for a discussion in 2014, to our colleagues Peter Hammond, Marcus Miller and Herakles Polemarchakis in the Economics department at Warwick for their comments, and to Shafi Sardar for testing some of our ideas, notably on the Carnot cycle, by simulations in the academic year 2022/3. We gratefully acknowledge the support of the Economic and Social Research Council (ESRC), via the Rebuilding Macroeconomics Network (Grant Ref: ES/R00787X/1, subproject: Why are economies stable?); and through the ESRC Network for Integrated Behavioural Science (grant number ES/P008976/1).

## A Derivation of entropy for the Cobb-Douglas economy

Here we derive the formula (4) for the entropy,

$$S = N \log \bar{g}^\alpha \bar{m}^\eta,$$

for the toy economy of Section 10, from the construction in the proof of Theorem 1 in [LY02], where  $N$  is the number of agents, and  $\bar{g} = G/N$  and  $\bar{m} = M/N$  are the amounts of goods and money per agent.

We have to find which states  $(G, M)$  are reversibly accessible from a subdivision of the agents in the ratio  $\lambda_0 : \lambda_1$  (with  $\lambda_i > 0$  and  $\lambda_0 + \lambda_1 = 1$ ) into two parts with initial states  $(G_i, M_i)$ ,  $i = 0, 1$ . Then  $\lambda_1$  (up to an affine transformation to take into account the choice of initial states) will be the entropy of state  $(G, M)$ .

First we show that the trader can move each subsystem reversibly to any state  $(G, M)$  with the same value of  $G^\alpha M^\eta$ , by offering to buy and sell goods at a price  $\mu$  just above or below

$$\mu_c = \frac{\alpha M}{\eta G}.$$

As in Section 10, the result is nett purchase of goods by the subsystem if  $\mu < \mu_c$ , sale if  $\mu > \mu_c$ . In the case of equality, we have reversible accessibility of the states infinitesimally close along the price line  $\Delta M + \mu_c \Delta G = 0$ . Integrating  $dM + \mu_c dG = 0$  produces  $G^\alpha M^\eta = \text{cst}$  (this curve is therefore isentropic).

In the above process, the trader has to buy or sell goods, but we ask the trader to buy exactly the same quantity of goods from one subsystem as it sells to the other. That way it may gain or lose money but will not change its quantity of goods, so the resulting pair of states is reversibly accessible from the original pair.

We choose the movement by the trader of the two subsystems along their isentropic curves to end with the money per agent in each subsystem becoming the same. This is feasible because the prices are positive and the changes in amounts of goods in the two subsystems have opposite signs, so the changes in amounts of money in the two subsystems have opposite signs.

Having achieved equal money per agent, we now put the two subsystems into financial contact and in contact with the trader. The trader buys goods from one at its market price and sells them to the other at its market price, in the direction to make the goods per agent the same in each, and in addition agents transfer money according to their utility for money, which keeps the money  $\bar{m}$  per agent the same in both. This results in  $d(M_0 + M_1) = -\mu_0 dG_0 - \mu_1 dG_1$  with  $M_i = \lambda_i \bar{m} N$  and  $G_0 + G_1$  constant. Then, using  $\mu_i = \frac{\alpha M_i}{\eta G_i}$ , we obtain

$$\frac{d\bar{m}}{\bar{m}} = -\frac{\alpha}{\eta} \left( \lambda_0 \frac{dG_0}{G_0} + \lambda_1 \frac{dG_1}{G_1} \right).$$

As a consequence,

$$\eta \log \bar{m} + \alpha (\lambda_0 \log \bar{g}_0 + \lambda_1 \log \bar{g}_1)$$

is conserved, where we took the opportunity to write  $G_i = \lambda_i N \bar{g}_i$ .

The process ends when both subsystems have the same goods per agent, because they already have the same money per agent and so can then be merged reversibly. It follows that

$$\log \bar{g}^\alpha \bar{m}^\eta = \lambda_0 \log \bar{g}_0^\alpha \bar{m}_0^\eta + \lambda_1 \log \bar{g}_1^\alpha \bar{m}_1^\eta,$$

where all quantities are per agent.

Thus we deduce that  $X \sim (\lambda_0 X_0, \lambda_1 X_1)$  iff the entropy is as above, up to affine transformation. To respect scaling symmetry it must be linear in the number of agents. Hence, up to addition of a possible but irrelevant constant per agent, we obtain the formula (4).

## References

- [A51] Arrow, KJ, An extension of the basic theorems of classical welfare economics, Proceedings of the second Berkeley symposium on mathematical statistics and probability, 2, 507–533, (University of California Press, 1951)
- [AD] Arrow KJ, Debreu G, Existence of an equilibrium for a competitive economy, *Econometrica*, 22(3) (1954) 265–290.
- [B+] Beck C, Benedek G, Livadiotis G, Rapisarda A, Tirnakli U, Tsallis C, Nonextensive statistical mechanics, super- statistics and beyond: theory and applications in astrophysical and other complex systems, *Eur Phys J Spec Topics* 229 (2020) 707–9.
- [BH] Brody DC, Hughston LP, Geometric quantum mechanics, *J Geom Phys* 38 (2001) 19–53
- [CMIM] Candeal J, Miguel J, Indurain E, Mehta G, Utility and entropy. *Economic Theory* 17 (2001) 233–238.
- [CT] Charlesworth HJ, Turner MS, Intrinsically motivated collective motion, *Proc Nat Acad Sci* 116 (2019) 15362–7.
- [C+] Cockshott WP, Cottrell AF, Michaelson GJ, Wright IP, Yakovenko VM, *Classical econophysics* (Routledge, 2009).
- [CR11] Cooper JB, Russell T, On the mathematics of thermodynamics, arxiv:1102.1540 (2011).
- [CR01] Cooper JB, Russell T, Samuelson PA, Characterizing an area condition associated with minimizing systems. In *Economic Theory, Dynamics and Markets: Essays in Honor of Ryuzo Sato*. 391–403 (Boston, MA: Springer, US, 2001).
- [Co] Corning PA, Thermoeconomics: Beyond the second law. 4(1) (2002) 51-88.
- [CK] Corning PA, Kline SJ, Thermodynamics, information and life revisited, part II: ‘Thermoeconomics’ and ‘Control information’, *Syst Res* 15 (1998) 453–82.
- [Da] Davis HT, *The Theory of Econometrics*. (Principia Press, Bloomington, IN, 1941).
- [D51] Debreu G, The coefficient of resource utilization. *Econometrica* 19(3) (1951) 273–292.
- [Di] Dixit A, Paul Samuelson’s legacy, *Ann Rev Econ* 4 (2012) 1–31.
- [DG] Demirel Y, Gerbaud V, Chapter 5 Thermoeconomics. In Demirel Y, Gerbaud V, *Nonequilibrium Thermodynamics (Fourth Edition)* (2019) 267–294.
- [DGMS] Düring B, Georgiou N, Merino-Aceituno S, Scalas E, Continuum and thermodynamic limits for a simple random exchange model, *Stoch Proc & Appln* 149 (2022) 248–77.
- [DGS] Düring B, Georgiou N, Scalas E, A Stylised Model for Wealth Distribution, in: Y. Aruka, A. Kirman (eds.), *Economic Foundations for Social Complexity Science, Evolutionary Economics and Social Complexity Science* 9 (Springer, 2017).
- [E81] Edgeworth FY, *Mathematical Psychics* (London: Kegan Paul, 1881).
- [ER] Elster J, Roemer JE, (Eds.) *Interpersonal comparisons of well-being*. (Cambridge University Press, 1993).
- [FF] Fehr, E, Fischbacher U, Why social preferences matter: The impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal* 112 (2002) C1–C33.
- [FM] Farjoun E, Machover M, *Laws of chaos: a probabilistic approach to political economy* (Verso, 1983, 2020).
- [FG94] Farmer R E A, Guo J-T. Real business cycles and the animal spirits hypothesis, *Journal of Economic Theory*, 63:1 (1994): 42-72.
- [Fa] Farmer R, *The macroeconomics of self-fulfilling prophecies* (MIT Press, 1993).
- [Fay] Fay CR, Newton and the gold standard, *Camb Hist J* 5:1 (1935) 109–117.
- [Fi] Fisher I, The debt-deflation theory of great depressions, *Econometrica* 1 (1933) 337–57.
- [FE] Freitas JN, Esposito M, Emergent second law for non-equilibrium steady states, *Nat Comm* 13 (2022) 5084.
- [Ga] Gali J, *Monetary policy, inflation and the business cycle* (Princeton Univ Press, 2015).
- [Gab] Gabaix, X, Power laws in economics and finance, *Annual Review of Economics*, 1 (2009): 255-93.
- [GR] Georgescu-Roegen N, *The entropy law and the economic process* (Harvard Univ Press, 1971).
- [Go] Goddard JD, The second law of thermodynamics as variation on a theme of Carathéodory, *Proc Roy Soc A* 477 (2021) 20210425.
- [H96] Hartley JE, Retrospectives: The Origins of the Representative Agent, *Journal of Economic Perspectives*, 10(2) (1996) 169–177.

- [H55] Harsanyi JC, Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, *Journal of Political Economy*, 63 (1955) 309–321
- [Hot] Hotelling H, Edgeworth’s taxation paradox and the nature of demand and supply functions, *J Political Economy* 40 (1932) 577–616.
- [Hou] Houck JP, *J Farm Economics* 47 (1965) 789–92.
- [J] Jaynes ET, How should we use entropy in economics?, 2 Jan 1991.
- [KI] Kleidon A, *Thermodynamic foundations of the earth system* (Cambridge U Press, 2016).
- [K92] Kirman AP, Whom or what does the representative individual represent?, *Journal of Economic Perspectives*, 6(2) (1992) 117–136.
- [Kr] Krugman P, <https://quotefancy.com/paul-krugman-quotes>.
- [Ku] Kümmel R, *The second law of economics* (Springer, 2011)
- [LY98] Lieb EH, Yngvason J, A guide to entropy and the second law of thermodynamics, *Notices Am Math Soc* 45 (1998) 571–81.
- [LY99] Lieb EH, Yngvason J, The physics and mathematics of the second law of thermodynamics, *Phys Rpts* 310 (1999) 1–96.
- [LY02] Lieb EH, Yngvason J, The mathematical structure of the second law of thermodynamics, in: *Current developments in mathematics 2001* (International Press, 2002), 89–130.
- [Li] Liggett TM, *Interacting particle systems* (Springer, 1985).
- [Lis] Lisman J, Econometrics and thermodynamics: a remark on Davis’ theory of budgets. *Econometrica* 17 (1949), 59–62.
- [M11] MacKay RS, Robustness of Markov processes on large networks, *J Difference Eqns & Applns* 17 (2011) 1155–67.
- [M20] MacKay RS, Sustainable energy storage – with hot air, cold air and liquid air, *Contemp Phys* 61 (2020) 1–11.
- [Ma] Maes C, Frenesy: time-symmetric dynamical activity in nonequilibria, *Phys Rpts* 850 (2020) 1-33.
- [M20] Marshall A, *Principles of Economics* (Eighth edition) (London: Macmillan, 1920).
- [MN] Maes C, Netocny K, Minimum entropy production principle from a dynamical fluctuation law, *J Math Phys* 48 (2007) 053306.
- [McK] McKenzie L, On the Existence of General Equilibrium for a Competitive Economy, *Econometrica* 27(1) (1959) 54–71.
- [Man89] Mankiw NG “Real Business Cycles: A New Keynesian Perspective.” *Journal of Economic Perspectives* 3, no. 3 (1989): 79-90.
- [MP] Mehra R, Prescott EC, The Equity Premium: A Puzzle, *Journal of Monetary Economics*, 15 (1985), 145–61.
- [MR] Milgrom P, Roberts J, The LeChatelier principle, *American Economic Review* 86(1) (1996) 173–179.
- [MillL] Mill JS “On Liberty.” (London: John W Parker and Son, 1859).
- [MillU] Mill JS “Utilitarianism.” (London: Parker, Son and Bourn, 1863).
- [MS] Murphy M, Savicki D, Obituary for Myron Tribus 1921-2016. *National Academy of Engineering*. <https://nap.nationalacademies.org/read/25543/chapter/58>
- [N50] Nash JF Jr, Equilibrium points in n-person games, *Proceedings of the National Academy of Sciences*, 36(1) (1950) 48–49.
- [NN] Nicolis G, Nicolis C, *Foundations of complex systems* (World Sci Publ Co, 2007).
- [Pag] Paganelli MP, Adam Smith’s Digression on silver, *Camb J Econ* 46 (2022) 531–44.
- [Ram] Ramsey, FP, A Contribution to the Theory of Taxation, *Economic Journal*, 37 (1927), 47-61.
- [RFP] Raine A, Foster J, Potts J, The new entropy law and the economic process, *Ecological Complexity* 3 (2006) 354–60.
- [R28] Robbins L, The Representative Firm, *Economic Journal*, 38 (1928), 387-404.
- [Ru] Russell T, Symplectic geometry: the natural geometry of economics? *Econ Lett* 112 (2011) 236–8.
- [S50] Samuelson PA, The problem of integrability in utility theory, *Economica* 17 (1950) 355–85.
- [S60] Samuelson, PA, 1960. Structure of a minimum equilibrium system. In: Pfouts, R.W. (Ed.), *Essays in Economics and Econometrics: A Volume in Honor of Harold Hotelling*. (University of North Carolina Press, 1960).

- [S72] Samuelson PA, Maximum principles in economics (Nobel lecture), *The American Economic Review*, 62(3) (1972), 249–262.
- [SZ] Samuelson W, Zeckhauser R, Status quo bias in decision making. *Journal of risk and uncertainty*, 1 (1988), 7–59.
- [Sas] Saslow WM, An economic analogy to thermodynamics, *Am J Phys* 67 (1999) 1239–47.
- [Sc] Schotter A, *Microeconomics: a modern approach* (Cengage Learning, 2008).
- [Sen] Seneta E, *Non-negative Matrices and Markov Chains* (Springer, Berlin, 2006)
- [Set] Sethna J, *Statistical mechanics: entropy, order parameters and complexity*, 2nd ed (OUP, 2021).
- [Sl15] Slutsky E, Sulla teoria del bilancio del consumatore, *Giornale degli Economisti e Rivista di Statistica* 51:1 (1915) 1–26.
- [Sl37] Slutsky, E, The summation of random causes as the source of cyclic processes, *Econometrica*, 5:2 (1937) 105–146.
- [SF] Smith E, Foley DK, Classical thermodynamics and economic general equilibrium theory. *Journal of Economic Dynamics and Control* 32(1) (2008) 7–65.
- [SDa] Sousa T, Domingos T, Equilibrium econophysics: A unified formalism for neoclassical economics and equilibrium thermodynamics. *Physica A: Statistical Mechanics and its Applications* 371(2) (2006) 492–512.
- [SDB] Sousa T, Domingos T, Is neoclassical microeconomics formally valid? An approach based on an analogy with equilibrium thermodynamics. *Ecological Economics* 58(1) (2006) 160–169.
- [St] Starr RM, *General equilibrium theory*, 2nd edition (Cambridge U Press, 2012)
- [Su04] Sugden R, The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. *American Economic Review* 94(4) (2004) 1014–1033.
- [Su] Sugden R, *The community of advantage: A behavioural economist’s defence of the market*. (Oxford University Press, 2018).
- [Ts] Tsallis C, Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics* 52 (1–2) (1988) 479–487.
- [TE] Tribus M, El-Sayeed YM, *Thermoeconomics of energy systems*. US Department of Energy Progress Report DOE/ER/1 0518-1, <https://www.osti.gov/servlets/purl/5377349>
- [V] Varian, HR, The Arbitrage Principle in Financial Economics, *Journal of Economic Perspectives*, 1(2) (1987) 55–72.
- [VI] Vlaev I, Seymour B, Dolan RJ, Chater N, The price of pain and the value of suffering, *Psychological Science* 20 (2009) 309–317.
- [Wa] Walters P, *An introduction to ergodic theory* (Springer, 1982).
- [Wr] Wright P, A simple piston problem in one dimension, *Nonlinearity* 19 (2006) 2365–89.