

“Good Questions Outrank Easy Answers”

Thought experiments in economics

Harro Maas
Centre Walras-Pareto
University of Lausanne
Harro.maas@unil.ch

Many doctrines that are afloat, may be truths, but they are at least *paradoxical* truths; —they may be sense, but at least they are not common-sense.

Richard Whately, *Introductory Lectures on Political Economy*, 1832.

Thought experiments have received quite some attention over the past decades from historians and philosophers of (social) science, with mixed messages. In her recent monograph on models and modeling in economics, Mary Morgan steered away from a discussion of thought experiments altogether, even though some of her earlier papers, invite for such a discussion (Morgan 2012, see especially Morgan 2003).

Scholarship on thought experiments has for the most part considered examples from physics, such as Newton’s buckle, Simon Stevin’s famous inclined plane, or Galileo’s fall experiments.¹ Simon Stevin was so taken with his thought experiment that he ornamented his book on the art of measurement with it, adding the epitaph, “Wonder, is gheen wonder”, that is wonder, no wonder, capturing the element of paradox that sticks to thought experiments to this day. Following Kuhn’s famous article on thought experiments (Kuhn 1977), some argue for the central importance of thought experiments for progress in science, while others put stress on the dependence of thought experiments on other forms of research, material experiments or field work, where in their view the real work of the scientist is done. In this view, thought experiments don’t add much, or at least they

¹ The literature is vast. See for example, Sorensen 1991, 1992, McAllister 1996, Reiss 2002, Kühne 2005, and references therein.

don't deserve the enthusiast endorsement by Sorensen and others, who consider thought experiments indispensable indeed for science. If there is agreement on general characteristics of thought experiments, then it is that they aim at debunking, or rather *testing*, existing beliefs, and do so by means of a conceptual, not a material, intervention, that leads to paradox or straightforward contradiction (the difference between paradox and contradiction is not always clear in discussions on thought experiments, but seems to hinge on the importance of commonly held beliefs that get undermined in the experiment). Whether pursued with formal or informal means does not seem to touch these general characteristics of a thought experiment. Thought experiments may lead to further research, but whether it does or not, does not alter the internal dynamics of a thought experiment.

In a useful critical examination of common arguments in favor of thought experiments, Julien Reiss (2002) wonders why philosophers would be surprised thought experiments contribute *something* to science, as all scientific endeavor involves hard thinking, but the very limited number of thought experiments usually referred to should caution anyone about claiming a special status for them. Just such a special status was argued for by Margaret Schabas (2008) who uses a well-known thought experiment of David Hume (on money supply) to distinguish between common modeling practices of economists and thought experiments that do something else. In her account, while models serve purposes of representation, difficult though this might be to argue for in and by itself, thought experiments serve to debunk commonly held presuppositions about our representations of the world and thus shift perspective. Schabas follows Kuhn in emphasizing the centrality of paradox in thought experiments. In contrast with models, she argues, thought experiments hinge on a “jarring paradox” that destabilize commonly accepted conceptual frames.

For Reiss, that may be so, but is simply not enough to attribute a special status to thought experiments as for him, it is in the subsequent collection of supporting evidence – in materially testing propositions – that the action is located. On their own, thought experiments don't lead to change. In addition, and almost taking this for granted, Reiss considers thought experiments less

“exact” than a deductive argument, a criterion that goes somewhat in the same direction as Schabas’ distinction between (formal) models and thought experiments. Reiss does not explain in what sense thought experiments are less exact than (formal) deductive arguments, and this is an issue I will discuss in what follows.

I will grant Reiss that thought experiments on their own seldom – not never – lead to (scientific) change, because after Quine introduced the notion of a “web of belief”, it has become common currency that no scientific activity stands on its own. Neither do the collection of data, material experiments, field work, statistical testing, seminar presentations or letters to the editor. Saying that thought experiments lean on other scientific activities hardly helps to understand their structure, as Reiss acknowledges, neither why they stubbornly keep attracting so much attention.² To think of thought experiments as less “exact” than formal, deductive models however misses their point, and assumes more rigor for models and less rigor for thought experiments than is actually present.

While Schabas discussed an acknowledged example of a thought experiment in economics, I will investigate the internal dynamics of thought experiments using two examples that have not been recognized as such. The first I take from Robert Malthus’ *Essay on Population*, which received overwhelming attention since its first publication in 1798, however, not as a thought experiment, but as an empirical conjecture that failed to be empirically validated, and yet became referred to, and used, as Malthus’ population law. The second is referred to as Samuelson’s overlapping generations model, which has been thoroughly discussed by Daniel Hausman as an example of economic modeling that scorned the “hard empirical work of the other social scientists” in avoiding any attempt to empirical testing (1992, 119). For Hausman, Samuelson’s model was a stellar example of economic modeling in which mathematical finesse trumped empirical relevance. Yet, as I will argue,

² For example, the philosophy department at Aarhus University recently advertised for a PhD-position on thought experiments as part of its Sapere Aude project on the role of intuitions in science and philosophy. In contrast, the logo of the also recently established society for experimental philosophy chose as its logo an armchair on fire.

Hausman wrongly attributed representational characteristics to Samuelson's model, while Samuelson explicitly developed the model to question existing beliefs within the economics profession about the functioning of markets. This is not denied by Hausman, but he does not give it the full weight it deserves.

My discussion will begin with Samuelson's overlapping generations model that I will develop halfway. I will then turn to Malthus' *Essay on Population*, to return via an interlude on nineteenth century Drummond professor in political economy, Richard Whately and the originally Dutch physician Bernard Mandeville to Samuelson's model to reconsider Hausman's verdict. In both cases, of Malthus and Samuelson, common sense beliefs were questioned by showing them inconsistent. The tension between commonly held beliefs can be played out formally (Samuelson) or informally (Malthus), but in fact such a distinction is misleading, because in both cases the thought experiment destabilizes previously held answers. A thought experiment can be seen as posing a question to which it suggests an answer by way of paradox.

Paul Samuelson's "Exact Consumption-Loan Model"

Paul Samuelson published "An Exact Consumption-Loan Model with or without the Social Contrivance of Money" in *The Journal of Political Economy* in 1958. As said, this article has been thoroughly scrutinized by philosopher of science Daniel Hausman in 1992, and his discussion will later serve as a reference point in my efforts to clarify the difference between thought experiments and the testing of empirical models.

A commonly maintained myth about Samuelson is that all through his distinguished career he paid little attention to empirical research, portraying himself as a theoretician who reveled in purely abstract mathematical models. Well into the 1950s Samuelson was involved in empirical research projects. Shortly after the war Jacob Marschak and Tjalling Koopmans of the Cowles Foundation joint forces with Henry Schultz, then head of the Chicago economics department, exactly because of

Samuelson's skills, interests and dexterity in macro-economic statistical investigations of the sort that were then undertaken at the Cowles Foundation (Maas 2013). It should not be forgotten that Lawrence Klein, who later would receive the Nobel memorial prize for his macro-economic modeling of the United States, was one of Samuelson's first PhD candidates at MIT.

It is important, I think, to be aware of Samuelson's heavy background in empirical statistical research, because one can easily forget Samuelson's empirical interests if one looks at his many publications for which it is hardly imaginable they have any practical application. Samuelson's "Exact Consumption Loan Model" is rightly singled out by Daniel Hausman as a premier candidate of such a model.

As explained by Hausman, and as Samuelson emphasized himself, Samuelson constructs in this article an extreme economy. In Samuelson's model world there are no durable goods and the lifetime of workers is divided into three phases, two in which they work and one in which they are in retirement. Childhood nor gender exist. All goods that are produced endure for one period only and therefore cannot be saved up for the next period. That is, there are no durables that can serve as medium of exchange (money) either. Since there are no durable goods, the only items that can be exchanged are those produced in the same period. This model is referred to nowadays as Samuelson's overlapping generations model (at any given moment, three different generations are alive, at least as long as there is no beginning or end to the chain as a whole). The idea of overlapping generations is often used by present-day economists to model an economy, so it has undoubtedly found acceptance in economics as a profession, but does Samuelson's model say anything about the world we live in?

The question that Samuelson investigates in his article is whether he can find in this very strange economy an equilibrium pattern of individual savings (and the interest rates that go with them) that can be brought into being purely by market transactions. The surprising answer is: no. That answer was not just a surprise to Samuelson, it ought to amaze any economist. After all, it is a central tenet

of economics that market forces are beneficial, and among economists it is usual (although historically incorrect) to reduce this doctrine to Adam Smith's metaphor of an "invisible hand" that guides people's behavior in the market such that the outcome is optimal for everyone. In this economy, that relies on market forces only, the elderly will starve in their retirement age.

With this in the back of our minds, let us look in more detail at the page from Samuelson's article shown in Figure 1. Every economist will recognize from the equations the standard micro-economic theory that Samuelson uses. We also see the equation in which all markets clear, in other words in which all supply and demand orders are filled and the savings plans of all the actors in Samuelson's economy can be fulfilled. The solution renders up an equilibrium interest rate that is equal to the growth rate of the population. Samuelson discusses the shape of preference orderings and assumptions about the time preferences of the actors in the economy (he makes sure actors have no time preferences). All this was part of the standard equipment of the academic economist in the 1950s, as it still is today.

Philosophers of science as Hausman, but also scientists and economists outside of the mainstream look upon this practice with suspicion. Herbert Simon, who later won a Nobel Memorial Prize for Economics, complained as early as 1954 that the science of economics made all kinds of psychological assumptions from the "comfort of the armchair". As said, one of the founders of cognitive psychology, Ward Edwards, wrote in that same year that economists generally use "an armchair method" in their efforts to understand human behavior: "They make assumptions, and from these assumptions they deduce theorems which presumably can be tested, though it seems unlikely that the testing will ever occur" (1954, 381). The late historian and methodologist Terence Hutchison wrote a diatribe against those who considered it perfectly all right for economists to develop

mathematical models from which they could deduce consequences without any attention to empirical testing (1998).³

It does indeed seem improbable that Samuelson would ever actually have wished to test his mathematical model, not only because of the assumptions he made but because he showed that the workings of the market mechanisms in his model economy resulted in a paradox. There is a market equilibrium (in fact there are several amongst which a Pareto-optimal outcome, i.e. an equilibrium in which no further improvement of positions of any of the agents is possible), but no means can be found by which it is possible to achieve this market equilibrium through market transactions: the ‘invisible hand’ does not work.

The paradox becomes all the more clear, Samuelson explains, if the life of each of the actors in the model economy consists not of three but of just two periods. An actor is either working or retired. If he works, he produces a product, but that product can be consumed only in that same period (after all, there are no durable goods). If he has retired, he produces nothing at all. What on earth could a retired person give to someone who was working to persuade that worker to give him, in return, part of the product he was working to produce? Anyone who, like an economist, is skeptical about the altruistic motives of human beings and tends to believe that people are driven by self-interest, could only answer: nothing at all. Samuelson explicitly invokes that intuition; that is how economists are trained. It was therefore “*impossible for any worker to find a worker younger than himself to be bribed to support him in old age.*” The only theoretically possible interest rate Samuelson could think of that would mean everyone at least had something to consume was a negative interest rate of minus 100% per period. Samuelson realized that such an interest rate was purely academic.

It is not hard to imagine why philosopher of science Daniel Hausman concluded that Samuelson’s model was a premier example of the way in which economic theorists had closed

³ Ironically implicating Daniel Hausman as an ultra-deductivist, while Hausman had accused Samuelson of exactly that.

themselves off from sound empirical research very soon after the Second World War. They had given priority to elegant scientific models over “the hard work of other social scientists who have to seek sometimes almost blindly for significant causal factors” (1992, 119). Samuelson’s model seems a paragon of a study that is neither based on meticulous research in the field nor amenable to the instruments of the econometrician and therefore to statistical testing. The model seems utterly devoid of empirical content of any kind. It was not just “armchair economics” because of the suppositions it relied upon but also, and perhaps primarily, because of the incredulous paradoxes in which his model world became entangled. How could such paradoxes ever teach us anything about the real world?

So did Samuelson find himself in the equivalent of a photographer’s darkroom in which he only tells mathematical truths that do not deal with real world issues? Or does Samuelson’s model learn us something about the world after all, something that we perhaps too easily overlook in our daily lives but that is nevertheless fundamental to any understanding of the working of a market economy?

Malthus’ *Essay as thought experiment*

Before getting a firmer grip on these questions, let me first turn to Robert Malthus’ *An Essay on the Principle of Population*, first published in 1798. In this essay, Malthus was responding to the utopian order described by political thinker William Godwin in *An Enquiry Concerning Political Justice* of 1793. Godwin wrote about his utopia at a crucial historical juncture. After the French Revolution (and the subsequent period known as the Terror), not just France but all of Europe was in turmoil. Godwin offered the general public a hopeful message and on condition that the book was sold for the high price of three guineas, it passed the censor. Nevertheless, it became a bestseller by the standards of the time.

Godwin wrote that the source of all evil lay in social institutions, and not in human nature. In this starting point more than an echo can be heard of Jean-Jacques Rousseau and the Enlightenment. Godwin’s utopia was a society not based on private property but on benevolence. When the

institutions of marriage and private property would be abolished, altruism would prevail over self-interest and the problem of overpopulation, to which Godwin was far from blind, would be resolved of its own accord, because of a lessening of the “commerce between the sexes.” Godwin even went so far as to predict that individual life expectancy could increase to infinity as a result of all these changes. Malthus’ strategy was to assume that Godwin’s utopia had been realized in full. He then showed that the population would very probably increase more quickly than food production. This would lead to a fight over food, and the institutions Godwin detested would quickly be reinstated.

The first edition of his *Essay on Population* appeared as a result of “conversations with a good friend” – his father – in a country house (the Rookery, where Malthus’ family lived) where he only had limited materials at his disposal. Hume, Smith, Richard Price’s *Observations on Reversionary Payments* (1771), and Süssmilch’s population statistics, were among his limited references.

However, the essay’s persuasiveness was not so much the result of its use of statistics, or its source-references, or the “postulates” on population and agricultural growth Malthus formulated in its beginnings. Its compelling force resided in a virtuoso chapter in which Malthus played out Godwin’s utopian vision of a society based on ‘benevolence’ rather than ‘self-interest’ against a tiny spot of evidence he took from Price’s *Reversionary Payments*: the message received from the Americas that in its “back settlements” population was doubling in 25 years (Price 1771, 203). That was in a situation where unlimited land was available. But now let’s change that condition. What would happen if resources were no longer plenty, but scarce?

In Chapter X of the *Essay* Malthus supposed that Godwin’s utopia was realised. In Godwin’s utopia there was no human want. Institutions like marriage and private property were abolished. But the missing element in Godwin’s reasoning was his failure to include the consequences of land-scarcity in his system. It was then easy to show how population pressure would soon produce scarcity because of the diverging growth rates of population and agricultural produce. Malthus did not need his postulates to prove this, he merely needed to point at the population growth in the back

settlements of America. As a result institutions like marriage and private property would quickly emerge again, and self-interest would gain prominence over benevolence. Within a period of “perhaps less than three months” Godwin’s ideal order would relapse into the existing order. Malthus thus gave, as it were, speculative Enlightenment philosophers such as Godwin a ‘reality-check’. He showed that his utopia led to a fundamental contradiction that undermined its basic assumptions. A society without institutions, where the governing principle of everyone’s dealings was altruism, could not exist. It was purely a product of the imagination:

We have supposed Mr. Godwin’s system of society once completely established. But it is supposing an impossibility. The same causes in nature which would destroy it so rapidly, were it once established, would prevent the possibility of its establishment.

Whereas Godwin had been able to attribute the population problem to the imperfection of social institutions, in Malthus’ hands it became an inescapable law of nature. “Nature’s mighty feast,” Malthus wrote in the second edition of 1803 did not welcome everyone uninvited because resources were scarce. Writing at the end of the eighteenth century Robert Malthus crushed the utopias of Enlightenment thinkers as Godwin, who, in Malthus’ eyes, put too much confidence in human reason and thus lost sight of the natural constraints that bend whatever benevolent passions might govern us to self-interested behaviour in a situation of scarcity. When “nature’s mighty feast” no longer offers a “vacant cover” for everyone, when nature is no longer plenty, an inevitable system of private property and institutional restraint of the passions produces an inevitable distinction between rich and poor. These institutions did not disrupt an otherwise benevolent state of nature. To the contrary, they protected mankind for an abhorrent natural order.

Malthus’ reasoning is an example of a thought experiment. It presumes a specific state of affairs. In your mind you change one or more factors and then declare what you believe the consequences of that change or changes would be. Or preferably: hidden in received views about the functioning of the (social) world are assumptions of which the thought experiment proves their importance. That is

what gives thought experiments their imprecise character. There is no fixed set of assumptions that then is altered, rather the thought experiment makes clear what was contained in received views, and puts that into question. To the test, if one likes the speak of philosophers of science.

In Malthus' case: we change abundance of land into its opposite. What will happen? It is not a real experiment because you do not actually manipulate factors, you do so only virtually, or mentally. This may be, although not necessarily, because it is impossible to exclude or materially control certain factors. This certainly holds for Malthus' thought experiment, for which we may guess that the contrast between the exploding population in his own Oakwood parish (to which he to his credit never refers) with Godwin's utopian state was so stark, that Price's reference to the back settlements in America pinpointed the issue at stake: limited resources. The strength of Malthus' thought experiment was that he fundamentally disrupted a line of argument that had seemed plausible (with the exception of eternal life), at least to adherents of Rousseau and the Enlightenment. It had been plausible to Malthus' father, who had been such an enthusiast, and it were conversations with his father, which had stimulated Malthus to put his pen on paper. The snippet of evidence on America linked Malthus' thought experiment back to the world, and his father back to earth. Such snippets, as Kuhn emphasized, were essential for thought experiments to succeed; there has to be a credible connection to the world. This does not mean that the theory, or in Malthus' case Godwin's utopian system, should be credible as well. Common sense might conceive of it as credible, but the point of the whole exercise is to show it is not. Thought experiments first aim is not to represent anything in the world, but to destroy existing representations. That is why Schabas' distinction between (representational) models and thought experiments holds water.

Reiss claims that thought experiments seldom have the impact nor appear as often as is claimed. I will grant him the last point – disruptive changes of conceptual frames don't appear on a daily basis, much of what happens in the sciences, including the social sciences and the humanities, is what Kuhn called normal science. And I will also grant him that thought experiments don't stand on

their own. The criticism Malthus received set him on a life long journey to gather evidence for his notorious postulates in which he failed miserably. But certainly up to the acceptance of the New Poor Laws in 1834, the year of his death, the doom scenario of this “ogre” or “mischievous reptile”, as Malthus was referred to,⁴ figured prominently, especially in the conservative press and even though the story is more complex, the nickname of Malthus’ infant discipline, the Dismal Science, figures until today.⁵ Hardin’s tragedy of the commons, itself a contested thought experiment, owes as much to Malthus as to the mathematician William Forster Lloyd (Hardin, 1986). Keynes not incidentally started his review of the long nineteenth century in *Economic Consequences of the Peace* (1919) with Malthus. I could mention Darwin, but I leave it to Reiss to judge whether Malthus’ thought experiment had any effect.

Thomas Kuhn believed that its paradoxical, dislocating character was typical of the thought experiment, which compels us to see our common assumptions in a new light, and this was also precisely what Malthus had in mind in denouncing Godwin’s utopia. Philosophers of the Enlightenment might have found it intuitively plausible to attribute social disorder to human institutions, but on further examination the argument does not stand up. What may have seemed trivial or irrelevant to their intuitions was in fact of far greater importance – the tension between scarce means and population growth is a law of nature independent of social institutions.

Shaking up an axiom

In the introductory lectures he gave as Drummond professor of political economy at Oxford (1832), British political economist Richard Whately argued that paradox was a vehicle for conceptual renewal in the sciences and that political economy in particular was riddled with paradoxes. For Whately the true political economist was a looker-on, a theorist, looking over the heads of the

⁴ These were worse than *ad hominem* attacks, as Malthus was known to have a haze-lip.

⁵ The history of the name is complex. See Maas 2008, and especially Levy 2002.

players, seeking out and explaining paradoxes they did not see (1832, 67-68). With his *Fable of the Bees*, Mandeville was to Whately a premier example of a such a theorist by showing the incoherence of eighteenth century theories of Shaftesbury and others, who took it as self-evident that human behavior was virtuous because it was by its nature directed to the common good. Mandeville had shown, by contrast, how selfishness – at the time an acknowledged sin - in a market-based society happened to serve the general interest and thus destabilized received views.

For Whately the issue was not whether Mandeville’s analysis was immoral, as many of Mandeville’s contemporaries complained, its true importance consisted in that Mandeville revealed an inconsistency in the common rhetoric of duties and virtues and society’s love of wealth; either society took this rhetoric seriously, with all consequences for the production of luxuries, or it had to accept that “vices” were constitutive of society. Its subtitle, *Private Vices, Publick Benefits*, expressed this startling paradox: even the most selfish behavior could be in the public interest. Almost all prominent moral philosophers of his day, from Berkeley and Hutcheson, via Hume and Smith to Kant, felt obliged to come to terms with Mandeville’s provocation in one way or another. Kant fully recognized the paradoxical nature of Mandeville’s vision of the source of sociability when he coined the phrase “ungesellige Geselligkeit”, unsocial sociability, in his outline of human history of 1784. Hayek considered that Mandeville had “asked the right question,” to which his political opponent Joan Robinson added that this question, “has never been answered properly until today.” (quoted from Maas 2004, 734).

In one of the basic texts of political economy, *The Wealth of Nations* of 1776 (pp. 26-27), we can read Smith’s answer, which in the image of the ‘invisible hand’ gained the status of an unshakeable axiom that replaced what Shaftesbury and others had presented as a truism:

It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages.

Samuelson's extremely successful textbook *Economics*, published in 1948, used an everyday example to illustrate this axiom. *Economics* had been written for students of technical sciences at the Massachusetts Institute of Technology (MIT), but it quickly grew to become the best sold textbook ever after the Second World War, which served as a model for all economic textbooks ever since. After having read the manuscript members of the board of trustees of MIT accused Samuelson of maintaining opinions that were subversive if not straightforwardly "communistic" and so he tweaked the text, but only slightly, to adjust to their unshakable belief in the virtues of the market (Giraud 2014).

At the start of the third chapter (of the fourth edition which I consulted), Samuelson discusses a "famous example" (which, without citing any sources, he transfers from London to New York; it had originally been used by Richard Whately in his Drummond lectures of 1832). He asks the reader to imagine that goods in New York need to be distributed by some kind of central planner. Such a thing would surely be inconceivable. To Samuelson this alone was sufficient proof that distribution is better done by individuals acting as individuals, motivated by their own self-interest. Markets coordinate selfish decisions by means of an invisible hand, so that they serve the general interest. But what Samuelson presents in his textbook as a foregone conclusion – the conviction broadly shared among economists and the businessmen at MIT's board of trustees, that markets work efficiently – was the subject of research in his academic work. Let us go over his overlapping generations model one more time.

The practical implications of a paradox

At the start of his article, Samuelson presents his subject as purely theoretical. If theory is defined as having a mathematical structure that needs to be confronted with statistical data, and otherwise remains an abstract play with symbols, then the reader may gain this impression indeed. But this is not how Samuelson reasons in his article, nor does this do justice to the purpose for which

Samuelson built his model. The subject of the paper concerns an issue that is not merely interesting from a theoretical point of view but ought to interest any economist who has faith in the working of the market, and indeed even a broader public, though Samuelson's choice of publication is limited to an academic audience of economists. But on the second page, he poses a very practical question that seems far off the theoretical stakes he started with: why would a person want to contribute to a pension that is paid out to someone he or she does not know? What does that person have to gain? From a theoretical question about the mutual consistency of saving and investment plans in an economy – a question dealing with a general equilibrium in a specific market – Samuelson moves to a mundane, but no less fundamental issue, as we know from present-day discussions about the problem of an aging population and the viability of the pension system: who will pay for the elderly?

In the article itself, Samuelson moves back and forth between the elementary mathematics and the common sense explanations that one could give for every result he derives. All these common sense explanations, Samuelson tries to convince the reader, fail, because they do not obey the limits set by the model. Hausman summarizes the barebones of this model, but has difficulty in making sense in the discussions on welfare considerations, common sense explanations of equilibria derived, which Samuelson inserts and which only seem to distract from the main point Samuelson is making: in a world without durables markets will fail. For convenience of the reader, Hausman makes Abraham, Isaac, and Jacob the protagonists of Samuelson's three generations model. But this means he takes no notice of the fact that such a family relation is precisely excluded by Samuelson from the beginning. Not as a theoretical assumption, but because it would describe a world we no longer live in. In earlier times, Samuelson writes, it was customary for children to look after their parents, but "that is now out of fashion." That is Samuelson's equivalent of Malthus' evidence on America. Economists believe that markets can instead optimally coordinate behavior. Is this true?

As a "test", Samuelson constructs a model world in which "cold and selfish markets" by themselves have to care for the elderly. No family ties exist between the actors in his model. There is

no money and no possibility of savings; there are no durables. Preferences are symmetric in their argument, so as to rule out any time-preference of actors. Samuelson does everything he can to make sure that is market transactions, and only market transactions that will bring savings and investments into equilibrium. What does the market do?

He investigates his model world for a static and for a growing population, and reaches surprising conclusions. Indeed they surprise even him. The market does nothing at all. The paradoxes in which he then revels revolve around the tension between (the economists') common sense belief in the market and the mathematics of the model. "How can we reconcile this with the mathematics?" (Samuelson 1958, 474). This tension teaches us something about not only the model world, but our world.

It turns out that in theory a social optimum exists in which individuals maximize their savings over their lifetimes, based on a what one could see as a reasonably normal interest rate. But when Samuelson then looks at how such a social optimum is to be achieved, the model economy ends up in the paradoxical situation in which although there is a social optimum, it cannot be achieved by market dealings – there is no invisible hand. Economically relevant interest rates (that is, interest rates that enable individuals to get what they need through market trading) diverge considerably from the optimum. There are negative interest rates and even interest rates of minus 100% and that are clearly only academic. So Samuelson is forced to conclude that in his model world "the social optimum configuration can never be reached by the competitive market, or even be approached in ever so long a time."

Samuelson's model world demonstrates that we cannot simply rely on markets for a pension system. This undermines, as Samuelson emphasizes, the "fundamental intuitions" of economists about the beneficial workings of markets. Samuelson thereby clarifies why it is that in *our* world there are compulsory pension schemes and, something he looks at in more detail in his concluding section, why such a thing as money is important in an economy. Money, as a store of value, can be a

means of transferring wealth to the future. As soon as money exists, a market can do its beneficial work, but then, it is not *just* the market that does its work; it is the institute of money as well, and that implies an institute that guarantees its value. Samuelson's model world teaches us lessons about the importance of enforced cooperation and of a stable currency. Such conclusions are obviously not just of importance to the theoretician; they have a real-world political charge. If selfish markets are not going to take care of the elderly, then politics needs to ensure there are institutions that will do so.

Economists (and not only economists) like to believe that markets work well, which is to say that they deliver the best outcomes for everyone. Economists and politicians regularly appeal to their own "experience" to substantiate this belief. But faith and experience will not get an economist or anyone else anywhere in Samuelson's model world. If in that world it is left to the market to achieve an optimal situation, then individuals will be disappointed and even die. The hope that there is an "invisible hand" steering an individual's decisions in the right direction is dashed by Samuelson, and an arithmetical example makes this abundantly clear. We cannot, is the message, simply trust to our – that is the economist's - common sense. More specifically, we cannot simply rely on the market to do its job in all circumstances. Without adequate institutions, markets do not work. So it will be no surprise to discover that Samuelson not only advised the Kennedy administration on economic policy but was an outspoken opponent of perhaps the greatest free-market ideologist in the United States, Chicago economist Milton Friedman. Consulted about monetary policy by his former teacher and mentor Alvin Hansen, Samuelson wrote, unusually impolite even in personal correspondence, that Friedman was "a nuisance" (quoted in Maas 2014).

Conceptual and empirical exploration

At the end of his article, Samuelson compares his working method with the performance of an experiment. In this connection economists use the expression *ceteris paribus*: if all other conditions remain the same, then the effect of a change to one variable in a model can be determined

unambiguously, as it would be in an experiment. The implicit assumption is that the model world and the real world resemble each other in some way, or at any rate exhibit a sufficient degree of resemblance for us to draw conclusions about our world. This is how many contemporary philosophers of science think about models. In the laboratory, attempts are made to isolate a connection between cause and effect, while controlling for other factors. In their models economists isolate a single important aspect that they then modify and their model worlds thereby become a fictional analogue for an experiment that could possibly or actually be carried out in a laboratory.

Hausman uses Samuelson's example to argue that models like this have no empirical pretensions, indeed cannot have any, since use is made of concepts that resist all empirical testing. Samuelson constructs an artificial, mathematical world and draws conclusions from it. Those conclusions may be surprising or unsurprising, but they relate only to the internal workings of the model. What Samuelson does is therefore purely a matter of conceptual exploration, of research into the conceptual implications of a formal model world. Whether or not the outcomes produced by such a model can be tested empirically (and thereby gain empirical significance) is a separate issue. Samuelson makes no attempt to do so, and as Hausman underlines, it is difficult to see how his model could ever be subjected to an empirical test, since the assumptions Samuelson makes are too unrealistic and the outcomes of his models too remote from what we know empirically about our economies. Samuelson's exercise is therefore of purely hypothetical significance, not in the sense of an empirical hypothesis but in the sense that it is irrelevant with respect to the real world.

This, Hausman claims, is a deficiency of all such models, and he concludes that economists are more interested in the "elegant finesses" of their mathematical formulations than in the "hard work" that social scientists carry out in other disciplines. This lack of interest in empirical research culminates in the misguided attention of economists on how general equilibrium is reached in precisely defined markets. To anyone who agrees with Hausman, the comparison Samuelson makes between what he does and an experiment can only be misleading.

Paradoxes change the world

So is Hausman's verdict on Samuelson justified? At the start of his article, Samuelson makes a number of observations about our world (most importantly: children in the West no longer look after their parents, so strangers have to be 'bribed' to do so), but at the same time he makes clear that he needs to introduce several rigorous abstractions to make the problem he wants to address visible. Hausman's misconception is that Samuelson is developing a hypothesis that can then be tested empirically.

Nowhere does Samuelson claim that his model represents our world in such a way that it can be tested, either experimentally or statistically. On the contrary, his model world is fundamentally unlike our own. It is precisely by taking us with him into a quite different world that he can offer us insights into the characteristic features of *our* world, features that we, as economists who believe in the beneficial workings of markets, might easily overlook: markets have to be tamed by institutions. This becomes clear in Samuelson's article only when he demonstrates that our "common sense" leads us to conclusions that cannot be valid (despite all our faith in self-regulating markets, whether or not based on experience). What Samuelson puts to the test is the economists' intuition that selfish markets produce efficient results. This is what the model is intended to demonstrate. Only then are we confronted with a surprise: they don't. Markets do not automatically work. The lessons of such models, to quote Whately, "may be sense but at least they are not *common-sense*" (1832, p. 65).

For this reason it is more to the point to compare Samuelson's theoretical modelling with thought experiments than with actual experiments. Thought experiments can be seen as a form of conceptual exploration, but they are more than that; they change our worldview. It is not so much a matter of a conceptual apparatus that is refined but rather of our understanding of what we observe. A thought experiment begins with everyday observations, then takes us into a world in which those observations are put to the test before returning us to the everyday world armed with concepts that alter the way we see it. After Malthus, it becomes difficult if not impossible to rule out scarcity of

resources as a constitutive factor of the social order. Samuelson's model may not have had such a dramatic impact, but if we consider the moment of publication, when building the welfare state was hotly discussed in the academic and public domain, after Samuelson's contribution it is too easy to dismiss forced pension payment schemes as state paternalism. Markets will not automatically provide the elderly with a viable alternative.

Kuhn stresses that thought experiments are not primarily useful in putting confused or inconsistent concepts to the test (that would merely amount to a clarification of our conceptual apparatus) but in making clear that our observations do not meet our expectations and that those observations must therefore be understood differently (and our expectations of them adjusted). This is the reason why Kuhn makes a connection between his thought experiments and his ideas about scientific revolutions. It is precisely because paradoxes have such a central place in thought experiments that it may be a long time before readers are prepared to accept the lessons that need to be drawn.

This applies to Mandeville's *The Fable of the Bees*, a book condemned as a 'public nuisance' a decade after it appeared. The assertion that acting out of self-interest could benefit the community as a whole was still anathema in 1723. The fact that this same assertion became one of the guiding principles of Adam Smith's *The Wealth of Nations* (1776), and in the nineteenth century was even seen as the axiom that every political economist should take as a starting point, shows that what at first goes against everyday knowledge can later become part of it.

To a degree something similar applies to Samuelson's model. Since the Second World War, belief in the free market has been regarded more as knowledge than as faith. What Samuelson makes clear is that the market does not always do its beneficial work unconditionally, one example being the transfer of income from one generation to another. By taking us with him into a world in which he takes this issue of the handing down of wealth and looks at it in a naked form, he is able to call into question the assumption that markets work perfectly. He can then go on to show that a market

society needs compulsory forms of cooperation (taxes, state pensions) in order to function efficiently.

Conclusions

It is in fact surprising that philosophers have devoted most of their attention to thought experiments in physics, as for long it was a common trope that the social sciences (or their predecessors) were unable to perform material experiments, if only because of the complexity of their subject matter. John Stuart Mill made this, it is well known, a stronghold in his defense of political economy as a science of tendencies, an “inexact science” as Daniel Hausman had it. Before Mill, first Drummond professor in political economy Richard Whately, recommended something quite similar to thought experiments as the method of preference for political economy.

In this essay I discussed two examples of thought experiments that can be seen as a test of received views. Malthus wrote his essay in answer to what he considered a misguided and destructive reliance on a benevolent natural state of mankind. Samuelson wrote against the intuition of economists that the workings of the market produce the best outcomes for actors in the market. Just how deep that intuition lies is clear from the lamentation of a Dutch professor of economics and prominent social democrat in 2001 that the general public really ought now finally to take account of the “truths of economic science,” namely that markets lead to efficient outcomes.

True, Samuelson’s model rests on completely unrealistic assumptions, yet it teaches us something about our world. Not for the reason Milton Friedman would give, since he would claim that the reality content of assumptions is irrelevant as long as the predictions are correct. Nor according to the demand that Hausman makes of models, insisting that they must produce hypotheses that can actually be tested. Obviously the criterion for the usability of Samuelson’s model does not lie in its predictive power, since it is clear from the start that his model leads to outcomes that are beyond the boundaries of observation. So what is the value of such models?

For philosopher of science Daniel Hausman, the answer lies in conceptual exploration.

Samuelson's model calls into question the meaning of concepts. His analytical exercise reveals the formal implications of the concepts he examines. According to Hausman, Samuelson's model has no empirical consequences, indeed he stresses that it is difficult to see how that model could have empirical implications and he concludes that ultimately it amounts to the same mathematical playing around as the theory of general equilibrium that ran into the sand in the late 1970s.

If we take Samuelson's model to be a thought experiment, we arrive at a different picture. Samuelson constructs a world that is deliberately made different from ours with one exception: transfers of wealth are coordinated through markets and not via family ties. Samuelson demonstrates that in such a world, markets can fail. If we then look at our world, institutions that (for those who believe in the beneficial workings of the market) seemed superfluous suddenly make sense: compulsory pension savings; money as an instrument for transferring value to the future.

Anyone looking for predictions, like Friedman, or testable hypotheses, like Hausman, will wonder what is empirical about such conclusions. Samuelson's model neither makes predictions nor represents a real economy. It is a thought experiment that calls into question fundamental intuitions that economists have about how market economies work, and at the same time it explains why the world in which we live does not, indeed cannot, accord with those intuitions. It shows why our world is not the way we (or some of us) would like to think.

Samuelson formulated his thought experiment mathematically, since he believed mathematics to be the language in which economists need to express themselves. We have seen from the example of Malthus that a thought experiment does not necessarily have to be formulated in mathematical terms. Thought experiments are about asking the right questions, and good questions, as Samuelson

once said, outrank easy answers. Malthus and Samuelson destabilized easy answers by asking the right question: whether nature and markets are benevolent by themselves.⁶

References

- Edwards, W. (1954). The theory of decision making. *Psychological bulletin*, 51(4), 380.
- Giraud, Y. (2014). Negotiating the “Middle-of-the-Road” Position: Paul Samuelson, MIT, and the Politics of Textbook Writing, 1945-55. *History of Political Economy*, 46 Annual Supplement. Durham, Duke University Press, 134-152.
- Godwin, W. (2013). *An enquiry concerning political justice*. Oxford: Oxford University Press.
- Hardin, G. (1968). The tragedy of the commons. *science*, 162(3859), 1243-1248.
- Hausman, D. M. (1992). *The inexact and separate science of economics*. Cambridge: Cambridge University Press.
- Hutchison, T. (1998). Ultra-Deductivism from Nassau Senior to Lionel Robbins and Daniel Hausman. *Journal of Economic Methodology*, 5(1), 43-91.
- Kühne, U. (2005). *Die Methode des Gedankenexperiments*. Frankfurt am Main: Suhrkamp.
- Kuhn, T.S. (1977). “[A function for thought experiments](#).” In *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press. 240-265.
- Levy, D. M. (2002). *How the dismal science got its name: Classical economics and the ur-text of racial politics*. Ann Arbor: University of Michigan Press.
- Malthus, T. R. (1966). *The travel diaries of Thomas Robert Malthus*. P. James (Ed.). Cambridge: Cambridge University Press.
- Malthus, T. R. (1798). 1986. *An Essay on the Principle of Population*. In *The Works of Thomas Robert Malthus*, EA Wrigley and D. Souden, eds, vol 1. London: Pickering.
- Malthus, T. R. (1803). An essay on the principle of population; a view of its past and present effects on human happiness; with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions.
- McAllister, J. W. (1996). The evidential significance of thought experiment in science. *Studies In History and Philosophy of Science Part A*, 27(2), 233-250.
- Maas, H. (2004) “Bernard de Mandeville.” In *Dictionary of British Economists*. Bristol: Thoemmes Continuum, pp. 730-734.
- Maas, H. (2008). “A Hard Battle to Fight”: Natural Theology and the Dismal Science, 1820–50. *History of political economy*, 40(5), 143-167.
- Maas, H. (2014). Making Things Technical: Samuelson at MIT. *History of Political Economy*, 46, annual supplement. Durham: Duke University Press, 272-294.
- Morgan, M. S. (2003). *Experiments without material intervention: Model experiments, virtual experiments and virtually experiments* (pp. 216-235). Pittsburgh: University of Pittsburgh Press.
- Morgan, M. S. (2012). *The world in the model: How economists work and think*. Cambridge: Cambridge University Press.
- Price, R. (1771). *Observations on reversionary payments ...* (London, Printed for T. Cadell).
- Reiss, J. (2002). *Causal inference in the abstract or seven myths about thought experiments*. London School of Economics, Centre for Philosophy of Natural and Social Sciences.

⁶ I would like to thank Steve Medema and Roger Backhouse for pointing me to Samuelson’s quote that Steve Medema saw used at the University Church of the University of Chicago.

- Samuelson, P. A. (1958). An exact consumption-loan model of interest with or without the social contrivance of money. *The Journal of Political Economy*, 467-482.
- Schabas, M. (2008). Hume's monetary thought experiments. *Studies in History and Philosophy of Science Part A*, 39(2), 161-169.
- Simon, H. A., & Bartel, R. D. (1986). The failure of armchair economics. *Challenge*, 18-25.
- Sorensen, R. (1991). Thought experiments. *American Scientist*, 250-263.
- Sorensen, R. (1992). *Thought Experiments*. Oxford: Oxford University Press.
- Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, 7(1), 1-31.
- Whately, R. (1832). Introductory lectures on political-economy, delivered at Oxford. *Easter term, MDCCCXXXI. With remarks on tithes and on poor-laws and on penal colonies*, JW Parker and son.
- Winch, D. (1996). *Riches and poverty: an intellectual history of political economy in Britain, 1750-1834*. Cambridge University Press.

