

# Avant-Garde Philosophy of Economics

Graham Joncas

## Introduction

To most people, the title of this paper is a triple oxymoron. Those left thoroughly traumatized by Econ 101 in college share their skepticism with those who have dipped their toe into hybrid fields like neuroeconomics and found them to be a synthesis of the duller parts of both disciplines. For the vast, vast majority of cases, this sentiment is quite right: ‘philosophy of economics’ tends to be divided between heterodox schools of economics whose writings have entirely decoupled from economic formalism, and—on the other side of the spectrum—baroque econophysicists with lots to say about intriguing things like ‘quantum economics’ and negative probabilities via  $p$ -adic numbers, but typically within a dry positivist framework. As for the middle-ground material, a 20-page paper typically yields two or three salvageable sentences, if even that. Yet, as anyone who follows my Twitter knows, I look very hard for papers that aren’t terrible—and eventually I’ve found some.

Often the ‘giants’ of economic theory (e.g. Nobel laureates like Harsanyi or Lucas) have compelling things to say about methodology, but to include them on this list seems like cheating, so we’ll instead keep to scholars who most economists have never heard of. We also—naturally—want authors who write mainly in natural language, and whose work is therefore accessible to readers who are not specialists in economic theory. Lastly, let’s strike from the list those writers who do not engage directly with economic formalism itself, but only ‘the economy’. This last qualification is the most draconian of the lot, and manages to purge the philosophers of economics (e.g. Mäki, McCloskey) who tend to be the most well-known.

The remaining authors make up the vanguard of philosophy of economics—those who alchemically permute the elements of economic theory into transdisciplinary concoctions seemingly more at home in a story by Lovecraft or Borges than in academia, and who help us ascend to levels of abstraction we never could have imagined. Their descriptions are ordered for ease of exposition, building from and often contradicting one another. For those who would like to read more, some recommended readings are provided under each entry. I hope that readers will see that people have for a long time been thinking very hard about problems in economics, and that thinking abstractly does not mean avoiding practical issues.

## 1 M. Ali Khan

Khan is a fascinating character, and stands out even among the other members of this list: by training he is a mathematical economist, familiar with some of the

highest levels of abstraction yet achieved in economic theory, but at the same time an avid fan of continental philosophy, liberally citing sources such as De Man (a very unique choice, even within the continental crowd!), Derrida, and similar figures on the more literary side of theory, such as Ricoeur and Jameson. It may be helpful to contrast Khan to Deirdre McCloskey, who has written a couple of books on writing in economics: McCloskey uses undergraduate-level literary theory to look at economics, which (let's face it) is a fairly impoverished framework, forcing her to cut a lot of corners and sand away various rough edges that are very much worth exploring. An example is how she considers the Duhem-Quine thesis to be in her own camp, which she proudly labels 'postmodern'—yet, just about any philosopher you talk to will consider this completely absurd: Quine was as modernist as they come. (Moreover, in the 30 years she had between the first and second editions, it appears she has never bothered to read the source texts.) Khan, by contrast, has thoroughly done his homework and then some.

Khan's greatest paper is titled "The Irony in/of Economic Theory," where he claims that this 'irony' operates as a (perhaps unavoidable) literary trope within economic theory as a genre of writing. Khan likewise draws from rhetorical figures such as synecdoche and allegory, and it will be helpful to start at a more basic level than he does and build up from there. The prevailing view of the intersection of mathematics and literary theory is that models *are* metaphors: this is due to two books by Max Black (1962) and Mary Hesse (1963) whose main thesis was exactly this point. While this is satisfying, and readily accepted by theorists such as McCloskey, Khan does not content himself with this statement, and we'll shortly see why.

Consider: a metaphor compares one thing to another on the basis of some kind of structural similarity, and this is a very useful account of, say, models in physics, which use mathematical formulas to adequate certain patterns and laws of nature. However, in economics it often doesn't matter nearly as much who the particular *agents* are that are depicted by the formulas: the Prisoner's dilemma can model the behaviour of cancer cells just as well as it can model human relations. If we change the object of a metaphor (e.g. cancer cells → people), it becomes a different metaphor; what we need is a kind of rhetorical figure where it doesn't matter if we replace one or more of the components, provided we retain the overall framework. This is precisely what **allegory** does: in one of Aesop's fables, say "The Tortoise and the Hare," we can replace the tortoise by a slug and the hare by a grasshopper, but nobody would consider this to be an entirely new allegory—all that matters here is that one character is slow and the other is fast. Moreover, we can treat this allegory itself as a metaphor, as when we compare an everyday situation to Aesop's fable (which was exactly Aesop's point), which is why it's easy to treat economic models simply as metaphors, even though their fundamental structure is allegorical.

The reason this is important is because Khan takes this idea to a whole new level of abstraction: in effect, he wants to connect the allegorical structure of economic models to the allegorical nature of economic texts—in particular, Paul Samuelson's *Foundations of Economic Analysis*, which begins with the

enigmatic epigraph “Mathematics is a language.” For Khan: “the *Foundations* is an allegory of economic theory and...the epigraph is a prosopopeia for this allegory” (1993: 763). Since I had to look it up too, *prosopopeia* is a rhetorical device in which a speaker or writer communicates to the audience by speaking as another person or object. Khan is quite clear that he finds Samuelson’s epigraph quite puzzling, but instead of just saying “It’s wrong” (which would be tedious) he find a way to *détourne* it that is actually quite clever. He takes as a major theme throughout the paper the ways that the same economic subject-matter can be depicted in different ways by using different mathematical formalisms. Now, it’s fairly trivial that one can do this, but Khan focuses on how in many ways certain formalisms are observationally equivalent to each other. For instance, he gives the following chart (1993: 772):

<i>probability space</i> (pr. space)	normed measure space
<i>elementary event</i>	point belonging to the space
<i>event</i>	measurable set
<i>sure event</i>	whole space
<i>impossible event</i>	empty set
<i>probability</i> (pr.)	normed measure
<i>almost sure, almost surely</i> (a.s.)	almost everywhere
<i>random variable</i> (r.v.)	finite numerical measurable function
<i>expectation</i> E	integral $\int$

**Fig. 1.** Correspondence between probability & measure theory

That is to say, to present probabilistic ideas using the formalism of measure theory doesn’t at all affect the content of what’s being said: it’s essentially just using the full toolbox of real analysis instead of only set notation. What interests Khan here is how these new notations change the differential relations between ideas, creating brand new forms of Derridean *différance* in the realm of meaning—which, in turn, translate into new mathematical possibilities as our broadened horizons of meaning let us develop brand new interpretations of things we didn’t notice before. Khan’s favorite example here is nonstandard analysis, which he claims ought to make up a third column in the above chart, as probabilistic and measure theoretic concepts (and much else besides) can likewise be expressed in nonstandard terms. To briefly jot down what nonstandard analysis is: using mathematical logic, it is possible to rigorously define infinitesimals in a way that is actually *usable*, rather than simply gestured to by evoking marginal quantities. While theorems using such nonstandard tools often differ greatly from ‘standard’ theorems, it is provable that any nonstandard theorem can be proved standardly, and vice versa; yet, some theorems are far easier to prove nonstandardly, whence its appeal (Dauben, 1985). In economics, for example, an agent can be modelled as an infinitesimal quantity, which is handy for general equilibrium models where we care less about particulars than about

aggregate properties, and part of Khan's own mathematical work in general equilibrium theory does precisely this.

To underscore his overall point, Khan effectively puts Samuelson's epigraph through a prism: "Differential Calculus is a Language", "Convex Analysis is a Language", "Nonsmooth Analysis is a Language", and so on. Referring to Samuelson's original epigraph, this lets Khan "interpret the word 'language' as a metonymy for the collectivity of languages" (1993: 768), which lets him translate it into: "Mathematics is a Tower of Babel." Fittingly, in order to navigate this Tower of Babel, Khan (following Derrida) adopts a term originating from architecture: namely, the distinction between keystone and cornerstone. A *keystone* is a component of a structure that is meant to be the center of attention, and clinches its aesthetic ambiance; however, a keystone has no real architectural significance, but could be removed without affecting the rest of the structure. On the other hand, a *cornerstone* is an unassuming, unnoticed element that is actually crucial for the structural integrity of the whole; take it away and the rest goes crashing down.

Like many of Khan's influences, it's a charmingly idiosyncratic choice of focus, since most continental philosophers (even card-carrying Derrideans) seldom if ever use the term. But the role it plays for Khan's argument is to break from the notion of a unitary 'master signifier', and instead note the simultaneous co-dependence and conflict between structural integrity and aesthetic *gestalt*. This abandonment of 'master signifiers' is intimately linked to allegory, as "allegory can never be reduced to a metaphor, to a symbol nor even to a metonymy or a synecdoche which would designate the totality of which they are a part." (1993: 791). Allegory as a figurative structure is therefore inherently "disjunctive and non-totalizing" (1993: 794) in a way closely parallel to irony as disjunction of a statement with its context.

The problem with viewing economic models as allegories is that this also implies that they are meta-allegories: since allegories (such as "The Tortoise and the Hare") can be treated as a metaphor, and metaphors can act as the objects within an allegory, we may thus arrive at a *mise en abyme* of allegories stacked on allegories stacked on allegories...or we would, anyway, if it weren't for Khan's keystone/cornerstone distinction. For Khan, "what makes the allegories of mathematical economics particularly interesting is that one can locate their cornerstones in the economics as well as in the mathematics" (1993: 792). He lists as illustrations the notions of a *commodity space* and of *agents*, both of which are amenable to nonstandard tools, where for all practical purposes we can get away with assuming that the number of commodities and/or agents is infinite—a far cry indeed from Marx's definition of commodities as goods that acquire exchange value in markets! What these cornerstones reveal to us is how, through the use of new mathematical approaches, great towers of allegories can be shown to simply be variations on the same allegory (*à la* Tortoise/Hare → Slug/Grasshopper). After developing suitable pidgins and creoles, the workers at the Tower of Babel may now begin anew, only to have their work torn down

again by future mathematical tools. And this brings us (finally!) to the irony in/of economic theory (1993: 795-6):

The irony of economic theory lies in the fact that it has to save itself by seeking to assure its ability to mean within a sphere posited as independent of economic theory. The imperative of theory thus commands—and this constitutes its profound paradoxy—the abandonment of the uncertain basis of theory in order to seek within a region, itself no longer theory, the stable ground of its capacity for truth and the reliable confirmation that theory really and indeed is theory. [...] The imperative of economic theory speaks only in the break of economic theory.

To elaborate: in *The Economy of Literature*, Shell (1978: 5-6) writes: “a metaphor about language and a metaphor about money are both metaphors of metaphorization”; this sounds impressive until we realize that all he’s saying is that money, through its function of ‘equating’ disparate objects, operates in a ‘metaphorical’ fashion. Khan’s position, more subtly, can be characterized as saying that economic theory is an allegory of allegorization. However, this is in no way a ‘master’ allegory, as its components, Babel-esque, do not cohere. Economics is not self-sufficient, but must draw from without itself: “The metaphors must keep moon-lighting” (1993: 978). Likewise, every economic proof operates as part of its own disavowal: “the language it speaks is the language of self-resistance” (De Man, in Khan, 1993: 798). Hence, irony. It is an irony **in/of** because of this feedback loop of disavowal between economic theory (considered as an edifice) and its objects, where the boundary between economics and non-economics is continually ruptured. Further, Khan’s essay “ironizes the irony in irony” (1993: 798) because [1] identifying this irony of economic theory as such, it produces a meta-irony in that this “crucial structural characteristic” (1993: 798) purports to ‘explain’ economic theory, but itself simply becomes one most aspect of this boundary between economics and non-economics, to be ruptured in turn; moreover, [2] as his paper is itself a model, tracing out the processes by which this irony manifests itself, it can only open up economic theory to brand new ironies (perhaps in the philosophical realm) that in turn must continually undermine themselves.

The major drawback of Khan’s work is that at times he is purposely enigmatic, and so his points are not always as clear as they could be. It’s also a pity that most of his other philosophical papers don’t measure up to his 1993 one. Most of his (open-access) papers in the *Pakistan Development Review* are quite good, and—if one is so inclined—his more didactic papers on mathematics (vs. actual math papers) are also very stimulating. His philosophical papers are, sadly, fairly mediocre. He *tries* to apply a similar formula to theorists such as Marshall (Khan, 2004a), Hahn (Khan, 2004b), and Hayek (Khan, 2005), but these are all fairly unsatisfying. Khan’s papers in general tend to lack anything resembling a conclusion, and it’s difficult to put your finger on precisely what you’ve gotten out of reading a paper of his. However, a very recent paper published in 2014 is very impressive, a worthy follow-up to his 1993 paper: it deals

with Georgescu-Roegen, well-known for his programme of ‘thermoeconomics’, and also dwells more at length upon the role of nonstandard analysis in economics (though less on literary theory). For more on the theme of irony, see Élie Ayache’s “The Irony in the Variance Swaps” (2006). Out of all the people trying to apply literary theory to economics—and there are a lot of them—Khan is the best we’ve got, full stop. And at the very least, he’s the only person in the world who can make general equilibrium theory sexy.

- Khan, M. (1993). “The Irony in/of Economic Theory.” *Modern Language Notes* 108(4), 759-803
- Khan, M. (1991). “On the Languages of Markets.” *Pakistan Development Review* 30(4), 503-4
- Khan, M. (2014). “Representation, Language, & Theory: Georgescu-Roegen on Methods in Economic Science.” *Journal of Economic Issues* 48(1), pp. 49-87

## 2 Jean-Pierre Dupuy

Dupuy is interested in the intersections between analytic and continental philosophy, and suggests that the notions of ‘game’ and ‘play’ provide an optimal bridge between the two (1989: 37). His main connection to economics is how he draws from game theory and decision theory, in particular the work of Robert Aumann. By the standards of philosophy of economics, this is an incredible amount of due diligence; among philosophy papers treating game theory, it’s not unheard of for a paper to cite the introduction to von Neumann & Morgenstern’s *Theory of Games and Economic Behaviour* and state that they will treat it as representative of the entire corpus of game theory, notwithstanding the fact that it was written in 1944, before even the concept of Nash equilibrium was first published in 1951. Dupuy’s work, by contrast, is interesting from both a philosophical and economic point of view, and may even prove to be the first step toward formalization of continental ideas. Dupuy’s main influence on the continental side is Lacanian psychoanalysis, which has become almost a staple in North American humanities departments. While combining game theory and Lacan likely seems incongruous at first, there is in fact ample textual evidence that Lacan was himself influenced by game theory. As Liu observes in a very eye-opening paper (2010: 291):

So much has been written about how Lacan rejected American ego psychology that we have nearly lost sight of how he simultaneously engaged with American game theory and cybernetics. Contrary to common belief, a great deal of what we now call French theory was already a translation of American theory before it landed in America to be reinvented as French theory. For example, it is startling to ponder how the English word game from game theory metamorphosed into the noun *play* in literary theory through the round-trip intermediary of the French words *jeu* and *jeux* in translation.

This very much goes against current anti-mathematical trends in the humanities, which tends to think of game theory as a deeply impoverished account of human subjectivity, since in their view it throws out all the good stuff. Yet, from someone like Dupuy's point of view, this is a straw man argument. If game theory works for cancer cells, semantics and evolution, the idea that it doesn't apply to humans because they "aren't rational" misses the point. The object of game theory is not 'rationality' in any anthropic sense, but something blurring structure and agency that we have no words for. Lacan realized this as well, leading to his interest in mathemes, as Leader (2000: 174) nicely explains: "Logical relations come into play at exactly those points in a subject's life where meaning and understanding break down, and thus at the horizon of such structures will always be a set of impossibilities, impossible in the sense of contradictions and in the sense of impossible to say or make mean." Both game theory and psychoanalysis, then, help to elucidate the limit experiences of subjectivity: both help us to see that profound emotional reactions—love as a solution to the iterated Prisoner's dilemma, for instance—are often deeply 'rational' in character.

Now we can focus on Dupuy's main ideas. Some people say that the most radical innovation of game theory is simply the notion of a payoff matrix: it forces you to step into the other player's shoes, and helps you to realize that the other player will be reasoning in the exact same way as you are. So you'll be stepping into their shoes, and they'll be stepping into your shoes in full knowledge that *you* are stepping into *their* shoes, and you'll be stepping into their shoes in full knowledge that...et cetera. This creates a relation of "infinite specularity" (Dupuy, 1989: 45) known in game theory as common knowledge (CK). Another way of framing this is to take the phrase "I know that you know that *P*," where the variable *P* likewise stands for "I know that you know that *P*," thus creating an infinitely fractal structure. (Much like the old joke: the "B." in "Benoît B. Mandelbrot" stands for "Benoît B. Mandelbrot.") However, CK only occurs when this specular relation is truly infinite: if it breaks down at any point, it is simply higher-order knowledge and not CK. However, this kind of breakdown arises in many philosophically interesting situations. A simple example is how Žižek characterizes ideology (using Donald Rumsfeld's famous typology) as "unknown knowns," where you know that *P*, but don't know that you know that *P*. Next, Dupuy draws from the formalism of epistemic games to show its uncanny resonance with the Lacanian notion of the 'Big Other' (1989: 48-9):

Aumann has contributed an important finding which we would reformulate in the following way: there exists a subject of CK. In epistemic logic, we can define 'knowledge' by introducing an operator that satisfied certain axioms. Such a definition is purely syntactical. To this syntax, however, corresponds a semantic interpretation by means of which we associate a knowing subject with each knowledge operator. Aumann has shown that given *n* elementary knowledge operators, the operator of CK associated with them satisfies the axioms that define knowledge. The operator of CK is thus itself a knowledge operator. A subject must therefore

be associated with it, and this is the subject of CK. Following Lasry, we will call this subject the Other (is this Lacan's 'big Other'?)

By construction, we have the following property of this Other: *the Other knows that P if and only if P is CK*. [...] In fact, it is not hard to show that there is a single subject, the Other, the subject of CK, for whom the following holds true: *the Other knows that P if and only if everyone knows that the Other knows that P*.

[...] One could say along with Lasry that this Other is the 'symbolic instance'. But in saying this we must be aware of the consequences. The fundamental postulate of French structuralism is that the symbolic transcends the imaginary. The symbolic governs the movements or 'play' of the imaginary and is in no way affected in turn by the imaginary order. But in the case we have examined, the Other is produced by the specular game. The agents are guided by a reference point that they themselves have caused to emerge. This kind of 'tangled hierarchy' cannot be grasped within the terms of French structuralism.

We can summarize this more concisely as follows. Starting from an *imaginary* relation of vicariousness, strategic interactions with other agents redoubles this vicariousness until it reaches a fractal infinity. However, the human mind does not have the computational power to handle such infinities in normal cognition, and so it creates an abstract Other—a logical fixed point designed so that "the players no longer look to see what the others are doing; they no longer anticipate each other's thoughts, and each individual only relates to the Other" (1989: 49). Therefore this fractal specularity causes agents to reason in the same way as automata. Game theory shows us that infinite subjectivity is indistinguishable from zero subjectivity. Moreover, these 'symbolic instances' represented by the Big Other take the form of cultural conventions, organizing semiotic interactions into systematic codes. Thus we can say that the symbolic order is in fact a *product* of the imaginary. After providing a wealth of examples ranging from blue-eyed islanders to Rawlsian utilitarianism, Dupuy closes by noting that game theory lets us reconcile a fundamental antinomy of structuralist discourse: namely, its dual interpretation of the symbolic as "signifying convention" (in effect, a social contract) versus as the source around which convention is built, rather than just its effect (1989: 60).

The main drawback of Dupuy's work is that he has only written a handful of papers on philosophy of game theory and then seems to have lost interest, turning instead to focus on technological ethics, in addition to publishing some rather mediocre books denouncing neoliberalism and so on. While a couple of his other papers touch on topics such as utilitarianism, it's truly a pity that we never got to see his interpretation of other legends in game theory such as Harsanyi or Shapley. Also, partly because most of his decision theoretic papers were written so long ago, they're paywalled at best, buried in some obscure conference proceeding at worst; his more mediocre papers, as luck would have it, are openly available, though some of them do touch on CK (e.g. 2004, p.



18). “Common knowledge, common sense” is easily Dupuy’s best paper, and after a great deal of effort I’ve found the download link below. One of Dupuy’s few other papers that deals with Lacan is “Self-reference in Literature” (1989b), which also touches upon Gödel, Borges, Foucault, and various other interesting topics, and is a great follow-up to his CK paper.

There’s also a nice overview of Dupuy’s corpus here that ties in Dupuy’s theme of common knowledge to his broader philosophical interests. For the more mathematically-inclined reader, Jean-Michel Lasry is another French thinker interested in the intersection of psychoanalysis and CK, and he has published a paper in issue 30 of the Lacanian journal *Ornicar* entitled “Le common knowledge” (1984: 7593); he also gives a more formal treatment of CK in Lasry, Morel, & Solimini (1989). Another interesting read is Lacan’s essay “Logical Time and the Assertion of Anticipated Certainty: A New Sophism” (1945), which draws from the notion of CK to outline a notion of time based purely in logic.

- Dupuy, J.-P. (1989). “Common knowledge, common sense.” *Theory & Decision* 27(1), pp. 37-62
- Dupuy, J.-P. (1989b). “Self-reference in Literature.” *Poetics* 18, pp. 491-515
- Dupuy, J.-P. (1989c). “The autonomy of social reality: On the contribution of the theory of systems to the theory of society.” *World Futures* 27(2-4), pp. 153-175

### 3 Kevin Hoover

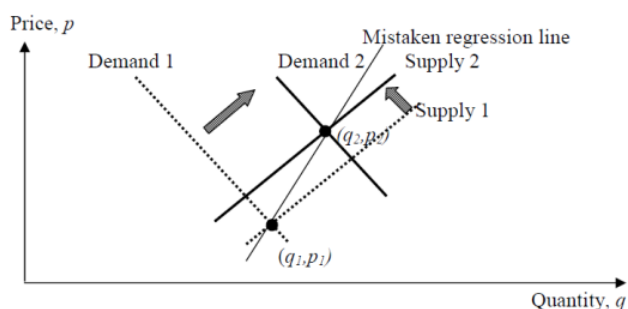
Hoover is definitely one of the more strait-laced members of this list, as well as the most well-known in academia, with joint positions at Duke University as professor of economics and of philosophy. He has also done plenty of rigorous work on econometrics, which he takes as his main philosophical focus. Most people familiar with the literature would be surprised to see Hoover characterized as ‘avant-garde’, but this is because Hoover takes great pains to cloak his (occasionally very subtle) ideas in insipid ‘realist’ rhetoric in order to appear respectable. Upon tweaking his rhetoric, however, we soon see that Hoover lays waste to many of the familiar chestnuts of economic methodology, from the common insistence that individual agents are the elementary ‘units’ of economic theorizing, to econometricians’ prejudice against data-mining. There are many possible ways to present Hoover’s philosophical ideas, but here we will focus on his rejoinders to the Lucas critique, methodological individualism, and the microfoundations project for macroeconomics.

The Lucas critique, published in 1976, is one of the defining events of contemporary economic theory, and most economic models today can be traced to one method or another of combating it. The gist of the Lucas critique is: 1) economists try to predict what a change in economic policy will do based on econometric (statistical) predictions; 2) the policy change (obviously) alters the economy, which in turn alters agents’ behaviour; 3) the parameters that the econometric model was based on also change, meaning that the original econometric model (according to which the policy was designed) becomes irrelevant

to the new economic landscape. Lucas characterizes his own argument in the following terms (1976: 41):

This essay has been devoted to an exposition and elaboration of a single syllogism: given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models.

Hence, any econometric model used for policy evaluation must be based on parameters that are invariant to policy changes, e.g. tastes and technology. That is to say, macroeconomic models ought to be *microfounded*: the aggregates they deal with must be traceable to the actions of individual economic agents in order to be properly invariant. This critique invites comparisons to how, in quantum mechanics, the observer must be considered as a part of the system being observed, and has spawned various quasi-philosophical readings (some of which we'll meet in later sections). While the popular reception of quantum physics and similar scientific themes may make the Lucas critique appear almost trivial to modern ears, it's enough to say that Lucas figured out a way to demonstrate his point using the same math that every other macroeconomist had been using at the time, which is a large part of why it was so revolutionary. Yet, Hoover is largely unfazed by the argument, seeing its chief merit as driving home various points made by econometricians as early as the 1940s, and also strongly opposes the microfoundations programme—an idiosyncratic viewpoint, but one worth taking the time to understand. First, however, we'll need a bit of background.



**Figure 1—Possibility 1: Shift in both Supply and Demand (in  $u_1$  and  $u_2$ )**

In econometrics it has long been known that we cannot separate theories from data, and that any configuration of datapoints is consistent with an infinite number of theories. Econometricians call this the *identification problem*, and its epistemological significance has been known at least since Koopmans (1947), anticipating by several years a similar formulation in philosophy of science called the Duhem-Quine thesis (Quine, 1951). The idea is straightforward to explain. Say we observe that good  $X$  currently costs  $\$P$ —we know that this must be the equilibrium price where the supply demand curves intersect, but we cannot infer solely from this equilibrium price the shape and angle of both curves.

Now say we come back to the store tomorrow and see that good  $X$  is trading at  $\$S$ . The easy thing to do would be to draw a line from the first price ( $P$ ) to the second price ( $S$ ); the problem is that this new price may have resulted from shifts in both the supply curve *and* the demand curve. When we just draw a line from the first price to the second price, we are assuming that the line we drew has stayed constant the whole time, and that only the other curve has shifted. One way to narrow down the infinite number of curves compatible with these two data points is to draw from theory: a simple example is that, in all but exceptional cases, the supply curve shifts upward (the higher the price, the more businesses want to sell) and the demand curve shifts downward (the cheaper it is, the more people want to buy). So in the case depicted above, we can tell that the demand curve has shifted more than the supply curve. There are various other appeals to theory that we can make, and a host of econometric techniques we can use, notably instrumental variables, but the main point here is that it is a non-trivial problem to make sure an econometric model is identified.

Hoover draws from various schematic accounts of causality in order to frame his argument. He defines a cause to be “an Insufficient, Non-redundant member of a set of Unnecessary but Sufficient conditions for the effect” (1994: 66), which is known as the INUS condition. The reason it is *insufficient/unnecessary* is that a cause requires various antecedent conditions in order to work; if the cause was *necessary*, these antecedents wouldn’t matter. Also, there are often multiple ways to bring about a given effect, and because each cause is unnecessary (but *sufficient*) we only have to draw from one of them. Another benefit of this INUS framework is that conditional claims (if  $P$ , then  $Q$ ; also written  $P \rightarrow Q$ ) can be true even if the antecedents ( $P$ ) are not satisfied. Hoover gives the example: “if it were true that the economy was at full employment and the money supply increased by 10%, then it would be true that prices would rise by 10%” (ibid.). While these conditions are not likely to be satisfied, they point to a ‘disposition’ that obtains whether or not the conditions are actualized. Thus it is nonsensical to say something like ‘a diamond is hard enough to scratch glass only if nobody uses it to scratch glass’ ( $\neg P \rightarrow (P \rightarrow Q)$ ), because “[s]uch dispositions necessarily presuppose invariance” (ibid.). Because economic policy changes can bring about the antecedents while falsifying a conditional statement, Hoover reads the Lucas critique as a statement that “existing macroeconomic models do not isolate causal relations—i.e., they do not assert correct conditional propositions” (ibid.).

Further, Hoover has shown that the INUS framework can be formally linked to an alternate representation of causality developed by Herbert Simon, which takes a series of equations and orders them so that “[a] variable ordered ahead of another...is said to cause the other” (1994: 67). Simon proves that “when a system of equations is causally ordered, it is identified econometrically” (ibid.); therefore the Lucas critique can also be read as a claim that existing macroeconomic models are not identified. Hence, according to Hoover, the Lucas critique doesn’t really say anything new, but only gives increased rhetorical force to an idea dating back to the origins of econometrics. He also points to empirical work

suggesting that invariance to policy changes (*superexogeneity*) is more common than not, so that there is no need for economics to drastically change how it operates (2006: 256):

The Lucas critique...can be seen as a possibility theorem: if the economy is structured in a certain way, then aggregate relations will not be invariant to changes in systematic policy. Tests of superexogeneity, based in LSE methods, have been used to test whether invariance fails in the face of policy change (for a survey, see Ericsson & Irons, 1995). That it does not in many cases provides information that the economy is not structured in the way that Lucas contemplates.

It is worth re-emphasizing that this is not at all a common position, and if an economic model is *not* microfounded then its creator has a lot of explaining to do. Part of the reason for the popularity of microfoundations is that it fits nicely into the well-known theme of ‘methodological individualism’, where individual economic agents are seen as the basic ‘units’ of economics. The canonical definition of economics dates as far back as 1935, from a paper by Lionel Robbins: economics is “the science which studies human behavior as a relationship between scarce means which have alternative uses.” Hoover notes (2010: 331): “Such a definition comes close to saying, ‘if it is not microeconomics, it is not economics’.”

\* \* \*

The term ‘methodological individualism’ was coined by the economist Joseph Schumpeter in a book of 1908. Yet, Friedrich Hayek himself notes in his introduction to the English translation that Schumpeter’s book aimed to describe the *Austrian* school of economics, whose methods bear scarcely any resemblance to orthodox economics. In particular, the Austrians reject the use of econometrics, believing that economic principles behaved like laws of nature that could, what’s more, be deduced *a priori*. Meanwhile, after recanting the ideas in his 1908 pamphlet, Schumpeter became one of the strongest advocates of the kind of macroeconomic and econometric tools that the Austrians rejected. Thus it’s an extremely bizarre historical mix-up that methodological individualism was attributed to orthodox economics in the first place, and most likely owes its persistence to its ease of being used as a scapegoat by sociologists and other critics. Yet, due to its commonality, this concept has served as a guide for many working economists, and has undeniably served as the metaphorical foundation for a wide variety of research programmes.

As a macroeconomist, Hoover considers the whole idea to be absurd (1995: 238 & 249):

David Levy (1985) argues that complete methodological individualism is impossible, because, given imperfect information, individual economic actors must make reference to collective entities as a part of their own decision-making processes.

[...] Levy's argument...is even more fundamental than he imagines. In evaluating the future, individuals must form expectations about real prices and real quantities. Independently of the uncertainty of the future, the Cournot problem implies that it is impracticable to solve good-by-good, price-by-price, period-by-period planning problems in all their fine detail. The information on which these are based is fundamentally monetary. Economic actors must use estimates and expectations of the general price level and real interest rates to form any practical assessment of their situation.

[...] Levy's argument demonstrates...that how people theorize about the economy is constitutive of macroeconomic phenomena. Since people cannot theorize about certain sorts of phenomena without appealing to macroeconomic categories...[t]he distinctiveness of the properties at the microeconomic and macroeconomic levels is breached,...because complete characterizations of the microeconomic must include characterizations of the macroeconomic on the part of individual agents.

Thus Hoover argues that macroeconomics should be viewed as *supervenient* on microeconomics: that is, that "if two parallel worlds possessed exactly the same configuration of microeconomic or individual economic elements, they would also possess exactly the same configuration of macroeconomic elements" (1995: 246-7), whereas two identical macroeconomic configurations do not imply identical microeconomic configurations. Hoover spells out in detail how various price indices cannot be precisely decomposed into prices of individual goods, and how common indicators such as real interest rates are deeply tied to these indices so that they, too, are non-decomposable. Thus: "macroeconomic entities...cannot be said therefore to exist only derivatively, despite their supervenience on microeconomic entities" (1995: 250), and ought to be thought of as distinct entities that exist in their own right.

Lastly, Hoover also looks at representative agent models, which represent all the agents in an economy as one or several paradigmatic ones; this is often portrayed as a relatively simple form of microfoundations. While (as he admits) these models are easy to scoff at, Hoover tries to be as charitable as he can, but ultimately finds them to be deeply flawed methodologically. In addition to the unrealistic mathematical assumptions required for the aggregates to have the same properties as the individuals comprising them, Hoover also makes the following point (2010: 345):

The representative agent is held to follow the rule of perfect competition, price-taking, which is justified on the idealizing assumptions that  $n \rightarrow \infty$ ; yet the representative agent is itself an idealization in which  $n \rightarrow 1$ . The representative agent is—inconsistently—simultaneously the whole market and small relative to market.

Hoover is surely the most articulate and rigorous philosopher of macroeconomics active today, and his work is read by many influential economists. For instance,

Milton Friedman heartily accepted Hoover’s description of him as a Marshallian rather than a Walrasian. Likewise, he is one of the few (along with Aris Spanos) willing to engage with the formalism of econometrics to draw out its implications for causality and scientific method. A drawback to his corpus, however, is that he tends to make lots of very subtle distinctions between things that (especially for the non-economist) seem the same. He also has a very odd commitment to realism: he insists that macroeconomic entities exist rather than being the product of theory or some other hyperstitional object—which is largely the consensus view—and this commitment forces him to make a great deal of (rather dull, and not especially convincing) arguments and excursions that he would otherwise not have to make. Still, Hoover opens up a great deal of connections between economic theory and topics at the frontiers of philosophy, such as Judea Pearl’s graph theoretic notation for representing causality, Bernt Stigum’s (morbidly opaque) attempts at axiomatizing econometrics, and Nancy Cartwright’s work in philosophy of statistics. Hoover’s lucid argumentation makes it easy to forget just how difficult the subject matter is to talk about, and anyone who disagrees with Hoover’s conclusions ought to have an equally lucid reason why not.

- Hoover, K. (1994). “Econometrics as Observation: The Lucas Critique & the Nature of Econometric Inference.” *Journal of Economic Methodology* 1(1), pp. 65-80
- Hoover, K. (2006). “The Methodology of Econometrics,” in Mills, T. & Patterson, K. (2006). *Palgrave Handbook of Econometrics*, vol. 1. London: Palgrave-Macmillan, pp. 61-87

## 4 Joshua Epstein

Epstein’s contribution arise from a relatively new computational tool known as agent-based models, which have been the main impetus behind the formation of computational social science. As much of the subject matter of ABMs and economics overlaps, Epstein’s vantage point lets him compare the differences between the two methods, and how their limitations manifest themselves as part of their respective ‘genres’. More than any other ABM advocate, Epstein takes great pains to codify how these separate methods give rise to different forms of reasoning (e.g. deductive vs. inferential), hence his project is deeply philosophical in nature. In brief, an agent-based model uses a set of agents pre-programmed with specific rules and protocols to simulate, piece-by-piece, some macro-level phenomenon. Agent-based models have a wealth of applications: in physics, ‘agents’ can be programmed to behave like elementary particles in order to simulate fluid dynamics; in economics, agents can take the form of individual people in an economy, or of firms within an industry. The idea, then, is to watch how these agents interact within a pre-specified environment, how these actions change that environment, how this new environment influences the actions of agents, how these new actions influence the actions of other agents as well as the environment, and so on and on until a macro-level state has been reached.

Almost always, the macro-level phenomenon could in no way have been predicted from the simple rules initially given to the agents, thus an ABM's results are considered to be *emergent* in a way that normal economic models are unable to emulate. Moreover, ABMs are fully micro-founded by design, sidestepping the Lucas critique entirely.

Of course, if an agent-based modeller's only goal is for their model to arrive at a certain macrostructure, they can just keep tweaking their parameters (a process known as specification-mining) until they arrive at that result, even if the rules given to each agent (called their *microspecification*) are completely absurd. Therefore, an ABM arriving at the result we want doesn't imply that it reflects what is going on in the actual economy, but should be thought of as a "constructive existence proof" (1999: 57, n. 15): we know that it's *possible* for a set of boundedly-rational agents to arrive at this result, whereas normal economics typically identifies the existence of equilibria without spelling out how this result can actually be arrived at. The Artificial Anasazi project nicely exemplifies the deep importance of existence proofs: the project uses ABMs to model—at very high levels of realism—a Native American tribe called the Anasazi, which existed from 800 to 1300AD before abruptly (and mysteriously) going extinct (1999: 44-6). What the researchers want to know is whether this extinction can be explained purely by environmental factors, or whether institutional factors (property, culture, war, disease) must be brought into the picture. The use of ABMs substitutes for the lack of extant data, and consequent unreliability of anthropological theory. If, after entirely sweeping the parameter space of their models, environmental factors are found never to suffice, then anthropologists will have good reason to conclude that there's something they haven't caught on to yet, and new ABMs that integrate environmental and institutional arrangements can be a great help in telling anthropologists what to look for.

Epstein takes these considerations in a philosophical direction by framing ABMs as radically *deductive* in nature, i.e. using general rules of logical validity to reach conclusions from a finite set of premises. Now, in this framework we can treat agents (and their microspecifications) as 'premises', but where the general rules come from is less obvious. For this, Epstein draws from the computational foundations of ABMs: namely, the theory of computation originating with Alan Turing. In particular, he draws from the famous Church-Turing thesis that any computable function can be performed by an abstract model known as a Turing machine. A philosophical corollary of this thesis is that a computer is not simply a collection of hardware (wires, chips, and so on), but instead an abstract idea; thus, humans and even the universe can be construed as computers without this being merely a metaphor. Since the agents in an ABM are just abstract robots carrying out pre-specified programs, it follows as a matter of course that agents can be represented as Turing machines. However, another famous result in computability theory states that "[f]or every Turing machine there is a unique corresponding and equivalent partial recursive function," from which it follows that "for any computation [in agent-based models] there exist equivalent equations (involving recursive functions)" [1999: 51]. And since the

economy is entirely composed of the actions of agents, it follows that the economy can *itself* be represented as a Turing machine. Therefore, the market is a meta-computational assemblage: a computer of computers. Epstein explains why it makes sense to think of the economy quite literally as a computer, and shows the isomorphism between economic ideas and the components of a Turing machine (1999, 49-50):

[M]arkets can be seen as massively parallel spatially distributed computational devices with agents as processing nodes. To say that “the market clears” is to say that this device has completed a computation. Similarly, convergence to social norms, convergence to strategy distributions (in  $n$ -person games), or convergence to stable cultural or settlement patterns, are all social computations in this sense. Minsky’s (1985) famous phrase was “the Society of Mind.” What I’m interested in here is “the Society as Mind,” society as a computational device.

[...] Now, once we say ‘computation’ we think of Turing machines (or, equivalently, of partial recursive functions). In the context of  $n$ -person games, for example, the isomorphism with societies is direct: Initial strategies are tallies on a Turing machine’s input tape; agent interactions function to update the strategies (tallies) and thus represent the machine’s state transition function; an equilibrium is a halting state of the machine; the equilibrium strategy distribution is given by the tape contents in the halting state; and initial strategy distributions that run to equilibrium are languages accepted by the machine.

Part of the point that Epstein wants to make here is that it’s inaccurate to distinguish between agent-based and equation-based models, since any agent-based model can in principle be represented as an equation. Epstein claims that the main accomplishment of ABMs is “*decoupling individual rationality from macroscopic equilibrium*” (1999: 48; his italics). This means, first of all, that equilibrium can be attained without rationality on the part of the agents. If, to borrow a line from Frank Hahn, “there is only one way to be perfectly rational, while there are an infinity of ways to be partially rational” (in Waldrop, 1992: 250-1), then ABMs may open up a lot of doors inaccessible to methods based on rationality. Second, ABMs can demonstrate that individual rationality is not always sufficient to reach equilibrium. The reason for this is that ABMs place agents’ computational capacity, i.e. their ability to solve problems, in the foreground. Some problems require far more computational power to solve than agents have on hand, even if they do everything else right.

The branch of computer science which deals with the computational hardness of problems is known as *complexity theory*, whose main technique is to represent how much time it takes to solve a problem in terms of a mathematical function, and look at how that amount of time scales as the modeller adds more components (e.g. agents in a model, or nodes on a network). Since an economy obviously consists of a great deal of agents, this scaling is crucial to see whether a problem is tractable to solve. If a problem’s scaling function is polynomial, i.e. of the



form  $ax^2 + bx + c$  (possibly including higher-order exponential terms), this usually means that it scales well for practical purposes; e.g. the function  $f(n) = n^2$  equals 100 for  $n = 10$  and 1000 for  $n = 100$ . However, if the scaling function is non-polynomial (NP) then it is typically intractable for high numbers of agents: an example is  $f(n) = 2^n$ , which gives us  $2^{10} = 10^{24}$  and  $2^{100} = 1.26765 \times 10^{30}$ , which is clearly intractable. The reason this is relevant for economics is that a fairly recent result (that Epstein could not have known about) proves that the problem of computing Nash equilibrium is actually in NP; see Daskalakis (2009).

Taking this at face value, it would seem that familiar solution concepts like Nash equilibrium ought to take agents the lifetime of the universe to compute—from which it follows that they cannot represent the world. While ABMs encounter this problem as well—there are some macrostructures that they just can’t reach (cf. Epstein, 2006: 1597)—they at least have the advantage of knowing that any result they reach is definitely computable, since they’ve just computed it! From the point of view of ABMs, therefore, orthodox economic methods are subject to a great deal of problems that ABMs need not worry about, such as economics’ deep rift between micro and macro, as rigorously shown by canonical results such as the Sonnenschein-Mantel-Debreu theorem (Rizvi, 2006). As we go further down the present list, however, we’ll come across new manifestations of these dilemmas as well as increasingly more profound ways of addressing them.

As we have in large part been treating Epstein as a metonym for the ABM programme in general, a satisfactory critique of ABMs from the point of view of economics would require a 10,000-word paper of its own. For now, it’s worthwhile to point out that the computational branch of orthodox economics has largely been tending toward integrating artificial intelligence algorithms (e.g. neural nets, wavelets) into economics, rather than ABMs; this has a lengthy history extending from repeated Prisoners’ Dilemma tournaments to cutting-edge econometric techniques. Yet, Epstein’s papers neglect computational economics, remaining at the level of abstract ideas rather than, say, directly comparing orthodox models to ABMs. Epstein’s papers are remarkably erudite and a pleasure to read, but if he had taken more time to query the specific genre differences of computational economics and ABMs, he might have been able to explain why AI appears to lend itself to the ontology of orthodox economics (or perhaps the other way around), and thus account for economic theory’s present trajectory.

Moreover, since many AI-based econometric models rely on numerical approximations, which usually scale far better than explicitly tracing out the actions of agents, it’s not clear that Epstein’s criticisms remain relevant. Even purely within machine learning, the concept of ‘overfitting’ makes a strong case for radical parsimony in a way reminiscent of Occam’s razor. If we keep specification-mining until we arrive at a specific result we want, then the parameters we end up with will contain a great deal of irrelevant ‘noise’ peculiar to the event we’re focusing on. (They’ll ‘overfit’ our dataset.) Therefore, even if an appropriately tailored ABM can explain a specific event, we can seldom generalize the results to similar events, but must build entirely new models. Conversely, simpler models are designed to eliminate as much noise as possible,

allowing them to be generalized to new contexts, and allowing for serendipitous connections between different contexts in a way that ABMs are not equipped to handle. From the perspective of orthodox economics, then, agent-based models operate at such a high level of granularity that they can't be mapped onto language—nor, perhaps, theory itself.

- Epstein, J. (1999). “Agent-based computational models and generative social science.” *Complexity* 4(5), pp. 41-60
- Epstein, J. (2006). “Remarks on the Foundations of Agent-Based Generative Social Science,” in Tesfatsion, L. & Judd, K. (2006). *Handbook of Computational Economics*, vol. 2. Amsterdam, The Netherlands: Elsevier, ch. 34

## 5 K. Vela Velupillai

Velupillai's core idea is: 1) In order for mathematics to be computable it must be constructive, i.e. given in the form of a finite algorithm; 2) The fundamental mathematical theorems of economics are non-constructive (proofs by contradiction rather than explicitly built from scratch); therefore: 3) Economic formalism cannot describe the world. Velupillai's own project, called Computable Economics, is to develop constructive foundations for economics, and thus bring about an 'algorithmic revolution in economic theory'. His entire corpus consists of variations on that theme, drawing from impressive erudition to base his new formalism on the work of Turing, Gödel, and the foundations of mathematics. Yet, despite his recondite subject matter, most of his prolific (and mostly open-access) array of papers involve no formulas, but are entirely conceptual; his mathematical work is largely confined to his books, notably *Computable Economics* (2000).

To spell out Velupillai's argument, there are two different ways of proving a mathematical theorem. The first is to assume the opposite of what you're trying to prove and derive a contradiction. This method relies on the *law of the excluded middle* in logic, i.e. the axiom ' $P$  or  $\neg P$ ' (i.e. any claim must be true or not true), which entails the principle of double negation, ' $\neg(\neg P) \Leftrightarrow P$ ' (if it's not the case that it is not the case that  $P$ , then it is the case that  $P$ ). The problem with this method is that while we can prove that a lot of claims are true, we don't always have a method of *finding* a solution for a given problem. Constructive proofs, on the other hand, build up a mathematical object from scratch, demonstrating its existence precisely by finding it. It rejects the law of the excluded middle, since even if we can prove 'not- $P$ ', it isn't satisfied until we can actually *construct* the object being talked about. An algorithm is by definition a set of instructions for bringing about a result, hence all algorithms are 'constructive' in this sense. In economics, John Nash used very general tools from topology to prove that any game has at least one Nash equilibrium; however, finding this Nash equilibrium is often very complicated in practice—his proof gives no general algorithm for finding that equilibrium. The tool Nash used was called Brouwer's fixed point theorem, which works like this (Daskalakis, 2009: 8)

Take two identical sheets of graph paper with coordinate systems on them, lay one flat on the table and crumple up (without ripping or tearing) the other one and place it any fashion you want on top of the first so that the crumpled paper does not reach outside the flat one. There will then be at least one point of the crumpled sheet that lies exactly on top of its corresponding point (i.e. the point with the same coordinates) of the flat sheet. The magic that guarantees the existence of this point is topological in nature, and this is precisely what Brouwer’s fixed-point theorem captures. The statement formally is that any continuous map from a *compact* (that is, closed & bounded) and *convex* (that is, without holes) subset of the Euclidean space into itself always has a *fixed point*.

Inspired by this result, Arrow & Debreu used Brouwer’s fixed point theorem in their proof of the existence of general equilibrium, and today every Ph.D student in economics is expected to understand these proofs. The problem, notes Velupillai, is that Brouwer’s fixed point theorem is non-constructive. This is because it relies on the Bolzano-Weierstrass theorem, which uses the law of the excluded middle in an infinitary context, creating, in Velupillai’s words, “an undecidable disjunction” (2012: 9). Therefore the foundations of economics are inherently non-algorithmic—and, he claims, *unconstructifiable*. That is: if Velupillai is right, then it’s mathematically impossible to compute equilibria using existing economic formalism, and we therefore have to toss it out and replace it with constructive foundations: his own project of Computable Economics.

Velupillai’s papers introduce the reader to an entirely new level of abstraction in thinking about why economics is the way it is, and illuminates its myriad interconnections with computability theory in ways that even orthodox computational economics seldom addresses. Computable economics as a research programme has made little impact on the mainstream, however—though from what I gather, Velupillai’s work is fairly well-respected, even in a discipline notorious for burning ‘heterodoxards’ at the stake. While Velupillai self-identifies as heterodox, and is quite dismissive towards orthodox economics, his arguments have appealed to a number of mainstream professors of economics, one example being Stephen Kinsella (who coedited a 2012 essay collection again entitled *Computable Economics*), another being Cassey Lee, who has written some interesting yet accessible essays of his own on algorithmic economics from a slightly different angle. Many of the people sympathetic to Velupillai’s arguments find themselves in the same situation as Axel Leijonhufvud (1993: 1) who acknowledges: “I am particularly grateful to my colleague, Kumaraswamy Velupillai, who has taught me all I know about computability, complexity and related matters.”

His impact is largely a philosophical one, however, as Velupillai (being a theorist) brings little in the way of practical methods for a new economics. Many of the flaws in his work derive from this excessive theoreticism; for example, he states that computable general equilibrium (CGE) models are, in fact, incomputable, since the mathematics behind them relies on an “undecidable disjunction in an infinitary context” (2012: 9). Yet, CGE models are still commonly used in econometrics, notably for climate modelling; Velupillai seems to indicate

that these models and any results derived from them are junk, and ought to be thrown out entirely. Perhaps he is right, but perhaps he is not, and no-one in the Computable economics camp offers any constructive criticism. Other flaws in Velupillai's work include excessive pedantry, little to no engagement with alternative computational methods such as agent-based models (which, as we saw above, bill themselves as avoiding precisely those problems that Velupillai inveighs against), and lastly, taking for granted the incomputability of economic theory in spite of other interpretations that render Velupillai's core idea a non-problem, as we'll see in a later section. Lastly, all of Velupillai's papers have so much in common, and it's very easy to feel like if you've read one of Velupillai's papers, you've read them all. Still, Velupillai is well worth reading, and by supplementing his writings with papers by some of the avant-garde economists described below, one can discern a lot of nuances to his argument that are otherwise easy to miss.

- Velupillai, K. “Towards an Algorithmic Revolution in Economic Theory,” in Zambelli, S. & George, D. (Eds.). (2012). *Nonlinearity, Complexity and Randomness in Economics*. Malden, MA: Wiley, pp. 8-35
- Velupillai, K. (2009). “Uncomputability and undecidability in economic theory.” *Applied Mathematics and Computation* 215(4), pp. 1404-16

## 6 Sheri Markose

Markose is heavily influenced by Velupillai, and adds a unique slant to the themes of computability, agent-based models, and the Lucas critique by viewing them through Gödel's incompleteness theorems. While there are a fair amount of papers hand-wavily citing Gödel as a way to claim we shouldn't bother to use math at all—e.g. Winrich (1984), which starts off well enough but then degenerates into “the poor man's Niklas Luhmann”—Markose does an impressive amount of due diligence, and makes one want to read Gödel just to better understand her arguments. There is some precedent in drawing from Gödel in order to plumb the foundations of economic theory: Binmore (1987) shows that defining game theoretic actants as Turing machines opens the way toward formalisms based in Gödel numbers; Albin (1982) applies Gödel's ‘metalogic’ to standard optimization problems in economics to show that they're subject to undecidability; and Anderlini & Felli (1994) use Gödel numbers and game theory to show that any legal contract depends on the existence of undescribable states of the world, and is thus “endogenously incomplete.” Markose draws from as many of these resources as she can find, integrating them with literature on genetic algorithms, complex systems theory, and automata theory. Yet, within this branch of ‘Gödelian economics’, Markose is the most overtly philosophical (though still very recondite), which is why she makes this list. She also has an impressive amount of accomplishments in applied economics, though we won't deal with those here.

One of Markose's primary themes is the Lucas critique, which we'll recall from earlier is the thesis that policy evaluations, designed on the basis of econo-

metric predictions, will change the parameters that the predictions were based upon in the first place. Unlike Hoover, however, Markose takes the Lucas critique very seriously, and sees it as a stark example of Gödelian reasoning. Note that Gödel's proof was inspired by the ancient Liar Paradox: "This statement is false." Thus, Markose's paper on the Lucas critique develops a game theoretic analogue of Lucas's argument by outlining a 'Liar strategy' that systematically falsifies preannounced events or predictable outcomes, and aims to show that this strategy can bring about uncomputable equilibria. To understand this better, we'll start off with a (lengthy, but very helpful) description of how Gödel's incompleteness theorem works, from Winrich (1984):

In his article "On Formally Undecidable Propositions of Principia Mathematica and Related Systems," Kurt Gödel...illustrated that any system powerful enough to encompass the whole numbers as well as being complete would be inconsistent, and if consistent must be incomplete. He attained these results by showing that metamathematics could be mapped into mathematics itself, eliminating the sharp demarcation of the two and subjecting the system to paradox.

This mapping (or Gödel Numbering) is based upon the theorem that any natural number greater than 1 is either a prime or a unique product of primes. The process of mapping an axiomatic system into arithmetic is quite involved but the overall mechanism can be outlined as follows: First, assign a number to each symbol in the system. Second, convert every formula in the system to a unique number; for instance, assign 1 to  $\in$ , 3 to  $a$ , and 5 to  $P$ , then for the formula  $a \in P$  we would have the sequence of numbers 3,1,5. Now assign to this formula a unique natural number by using the theorem of primes. Gödel's method would produce the number  $2^3 * 3^1 * 5^5 = 75,000$  by using the first three primes. Whenever we run across the number 75,000 it can be converted into a product of primes in only one way, that is,  $2^3 * 3^1 * 5^5$ , which gives us the sequence of numbers 3,1,5, which is the formula  $a \in P$ . In a similar manner, a proof could be converted into a unique product of primes. So, from the number of a proof, we could reconstruct the formulas of the proof. Gödel then converted metamathematical statements into arithmetic. By converting statements about arithmetic into arithmetic the self-referential nature of formulation becomes clear: It is like a snake swallowing its own tail.

Gödel then showed how to construct a statement  $G$  that says that the statement with the Gödel number  $k$  is not provable. But,  $G$  has the Gödel number  $k$ , so  $G$  says of itself that it is not provable. Like Eubulides's Liar, Gödel produced a 'mathematical liar' that asserts: "This statement is unprovable." If the statement is true it is not provable, and if it is not provable it is true. Hence, the statement is true if and only if it is provable. The formal system to which this assertion belongs is consistent only if it is incomplete.

After establishing his ‘mathematical liar’, Gödel then showed that these results apply to any system that can under some scheme be mapped into arithmetic. Closed systems maintain their consistency at the price of completeness.

It is also worth mentioning that Gödel’s result can be viewed as an uncomputable fixed point, and Markose takes advantage of this property to map it onto the notion of economic equilibrium. Recalling that game theoretic agents can be represented as Turing machines, it follows that for this Gödelian equilibrium to be truly uncomputable, it must be incapable of being calculated by a Turing machine. Markose glosses this in intuitive terms by claiming that the only proper response to the Liar strategy is what she calls a ‘surprise strategy’. To illustrate, during the financial crisis the Federal Reserve was very secretive about what its policy measures would be; they knew that if they announced their policies ahead of time, people would just behave strategically in order to profit from the policy, and this behaviour would dull or even annul the policy’s effects—thus they were forced to catch people off guard. In normal game theory, by contrast, action sets are ‘given’ from the start, containing a set number of possible strategies that do not change. A surprise strategy is one that “generat[es] a new action rule not previously within the given action sets—which may be difficult if not impossible to model outside the framework of recursion function theory” (2003: 3).

Markose illustrates the idea of surprise strategies using the Catch-22 paradox and a variation on Brian Arthur’s El Farol Bar problem. To tie this in with some of our earlier sections, her formalism is connected to the notion of common knowledge in an unexpected way: “In the absence of computable fixed points, rational agents even with the same information must agree to disagree” (2005: 177). The latter refers to a proof by Aumann (Dupuy’s main influence) that for Bayesian agents with common priors to ‘agree to disagree’ is inherently irrational; in the exceptional circumstance where equilibrium is not practically computable, sometimes the irrational becomes rational.

The connection to Hoover’s analysis is far more tenuous, and relies on Lucas’s own interpretation of his critique: Lucas thought that in order for keep the economy stable, the Federal Reserve ought to 1) be transparent in its policymaking and 2) act according to simple rules that economic agents can easily predict; only in this way will the prediction function for the Fed’s actions be econometrically identifiable (Markose, 2003: 5). Yet, Markose concludes that it is not possible for this prediction function to be identified, and that this impossibility results directly from Gödel’s theorem. Perhaps surprisingly, Markose’s recondite analysis therefore yields a straightforward policy conclusion: transparent policymaking only works if people cannot adopt a Liar strategy. Due to the abstractness of her framework, this result can be generalized to any mechanism or market design that permits Liar strategies (2003: 14):

For any rule that...involves predictable outcomes for asset prices or quantity positions, a Liar strategy against it furnishes us with conditions under which a transparent rule will fail to be a Nash equilibrium strategy.

The agent applying the transparent rule will be used as a money pump and/or have his desired objectives contravened. No rational agent can be assumed to operate such a transparent rule and no institution based on it will survive unless the transparency of the rule is attenuated. [Thus,] precommitment to transparent rules that are not Liar-proof in order to vitiate surprise equilibria is both an illogical and strategically irrational proposition despite the facile allure of determinacy.

In a 2005 paper, Markose goes on to link the idea of ‘surprises’ to the notion of markets as complex adaptive systems. Part of her agenda is to advocate for the use of genetic algorithms and agent-based models, which she finds interesting because they are able to use trial-and-error heuristics to arrive at outcomes that otherwise cannot be reached in polynomial time (2005: 160). Part of the appeal of complex systems is that they can be directly linked to computability theory, via the agents that make up the system. It is fairly common to represent economic agents not as Turing machines, but as less powerful abstract machines known as finite automata. Finite automata are effectively ‘nested’ in Turing machines, as the latter “can simulate the operation of machines of lesser computational capacity” (Mirowski, 2007: 226). The most obvious difference is that automata have no working memory, whereas Turing machines are posited as having an infinite memory (or: infinite tape).

The capacities of these different types of automata are gauged by what is known as the Chomsky hierarchy, which arranges types of language (from machine language to human language to birdsong) according to the computational power needed to replicate them. Wolfram (2002) takes this hierarchy and shows that it also applies to complex systems, where the computational power of the agents composing the system allow the system to reach more or less complex states. After presenting his results, Wolfram goes on to conjecture that “only agents with the full powers of Turing machines capable of simulating other Turing machines, which Wolfram calls *computational universality*, can produce...irregular innovation-based structure-changing dynamics associated with evolutionary biology and capitalist growth.” (Markose, 2005: 167). Markose’s paper is filled with ideas, and mostly lacks any linear narrative connecting them, but suffice it to say that she draws from a wealth of examples and a variety of results from various formal sciences to identify situations where the traditional logic of economics does not hold, and where a Gödelian approach is more fitting. Her conclusion, broader than her previous one, is that “problems involving strongly self-referential global/system-wide mappings, like an endogenously determined price or reward system, or those that involve coevolutionary contrarian or hostile agents, [are] impossible to solve by deductive means” (Markose, 2005: 188).

The main drawback to Markose’s work is that she’s not a very good writer, and most of her philosophical papers are abysmally formatted—enough that I’ve felt the need to typeset her “Liar Surprise and Strategy” paper in LaTeX and (lightly) edit for better readability. Also, while Markose is not a Velupillai clone by any stretch, she tends to inherit a lot of his flaws (mostly in connection to

their shared heterodoxy), and often interprets her own results in a conceptually facile way when more nuance—and direct engagement with the philosophical end of things—seems called for. Readers who want to explore Markose’s 2005 paper will find it much easier if they first read Cassey Lee’s “Emergence and universal computation” (2004), which goes over some of the same material in a much more accessible way, plus actually has a coherent argument.

- Markose, S. (2005). “Computability and Evolutionary Complexity: Markets As Complex Adaptive Systems.” *Economic Journal* 115(504), pp. F159-92.
- Markose, S. (2003). “The Liar Strategy and Surprise: So What is the Lucas Critique? A New Perspective from Computability Theory.” Paper given at Econ. Dept. Seminar, Carnegie Mellon.

## 7 Rohit Parikh

Parikh is a professor of logic at CUNY, and his contribution to economics is a project known as ‘Social Software’, which looks at societies by means of algorithmic and semantic tools. That is to say, Social Software treats social institutions and mechanisms as software, which lets him apply the whole gamut of tools in computer science to analyze economic problems. The main philosophical thread of this project is that Parikh takes it from a heuristic metaphor to a conceptual isomorphism: just as the economy *is* a computer (cf. Epstein above), so economics is computer science (or, at least, it can be!). In his 2002 ‘manifesto’, he names three fields in CS theory and traces out their parallels with economics. First off (2002: 189-90):

*Concurrency theory*, and *Distributed computing*...analyze the behaviour of several computing processes acting together and...ensure that different processes sharing some resources do not frustrate each other’s purposes and do share information so that when a process needs to act, it knows the facts that it needs to decide which action to take. [...]

Game theory is an analogue both to concurrency theory when we consider agents acting concurrently and in ignorance of each other’s moves, and to distributed computing, since we also consider situations where the agents are acting in turn, in full knowledge of each other’s moves.

Second, CS theory has developed formal methods to prove *program correctness*. This dates back to C.A.R. Hoare’s (1969) “Axiomatic Basis for Computer Programming,” which represents programs using logical notation; thus, the formalism for one component can be chained to the notation for other components in order to prove various properties about the program as a whole. Over the decades this has drawn from increasingly more complex formalism under the aegis of *abstract interpretation*—to the point where “there now more research effort in logics for computer science than there ever was in traditional logics” (Marek & Nerode, 1994: 2). Parikh (2002: 194) provides an interesting example here: airlines often overbook their flights because it is very likely that someone



will cancel their ticket or not show up, and the airline has calculated that it makes more money providing refunds for overbooked flights than it does leaving empty seats on its flights. The problem is that “the set of promises made by the airline is inconsistent with the physical fact of the plane’s capacity” (ibid.), which therefore requires that the airline’s database make use of a *paraconsistent logic* that can tolerate such inconsistencies. Many ordinary situations likewise draw from non-standard logics, even though they are entirely commonsensical on the part of humans.

Parikh’s economic analogue for proving program correctness is *game semantics*, which was inspired by game theory and is actually formally related to Hoare logic. In essence, it models a logical proof as a game between players that try to prove vs. disprove a given assertion. Pietarinen (2003: 325) gives a lucid description of how this takes place:

The semantic game is played on a model  $M$  consisting of a nonempty domain of individuals and an assignment function from terms of  $L$  to the domain of the model, restricted to free variables of every  $\varphi \in L$ . The Falsifier is trying to falsify the formula (i.e., to show that it is false in  $M$ ) and the Verifier is trying to verify it (i.e., to show that it is true in  $M$ ). The universal quantifier  $\forall$  and conjunction  $\wedge$  prompt a move by the Falsifier, and the existential quantifier  $\exists$  and disjunction  $\vee$  prompt a move by the Verifier. When the players come across negation, they change roles, and the winning conventions will also change. Each move reduces the complexity of a formula and hence an atomic formula is finally reached. The truth-value of an atomic formula determines who wins.

A strategy for any player is a function assigning to each subformula a player, and outputting the result of an application of each rule given the input of the rule. The input can be a quantified variable, a value in a two-element set corresponding to connectives, or an instruction to change roles at a subformula. A winning strategy is a strategy by which a player can make operational choices such that every play results a win for him or her, no matter how the opponent chooses. Finally, a formula  $\varphi$  is true in  $M$  if and only if there exists a winning strategy for the player who started the games as the Verifier, and false in  $M$  if and only if there exists a winning strategy for the player who started the game as the Falsifier. This truth-definition invokes the key notion of strategies, codified in Skolem (choice) functions. For example, if the formula  $S_{xy}$  is atomic, then  $\forall x \exists y S_{xy}$  is true in  $M$  if and only if there exists a one-place function  $f$  such that for any individual chosen by the Falsifier (say,  $a$ ) in the domain of  $M$ , the atomic  $S_a f(a)$  is true.

Similar ideas can be found in Wittgenstein, Peirce, and even Leibniz, who framed epsilon-delta proofs in calculus as a game between  $\epsilon$  and  $\delta$  (ibid, 318). The notion of ‘games’ turns out to be very helpful for proofs that are inaccessible via normal methods; specifically, “their non-compositional nature...evaluate[s] logical statements in an outside-in, top-down manner, starting from the outermost

ingredient and ending when atomic components are reached, [letting] the contextual information...reach inner evaluation points.” (ibid.). Parikh’s pet example is cake-cutting algorithms: while “one person cuts, the other person chooses” works effectively in practice for two siblings, things get much more complicated in the  $n$ -person case. Yet, Parikh was able to use game semantics to provide just such a general algorithm, which is one of the few adequate solutions yet devised.

However, a significant component of computer science takes the form of “analy[zing] the efficiency of programs in terms of resources utilized and the amount of time taken” (2002: 189)—that is, analyzing the algorithmic content of computer programs to see if they can be made better. The problem is that: “Since the structure of games in terms of simpler subgames is not analyzed very much in economics, there is no genuine analogue to the analysis of algorithms” (p. 190), hence a more general theory is required. This is precisely the niche that Social Software aims to carve out for itself in economic theory.

Most of this discussion has been quite abstract, but in fact Parikh makes a point of using everyday examples to illustrate his points, so as to stress the wide applicability of his project. The most accessible introduction to his work is “What is Social Software?” (coauthored with van Eijck), which takes the form of a Platonic dialogue. On the philosophical end, he actively draws from analytic philosophers such as Wittgenstein and Grice, which he charmingly blends with ancient Indian philosophy. His project as a whole demonstrates that even the upper echelons of formal logic can be framed so as to have deeply practical implications, and invites the economically-inclined reader to experiment with a host of formal tools that have largely been neglected even in economics’ furthest algorithmic frontiers.

The main flaw of Parikh’s project is that although he and his colleagues bill Social Software as an independent discipline, it is essentially algorithmic mechanism design + game semantics, but with a more CS-based rhetoric. Algorithmic mechanism design has up to now been very commercial (*à la* Silicon Valley), and has been a staple for online auctions such as Google’s AdWords. While several dissertations (e.g. Pacuit, 2005) have been written on Social Software, it would be far more fruitful to integrate this mode of thinking with the mechanism design literature. Still, Parikh’s project provides a fascinating new angle on what  $\text{CS} \cap \text{Economics}$  actually entails, and has the potential to make game semantics into a formidable tool for applied work.

- Parikh, R. (2002). “Social Software.” *Synthese* 132(3), pp. 187-211
- Parikh, R. (2009). “Knowledge and Structure in Social Algorithms,” in Berghammer, R, Jaoua, A., Möller, B. (2009). *Lecture Notes in Computer Science*, vol. 5827. Heidelberg: Springer-Verlag, pp. 1-12 [extended abstract]

## 8 Fernando Tohmé

Tohmé is a mathematical economist from Argentina, much of whose work involves experimenting with new mathematical formalisms. For example, he’s written multiple papers on alternate set theories as a means of overcoming formal

limitations within economic theory, which in some ways is reminiscent of Badiou's project. On a more directly philosophical front, he has collaborated with Rocco Gangale—a Laruellean most well-known for his reader's guide to Laruelle's *Philosophies of Difference*—in order to formalize abductive reasoning using category theory. He also provides an underexplored foil to Velupillai's thesis by using the concept of 'oracles' (or: Turing machines with 'advice') to make it into a non-problem—arguably salvaging the entire neoclassical paradigm. This latter argument actually helped me out of a tough spot in my (otherwise dismal) undergrad thesis, so let's start with that and go deeper into Tohmé's mathematical work as we go along.

To recap: computer scientists have proved that computing equilibrium—in the form of Brouwer fixed points—is an NP problem, meaning that it should take agents—in the form of Turing machines—the lifetime of the universe to compute. Tohmé's 2003 paper first demonstrates the incomputability of Brouwer fixed points, but then evokes the crucial concept of an *oracle*, which dates back to Turing's 1939 doctoral dissertation. An oracle amplifies the power of Turing machines in the following way (Tohmé, 2003: 6):

An oracle for a function  $\alpha : \mathbb{N} \rightarrow \mathbb{N}$  is a device that, given  $x \in \mathbb{N}$  responds with the value  $\alpha(x)$ . So, a Turing machine that requires as an intermediate step of its computation the value of an arbitrary function over  $\mathbb{N}$  can be empowered allowing it to consult an oracle for that function.

In other words, an oracle, as its name implies, is a 'black box' that is called upon whenever a Turing machine is unable to yield an answer; another (slightly weaker) way of putting this is that the Turing machine receives external 'advice' from an entity not subject to the same limitations, just as the ancient Greeks might have consulted the Oracle of Delphi. Turing himself wrote "We shall not go any further into the nature of this oracle apart from saying that it cannot be a machine" (1939: 173). While 'oracle' sounds esoteric, it's actually a common notion in computability theory, and Van Leeuwen & Wiedermann (2001: 16) have a lovely paper describing how oracle computing makes sense of various computational processes that appear at first glance to go beyond the limits of Turing machines:

In recursion theory and in complexity theory, several computational models are known that do not obey the Church-Turing thesis. As examples, oracle Turing machines, non-uniform computational models such as infinite circuit families, models computing with real numbers (such as the so-called BSS model), certain physical systems, and variants of neural and neuroidal networks can be mentioned. Yet none of these is seen as violating the Church-Turing thesis. This is because none of them fits the concept of a *finitely describable algorithm* that can be mechanically applied to data of arbitrary and potentially unbounded size. For instance, in oracle Turing machines the oracle presents the part of the machine that in general has no finite description. The same holds for

the infinite families of (non-uniform) circuits, for the real numbers operated on by the BSS machine or analog neural nets. So far no finite physical device was found that could serve as a source of the respective potentially unbounded (non-uniform) information.

Tohmé’s innovation is to define as oracles the heuristics by which economic agents make decisions—precisely the subject matter of behavioural economics (2003: 8-9). In a sense, this is completely obvious, and this is why it is such an effective rebuttal to Velupillai. Velupillai, framing economic agents as Turing machines, assumes that for them to compute equilibrium they have to take into account *everything*: the utility function is fundamentally *representational*, and must replicate the same functional forms that comprise the actual economy. Tohmé’s account breaks from the notion of direct representation: agents compute as far as they are able (which may include bounded rationality, which is formalizable by limiting the number of states to which the Turing machine may transition), and then are given ‘advice’ in the form of non-linguistic nudges from things like cultural norms, evolutionary instincts, and Schelling points more generally.

Therefore, contrary to Velupillai, economics is not in danger even if its formalism is not directly computable, because it is not intended as a map of the world. Non-constructive tools such as Brouwer fixed points provide a general way of *understanding* the limiting tendencies by which equilibrium takes place, while throwing out extraneous situational details that vary case-by-case. For a somewhat abstruse analogue in computer science, Solomonoff’s theory of inductive inference (which has many applications for predictive models in machine learning) relies on a formalism known as *algorithmic probability* that is “the *only* induction method known to be complete” (Solomonoff, 2009: 5) and which is “guaranteed to discover any describable regularities in a body of data, using a relatively small sample of the data” (ibid., abstract). While the completeness of algorithmic probability is fundamentally linked to its incomputability, this incomputability—paradoxically enough—becomes an *asset*, in that it provides an ideal limit for approximation. In the same way, Velupillai’s entire argument, while incredibly inspiring and intellectually stimulating, is thus a non-problem.

This actually segues quite nicely with Tohmé’s mathematical work on abductive reasoning. *Deduction* uses general rules of logical validity to reach conclusions from finite set of premises; *induction* generalizes from specific facts to general trends, via appeals to asymptotic properties. Conversely, what logician C.S. Peirce called **abductive reasoning** is a process of inferring directly from observation to hypothesis (Kapitan, 1992). While this is laid out in very abstract terminology, the best exemplar of abduction is, in fact, Sherlock Holmes. In avant-garde circles of continental philosophy this term has been used by people like Reza Negarestani and Fernando Zalamea, and is often used (somewhat hand-wavily) to characterize algorithmic reasoning processes. Negarestani (2013: 14, fn. 9) gives a more elaborate definition of abductive reasoning:

**Abductive inference**, or abduction, was first expounded by Charles Sanders Peirce as a form of creative guessing or hypothetical inference

which uses a multimodal & synthetic form of reasoning to dynamically expand its capacities. While abductive inference is divided into different types, all are non-monotonic, dynamic, and non-formal. They also involve construction and manipulation, the deployment of complex heuristic strategies, and non-explanatory forms of hypothesis generation. Abductive reasoning is an essential part of the logic of discovery, epistemic encounters with anomalies and dynamic systems, creative experimentation, and action and understanding in situations where both material resources and epistemic cues are limited or should be kept to a minimum.

Furthermore, according to Peirce, pragmatism “is nothing else than...the logic of abduction” (in Burks, 1946: 306). Digging below the surface, however, it’s actually quite difficult to put one’s finger on what abduction entails, and there is a whole literature (e.g. Kapitan, 1992) ultimately concluding that abduction is simply a dressed-up form of induction. Hence we can see why this question would be interesting to both a philosopher like Rocco Gangle (one of Tohmé’s coauthors, along with mathematician Gianluca Caterina) and to someone interested in the abstract tools of mathematical economics: as well as being a challenge even to formalize clearly, success at this would open up a wealth of applications, such as integrating machine learning processes directly into economic formalism. To solve this problem, a tool that Tohmé et al. have drawn quite heavily upon is *category theory*, characterized as “the mathematics of mathematics” (Cheng, 2015). However, let’s hold off from spelling out the implications of category theory for economics until our next, and last, avant-garde economist.

While I’m not yet qualified to comment on the mathematics behind Tohmé’s work, its main ‘flaw’ is of course its inaccessibility. While Tohmé’s subject matter is often very philosophically-tinged, his writing is aimed at other mathematical economists, and one is forced to read very hard between the lines in order to glean what his results *mean*. That said, his papers are far more accessible than those of people like Alain A. Lewis, whose results are equally profound but delivered almost entirely in symbols without any ensuing explanation (which is why Lewis didn’t make the present list). Tohmé’s most accessible philosophical paper is his one on Rolf Mantel (of Sonnenschein-Mantel-Debreu fame) which talks about computable general equilibrium models. His 2003 paper on oracles is a bit dense at times, but for readers familiar with basic economic notation it should be easy enough to follow. The rest of his corpus is quite varied, touching upon fuzzy sets, semiotics, non-ZFC set theories, and Cohen’s forcing in the context of game theory. It would be interesting to contrast Badiou’s ideas with Tohmé’s approach, and in any case it should be very interesting to see how far his collaboration with Rocco Gangle ends up going.

- Tohmé, F. (2003). “Oracles and economic behavior.” *The Journal of Management & Economics* 7, pp. 1-10
- Tohmé, F. (2006). “Rolf Mantel and the Computability of General Equilibria: On the Origins of the Sonnenschein-Mantel-Debreu Theorem.” *History of Political Economy* 38(S1), pp. 213-227

## 9 Viktor Winschel

Winschel is at the end of this list because he draws from just about all the different ideas sketched above. The framework he uses to integrate them is based in perhaps the highest level of abstraction the human mind has ever reached: **category theory**, known as ‘the mathematics of mathematics’, with which he aims to construct new categorical foundations for economics. Winschel’s aim is nothing less than to rhizomatically link economics to the furthest frontiers of theoretical computer science, formal logic, and mathematical research itself. It’s scarcely possible to be more avant-garde than Winschel; what makes him a philosopher is his focus on the essential role of recursion in making sense of the economy. Reminiscent of the system-theoretic sociology of Niklas Luhmann, Winschel’s project can be characterized as formulating an ‘economics of economics’. Winschel’s research has covered nonlinear econometrics, optimal currency areas, and recasting the foundations of game theory and decision theory in category theoretic notation. Perhaps most surprisingly, his recondite formalism lends itself perfectly to the functional programming language Haskell, meaning that far from being castles in the sky, his ideas are directly implementable in Haskell code (itself based on category theory) and thus directly usable in applications by anyone with sufficient imagination.

Category theory has received a fair bit of attention of late in Continental philosophy due to Alain Badiou’s *Logics of Worlds* & Fernando Zalamea’s *Synthetic Philosophy of Contemporary Mathematics*. It has also received some mainstream attention as a result of Eugenia Cheng’s *How to Bake  $\pi$ : An Edible Exploration of the Mathematics of Mathematics*, written for people who haven’t done any math since high school, as well as the recent death of the legendary Alexandre Grothendieck, which elicited many gorgeous eulogies and bloggers’ attempts at distilling his ideas for a lay audience. Mathematicians such as Abramsky (2012) have touted category theory as a new kind of ‘formal philosophy’ capable of brand new forms of abstraction. This need not be limited to analytic philosophy: a rhizome can be formalized as an indiscrete category, in which “there is exactly one arrow from every object to every other object” (Cheng, 2015: 260). But what is it, and what does it have to do with economics? Let’s start with the definition of a category itself, given to us by Cheng (2015: 199):

A *category* in mathematics starts with a set of objects and a set of relationships between them. Now, these relationships are not necessarily symmetric, so we need to change our wording a bit to bring this out. So instead of saying “a relationship between A and B” it would be better to say “relationship from A to B” to emphasize that it only goes one way. In fact, in category theory we sometimes say “arrow from A to B” to emphasize that direction even more, and to remind ourselves of the fact that we can draw helpful pictures of these relationships using arrows. We might also say “morphism” because sometimes these things are more like a way of morphing something into something else, like morphing a donut into a coffee cup.

Now we have to say what rules our relationship must obey.

1. (A bit like transitivity) Given an arrow  $A \xrightarrow{f} B$  and an arrow  $B \xrightarrow{g} C$ , this has to result in a *composite* arrow  $A \xrightarrow{g \circ f} C$ .
2. (A bit like reflexivity) Given any object  $A$  there has to be an ‘identity’ arrow  $A \xrightarrow{I} A$ , which means that for any other arrow  $f \circ I = f$  &  $I \circ f = f$
3. Given three arrows  $A \xrightarrow{f} B$ ,  $B \xrightarrow{g} C$ ,  $C \xrightarrow{h} D$ , we can make composites in various ways, and it all has to obey this rule:

$$(h \circ g) \circ f = h \circ (g \circ f).$$

In short, “Arrows show the relations among objects of a category, and functors show relations among categories” (Beheshti & Sukthankar, 2013: 281); relations among functors also have a name (‘natural transformations’), and so on, providing a handy vocabulary for any sort of formal relation you might want to deal with. The benefit of this kind of notation is that it allows connections between disparate branches of mathematics, such as Grothendieck’s famous work unifying “the discrete world of algebraic varieties and the continuous world of topology” (Jackson, 2004: 1038). It can likewise characterize alternate types of formal systems; for instance: “Logical systems can be represented as categories in which formulas are objects, proofs are arrows, and equality of arrows reflects equality of proofs” (Abramsky, 2012: 15). This is very helpful in fields like computer science that draw heavily from a wide variety of formal systems (various types of math, logic, and syntax), and so has become a standard tool for higher-level theoretical CS such as programming language theory.

Let’s see how this translates into economics. One of Winschel’s projects is to use a part of category theory called *coalgebras* to develop higher-order formal frameworks for game theory. The main advantage of category theory in a microeconomic context is that it “provides an exact notion of modularity and composition” (Beheshti & Sukthankar, 2013: 281). Hence, Winschel & Blumensath’s paper—effectively a manifesto for the ‘economics of economics’—proposes that their coalgebraic framework can solve four major problems in economic theory. Their framework can [1] provide a general method to “compose simple games into complicated ones, sequentially or in parallel” (Blumensath & Winschel, 2013: 4), as well as [2] give “a formal account of aggregation of games” (ibid.). Categorical formalism is, for similar reasons, far better at formalizing the notion of ‘emergence’, which as we saw above is the *raison d’être* for agent-based models; therefore their framework [3] allows “a formal semantics for agent-based models” (ibid.) that can be integrated with the orthodox literature on heterogeneous agents. Lastly, while the math used in economics tends to be continuous (calculus, topology, measure theory), the boundary between continuous and discrete mathematics is not a problem for category theory; one application is thus to [4] “generalize network economics,” since institutional economics places paramount importance on “[t]he interaction of the network structure and the games played in the network” (ibid., 5).

Perhaps the most important property of category theory is that it was designed precisely in order to unify disparate branches of mathematics, and so provides the machinery to integrate the coalgebra and formal logic involved in microeconomics with the algebraic geometry (e.g. topological fixed point theorems) used in macroeconomics. Moreover, since programming languages such as Haskell are explicitly designed according to category theoretic ideas, coding becomes a form of theorizing, and theorizing a form of coding. Philosopher Paul Humphreys (2009)—another honorable mention who didn’t quite make it onto the list—notes that the more computational economics leverages numerical approximation techniques, the more the distinction between theoretical models and the code used to fabricate them becomes erased. Winschel takes this deconstruction perhaps as far as it can go, where the highest echelons of economic theory fuse into the numerical legwork behind applied projects.

Much like Parikh’s social software, Winschel blurs the lines between economics and computer science itself: “My motto right now is that theories are code and formal methods of computer science can be used to analyse their properties” (2013: 54). Winschel also uses formalisms from computer science such as type theory and the lambda calculus, which he notes are seldom used even in computational economics. Like Parikh, Winschel hopes to use CS as a semantics for economic theories, but Winschel emphasizes their observational equivalence with economic formalism (rather than using CS-based ideas as tools like Parikh), e.g. drawing parallels between numerical approximation in economics and the logical methods of “abstract interpretation, which provides a notion of approximation of the semantics of programming languages” (Blumensath & Winschel, 2013: 32).

Winschel’s own favorite use of category theory is its way of expressing reflexive relationships, which the formalism is ideally suited for: “categories are in a sense fractal or hierarchical themselves since the functors form the objects of a category with natural transformations as arrows” (Blumensath & Winschel, 2013: 26). That is to say, while before we had that relations among objects of a category are arrows, relations among categories are functors, and relations among functors are natural transformations, we can shift this down a level in order to treat functors as categories and so on, applying category theory to itself on a lower level. To illustrate how this comes in handy in economics, the rational expectations school of economics holds that “agents in the model should be able to forecast and profit-maximize and utility-maximize as well as the economist...who constructed the model” (Sargent, 1993: 21). This creates a recursive structure that identifies the modeler with what is modeled. We can see how this is quite similar to the Lucas critique, which Winschel glosses as follows (Blumensath & Winschel, 2013: 27; slightly emended):

The Lucas critique can be formulated as the need for the economics of economics to be economics, or for the need to model economic agents as isomorphic to the econometrician, both being inside the modeled systems. The economics of economics studies the production function of economics itself.



By using this formalism we can identify brand new forms of observational equivalence, just as Khan did with nonstandard analysis; we can make sense of the  $n$ -person Russell paradoxes (known as Brandenburger-Keissler paradoxes) that arise in game theoretic belief formation, much like those addressed by Dupuy in Lacanian terms; we can use categories to identify isomorphisms between econometric models (e.g. Kalman filters and control systems) and even construct hybrid models (Beheshti & Sukthankar, 2013), all the while taking account of the Lucas critique in a way that Hoover would approve; we can integrate the agent-based modelling framework proposed by Epstein with orthodox treatments of heterogeneous agents; we can use categorical programming languages such as Haskell to sidestep the problems of computability identified by Velupillai; we can address the reflexive and paradoxical nature of economic phenomena delineated by Markose; we can coalesce the furthest frontiers of theoretical computer science with economic formalism, as advocated by Parikh; and we can use the work of people like Tohmé to integrate philosophical ideas such as abductive reasoning within the semantics of economic models. Thus, in many dimensions, Winschel's work is the most avant-garde you can get in economic theory.

It's hard to identify shortcomings in Winschel's work, but the main one that comes to mind is that he sometimes (unintentionally) frames his project as new foundations for economics, whereas Abramsky notes that "What category theory offers is an alternative to *foundational schemes in the traditional sense themselves*" (2012: 12), which is in many ways one of the most profound aspects of categorical formalism. Also, while Winschel expresses an interest in continental philosophy, it's regrettable that he doesn't engage with formally-inclined philosophers such as Badiou, who for example defines the 'event' as a set containing itself.

Winschel's 2013 interview provides an excellent overview of the scope of Winschel's project, though in its original version the editor didn't do his job and the English is very difficult to read. I've therefore emended the text for better readability, and the reader can download this new version below in pdf format. As I said above, Winschel's paper with Blumensath reads very much like a manifesto, and the reader will be safe skipping the formal parts in order to focus on the (*very* extensive) conceptual parts. While these papers will doubtless be quite difficult to follow for non-economists, they are filled with brilliant and original ideas from beginning to end, and are well worth the effort.

- Winschel, V. (2013). "Interview: Logical Simulations of Economic Phenomena and Computational Economics." *Studia Humana* 2(1), pp. 48-56
- Blumensath, A. & Winschel, V. (2013). "A Coalgebraic Framework for Games in Economics." 19th International Conference on Computing in Economics & Finance.

## TL;DR

- Khan uses literary theory to argue that economic models are (meta-)allegorical in structure.
- Dupuy uses Lacanian psychoanalysis to make sense of ‘common knowledge’ in game theory.
- Hoover wants to see what econometrics has to say about philosophical concepts like causality.
- Epstein argues that agent-based models involve a different kind of logic than normal economics.
- Velupillai insists that the math behind economic theory can’t be translated into algorithms, and so we should replace the foundations of economics with kinds of math that can be.
- Markose uses Gödel’s incompleteness theorem to identify self-referential economic scenarios where the traditional logic of economics doesn’t hold.
- Parikh’s project of social software looks at societies by means of algorithmic & semantic tools.
- Tohmé treats the cultural and behavioural elements of society as patching up the holes in what computers otherwise can’t calculate.
- Winschel uses category theory (“the mathematics of mathematics”) to integrate different areas of economic theory and to make sense of reflexive economic structures.