Taking stock of complexity economics: a comment

Ricardo Hausmann
Center for International Development at Harvard University
Harvard Kennedy School
Santa Fe Institute

It is an honor and a pleasure to speak at this important event and I thank the organizers for inviting me and to Eric Beinhocker and Doyne Farmer for thinking of me as part of this panel.

I am supposed to speak about Complexity Economics and the issues this theoretical approach is able to illuminate. Defined as such, it appears as if I should discuss methodology and I am reminded of Paul Krugman's quote: "It is said that those who can, do, while those who cannot, discuss methodology". So, having a discussion centered on methodology may give the impression that the panelists in this session are unable to make scientific progress.

I am happy to concede that it is easy to fall into true but rather useless methodological discussions. Making fun of the simplifying assumptions of neo-classical economics is surely one of them. We all know that life is much more complex than that. But simplification is at the core of the scientific process, including those that are based on complexity or network science. It is only by focusing on the important and disregarding the less relevant aspects of reality that understanding can emerge. This is also true about complexity science. For example, a lot can be said about networks if we disregard information regarding what the nodes and links of the network actually represent: people, products, genes, diseases, etc. So simplification per se is not the issue. Telling mainstream economists that the world is more complex than what they have assumed is unlikely to sound either controversial or interesting. The challenge is to uncover new facts about the world and new understandings that are hard to grasp in one paradigm, but are more easily observable through another approach.

The research questions that have motivated my recent work have to do with the questions of long-run growth and development. What underpins long-run growth? Why are some countries rich and others poor? Why do some laggards catch up quickly while others fall behind? More concretely, what caused the Great Divergence: the enormous increase in income differences between rich and poor countries since the early XIX century. Why has this process partially reversed in the last decade or two into something that may well become the "Great Convergence"? Why now and not earlier?

I would much rather discuss these issues than talk about methodology, and let the method speak for itself. But I will indulge myself first by quibbling about the neo-classical approach to these questions, before I "do", rather than discuss methodology.

Quibbles

Two ideas rub me the wrong way. First, is the idea that we fundamentally know the things the world is made of. In economics we seem to think that it is made of capital, labor, human capital productivity and output. How do we know that these are the right descriptors of reality? It is a bit like Aristotle telling us that the world is made out of earth, water, wind and fire, while the heavenly bodies are made out of the quintessence. Contrast this with the assumption that things are made out of atoms and that atoms may differ in nature and may combine into molecules according to yet unclear rules. Note that you do not really need to know what these atoms are. You need to let the world tell you. Mendeleyev figured out the periodic table without knowing what atoms were made of. He had no clue about protons, neutrons and electrons. The actual number of atoms – some 118 according to a recent count – was a bit greater and not as intellectually cute as the 5 substances that Aristotle posited. The idea that their chemical properties recur periodically, but with varying periods, was definitely not something that anybody would have guessed at the time by just thinking about it. It required looking at the world and letting it tell us what it is made of. It would take more than 70 years for the quantum-mechanic foundations of the periodic table to be understood.

In trying to understand the nature of economic reality, we have been much less willing to let the world tell us what it is made of and more willing to believe our theoretical contraptions. We think of capital indistinctly as a set of tools and buildings or as an amount of money, even though many of the things that go into production cannot be purchased by firms with money and cannot be optimized across their alternative uses. We think of human capital as years of schooling, maybe corrected by scores in standardized tests. We call all other things total factor productivity. I will try to show implicitly that these descriptors have moved us away from where the real action is.

The second idea that rubs me the wrong way is the notion of requiring that a macro theory be built on micro-foundations of individual choice under constraints. My complaint is not based on the many critiques that Simon, Tversky, Kahneman and other behavioral economists have been able to level against the traditional view of individual decision-making. It is just that it presumes too much about the relationship between micro and macro behavior. The properties of DNA are not contained in any of its atoms and DNA would have no function as a stand-alone molecule. The way DNA codes for amino-acids and the properties of chains of amino-acids that we call proteins can be explored without going back to the quarks, protons or even the atoms that are involved in the process. By the same token, requiring that the patterns of aggregate behavior be relayed back to individual decision-making is often an obstacle rather than an aid to scientific progress. It forces us to radically simplify the structure of the economy in order to make the aggregation problem tractable, and thus give a role to individual maximization under constraints.

Making it more complicated

So the approach Cesar Hidalgo and I have taken has been to take a Mendeleyev approach. Adrian Wood suggested we call our approach the Scrabble Theory of Production. In this, it is not unrelated, although somewhat different from Stuart Kaufman's grammar model. Essentially, products are like words: they are made with letters. In order to make a word you need to have all the right letters. If you miss a letter, you cannot write the word. Not all combinations of letters are meaningful: that is presumably where the word Scrabble comes from: it looks like a plausible word, but it is just meaningless, except as the name of the game. This is one way of capturing the idea of a rugged landscape: the word "music" means something, but the words musid, musib and usic do not. You can think of letters as productive capabilities of one sort or another. At this stage of the game we need not say more, just as Mendeleyev did not have to say what he meant by atomic number. I will have something more speculative to say about what I think they are later on. But that is not really necessary at this stage.

Think of the set of all possible letters as the alphabet. The alphabet may evolve with the introduction of new letters, not unlike the description that Brian Arthur makes of the incorporation of new scientific principles into the set of things technology can play with.

The set of all possible products is like a dictionary: it is the set of meaningful combinations of letters, which is a very small and capricious subset of the set of all possible letter combinations. In another metaphor, you can think of letters as atoms and products as molecules. Only a small set of all combinations are feasible. There are water molecules, made out of three atoms of two different kinds and things as complex as hemoglobin with over 100,000 atoms.

As I said before, we don't need to specify at this stage what the letters represent. But the words are observable: they are the goods and services we can see in the world. We can also see which countries or localities make which products. In the context of our theory, when we observe a locality making a product it is evidence that the locality has what it takes to make the product, except for the elements that can be shipped around, such as the tradable intermediate inputs.

To make matters simpler, let us assume that it is costless to make copies of the letters you have, so we can disregard the number of exemplars of each letter. Once you have a letter, you can use it as many times as you want. Another simplifying assumption is that the order of the letters does not matter: the words "act" and "cat" are the same. A product is defined by the capabilities that go into it, but not the order in which they are used.

What could such a minimalist set of assumptions tell us about the world? Well, lets see.

First, countries differ in the subset of the alphabet they have. Therefore, they differ in the subset of the dictionary that they can write: countries with a fuller set of letters will be able to make more products. We call *diversification* the number of products that a country can make. Countries with more letters will be more diversified.

Second, words that require more letters will be accessible to fewer countries because only those that have all the relevant letters can make the product. We call *ubiquity* the number of countries that can make a product. Products with more letters will be less ubiquitous.

Third, countries with more letters will be able to make products that require more letters, i.e. products that are less ubiquitous. This means that we should observe that the diversification of countries and the average ubiquity of the products they make should be inversely related, since both of these reflect the number of letters a country has. This is exactly what we find in the data (see Figure 1a and 1b). Here we show which countries export which products, using trade data. Moreover, this phenomenon is not only true about the export of goods. When we repeat this calculation for all types of establishments at the municipal level we find a similar result for the US, India, Chile and Turkey: Figures 2, 3, 4 and 5.

Figure 1a

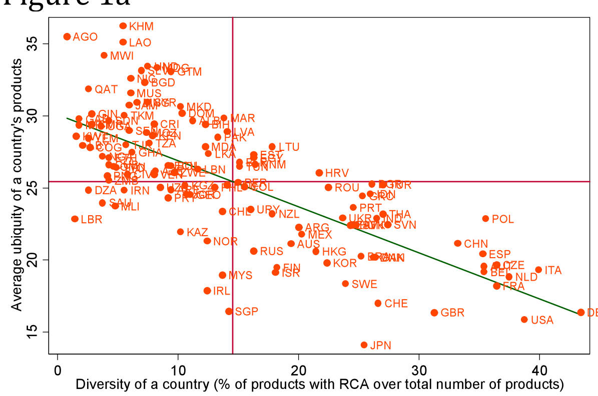


Figure 1b

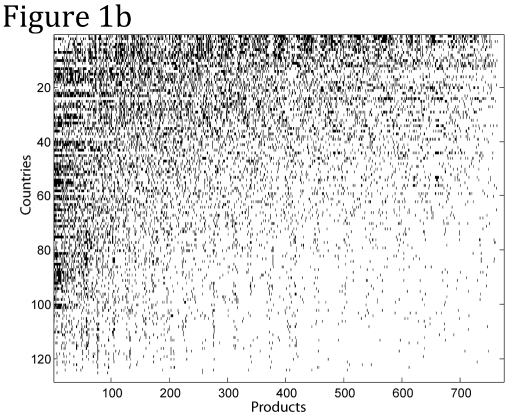


Figure 2 – US

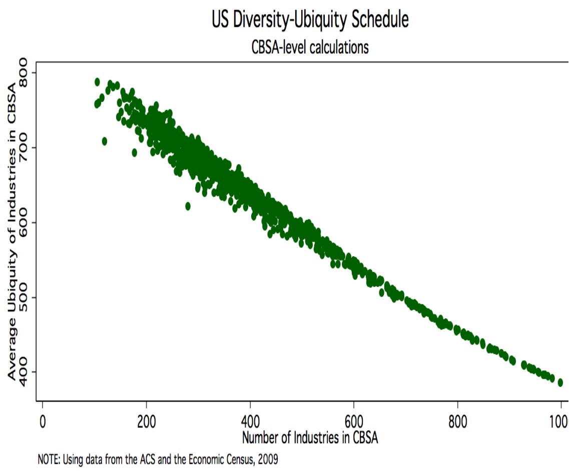


Figure 3 – India

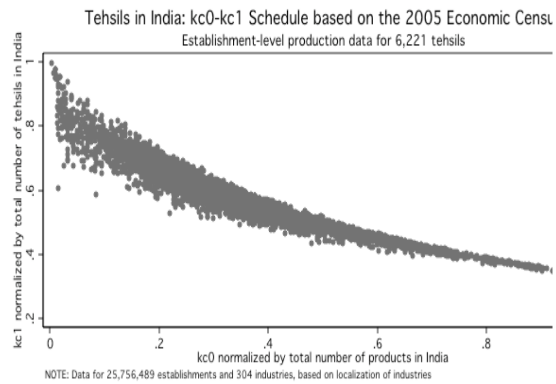


Figure 4 - Chile

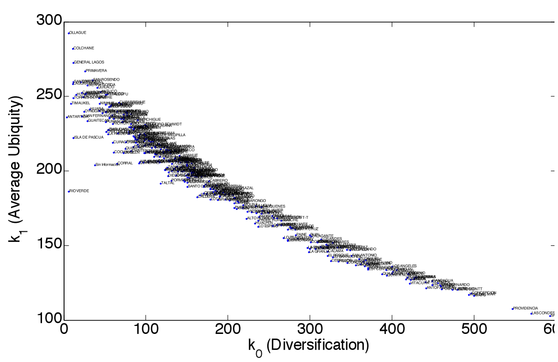
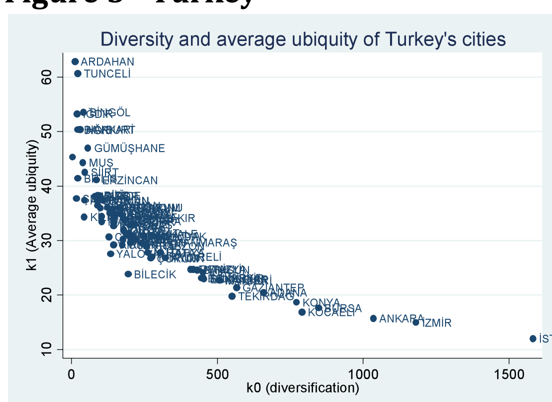


Figure 5 - Turkey



More formally, the model we develop assumes that the part of the world we can readily observe – the matrix of countries and products □□□ is the outcome of two matrixes we cannot readily see: a □□□ matrix that tells us which countries or locations have which capabilities (or letters) and the □□□ matrix which tells us which letters go into which products.

□□□=□□□⊙□□□, (1)

where

□□□⊙□□□=*1* if □□□□□□□□□=□□□□□□ and 0 otherwise. (2)

In fact, we can use this simple model to try to replicate several features of the world:

- the distribution of country diversification,
- the distribution of product ubiquity,
- the negative relationship between the average ubiquity of a country's products and its diversification,
- the distribution of the probabilities that pairs of products are simultaneously present

We do this by just estimating a few parameters:

- how big is the alphabet, *N*
- how sparse is the dictionary relative to the set of all possible letter combinations, parametrized by the probability that a letter goes into a word, *q*
- what percentage of the alphabet countries are estimated to have given how diversified they are and the two previous assumptions.

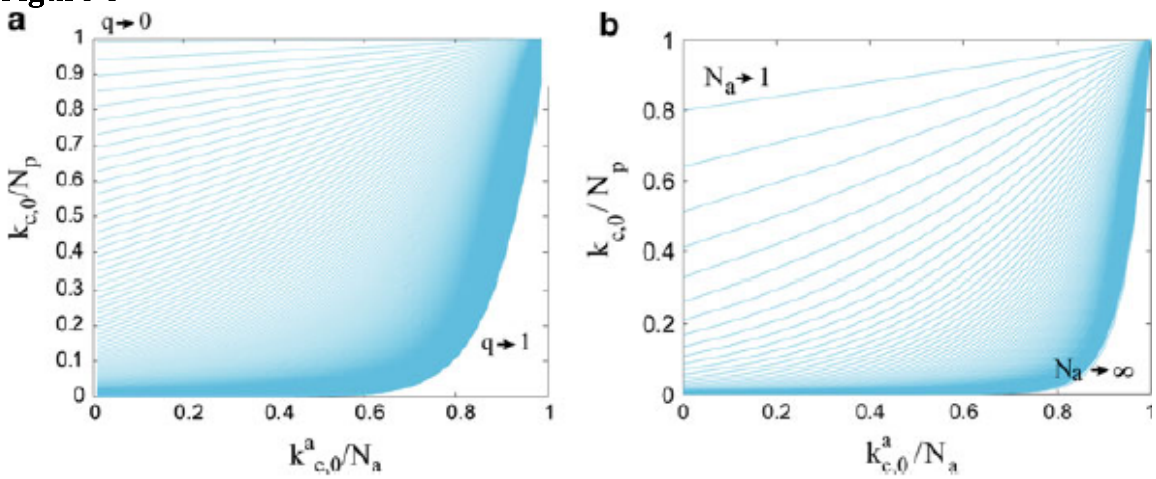
We do this in Hausmann and Hidalgo (2011) where we show how well the model fits these admittedly quirky stylized facts of the world, with the few parameters we estimate.

But the effort comes with a bonus: the estimated model predicts what fraction of the dictionary countries will be able to write with a given proportion of the alphabet. The combinatorial nature of words would predict that the relationship is concave: the proportion of the dictionary that is feasible grows exponentially with the proportion of the alphabet a country has.

□=(1−(□1−□)□

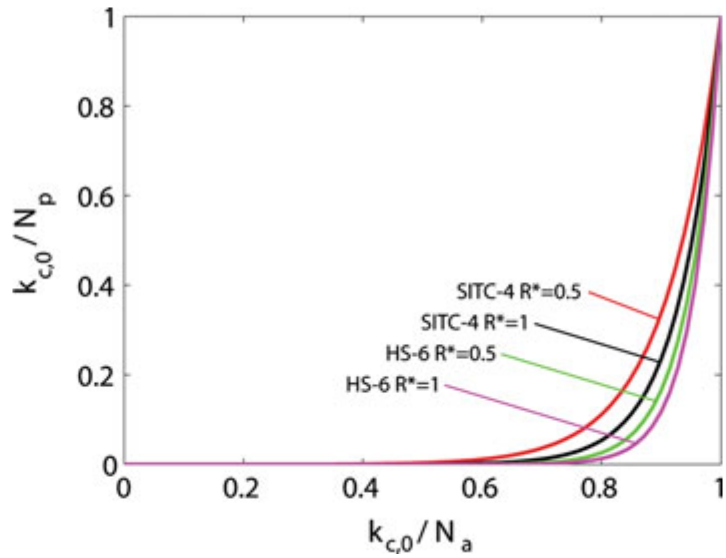
where *d* is the proportion of products that countries are able to make, *a* is the proportion of the alphabet they have, *N* is the size of the alphabet and *q* is the probability that a letter goes into a word. You can also think of it as how long are words, on average. The point is that the concavity of the relationship between *a* and *d* depends on how complex are the products (words) and how big is the alphabet. Figures 6 and 7 illustrate the relationship for different values of *N* and *q* as well as for the estimated parameters of the model when applied at different trade datasets.

Figure 6



Analytically calculated predictions from the Binomial Model used in Hausmann and Hidalgo (2011). **a** Fraction of all products that countries can make as a function of the fraction of all capabilities that countries have for values of q ranging from 0 to 1 and Na = 50. **b** Fraction of all products that countries can make as a function of the fraction of all capabilities that countries have for q = 0.2 and Na ranging from 1 to 1000

Figure 7



Relationship between fraction of capabilities that countries have (*x*-axis) and the fraction of products that countries make (*y*-axis) for the calibrated model values

We call this result the quiescence trap. In a world of many letters (large N) or long words (large q), only those with a very hefty chunk of the alphabet will be able to make many words. Those that start with few letters will not only be able to make almost no words, but will get scant returns from the efforts to accumulate more letters, as these are unlikely to make any additional words.

So, a world with a quiescence trap can lead to the Great Divergence of incomes that characterizes the past two centuries of global development. The countries that had an initial number of letters that was relatively high had many combinations of letters to explore and would get larger benefits from the accumulation of any additional letter. Increasing returns to productive capabilities is our story for the Great Divergence.

To show this we need to incorporate a few additional concepts. We would like to have a good measure of how many letters a country has. Part of this information is captured by the number of products it makes, i.e. its diversification because the larger the set of letters, the greater the number of products the country will be able to make.

But maybe a country's high diversification reflects the fact that it makes many relatively short words, implying that it has fewer letters than you might otherwise think. But if this were the case, then the products the country is able to make will be accessible to many countries, meaning that those products will have high ubiquity. Hence, the ubiquity of a product contains additional information that is relevant to the question of how many letters a country has.

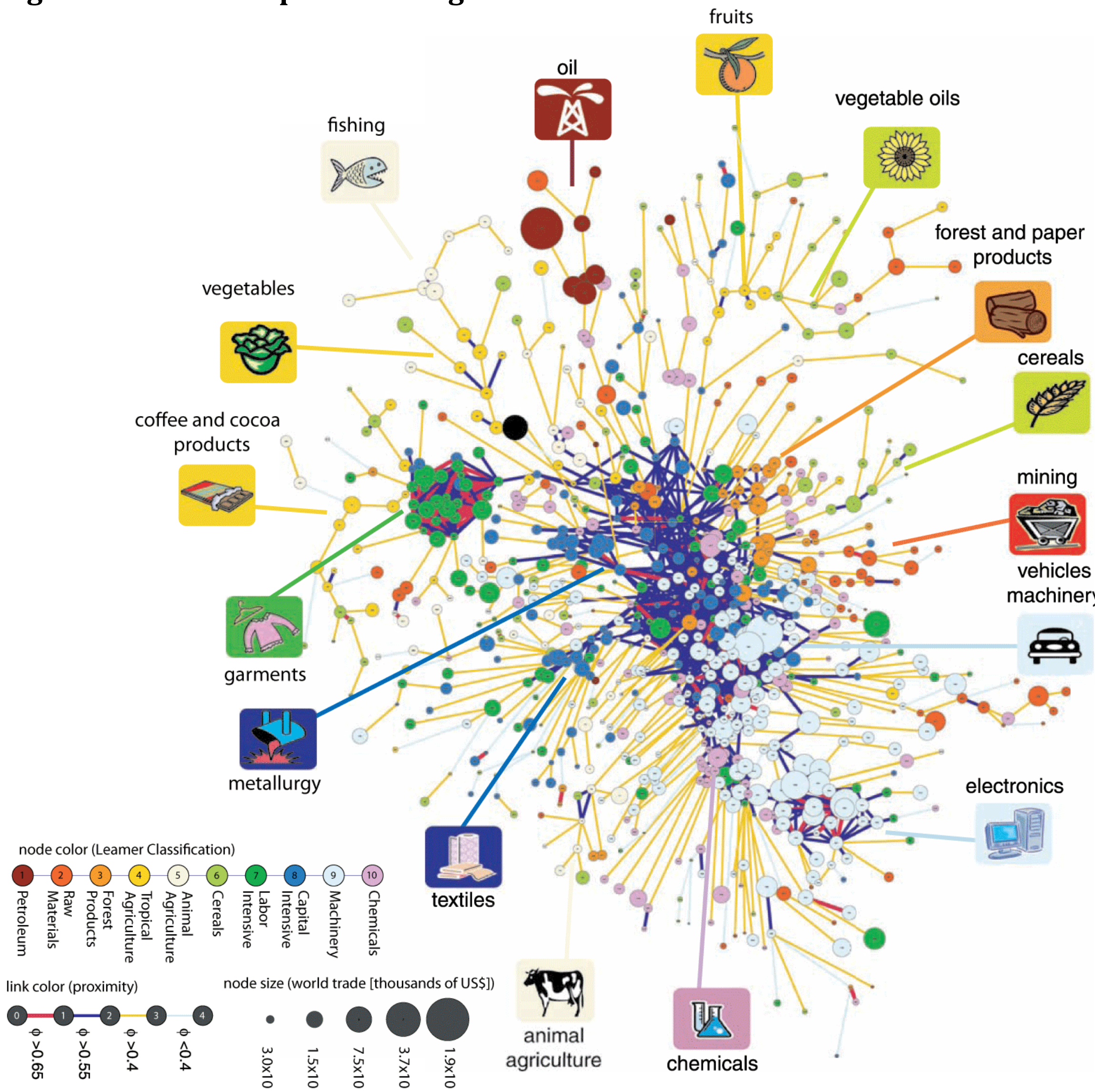
Now, some products like diamonds have low ubiquity, not because they require many letters but because they require rare letters like Q or X. How can we correct for rare letters as distinct from long words? Well, if diamonds were complex in terms of the number of letters they require, then countries that make raw diamonds should be able to make many other things, i.e. they would be diversified. But Botswana and Sierra Leone export few things besides diamonds. So we can use the diversification of a product's exporters to correct for the information about the number of letters the product requires that can be inferred from its ubiquity. So we can use ubiquity to correct for the information about the number of letters contained in a country's diversification and use diversification to correct for the information about the number of letters contained in a product's ubiquity. We can do this an infinite number of times and this becomes an eigenvalue problem. Hidalgo and Hausmann (2009) and Hausmann, Hidalgo et al (2011) use this to calculate an Economic Complexity Index: a measure of how complete is a country's subset of the alphabet. They find that the economic complexity index is highly correlated with the level of income of countries. More importantly, deviations in this prediction are correlated with future growth: countries tend to converge to the level of income that can be supported by the number of letters that they have.

A second idea is related to the topology of the product space. What is the proportion of valuable words that can be made with a certain number of letters? Is the distribution of words a purely random sampling of all the letter combinations that can be made? Or are there more products that can be made with some combinations of letters than others?

The topology of this space matters for the process of accumulating letters because this process faces a serious chicken and egg problem. You cannot make products (words) for which you lack capabilities (letters), but you do not want to develop capabilities for industries that do not exist. Moreover, you have no local exemplars of these capabilities that you can just copy, as they are not present in the locality. So the distance between products in terms of capabilities will affect the severity of this chicken and egg problem and hence the capacity to accumulate letters and express them into valuable new words.

The distance between products, or the topology of this space can be inferred from the information contained in the likelihood that any two products are co-exported by the same country. In principle, if two products require a very similar set of letters, then countries that are able to make one should be able to make the other with a probability much higher than that of making also a randomly chosen alternative product. I introduced the product space in Hausmann and Klinger (2006) and explored it using network science in Hidalgo et al (2007). We showed that the product space is highly heterogeneous with sparse sections and more densely connected parts (Figure 8). Countries diffuse in this space by moving preferentially towards nearby goods as captured by this very heterogeneous product space. Countries in better-connected part of the product space find it easier to diffuse in this space because products not currently being made are fewer missing letters away. The inference is that they should be able to grow more rapidly.

Figure 8 – Product Space – Hidalgo et al 2007



We developed a measure of how well positioned is a country in the product space which we call Opportunity Value. It is based on the sum of the distances between the products a country is not making and the products it is making, weighted by how complex those missing products are. The measure of distance is based on the conditional probability that the pair of products under consideration is co-exported.

The distance between any two goods in the product space is defined as:

$$\varnothing_{pp'} = \frac{\sum_c M_{cp} M_{cp'}}{max(k_{p,0}, k_{p',0})}$$

where $k_{p,0}$ is the ubiquity of each product.

The distance between a country and a product can be calculated as:

$$d_{cp} = \frac{\sum_{p'} (1 - M_{cp'}) \varnothing_{pp'}}{\sum_{p'} \varnothing_{pp'}}$$

Then opportunity value faced by a country is calculated is related to distance between the products that are not being made and the products that are, weighted by the Product complexity Index.

$$opportunity\ value_c = \sum_{p'} (1 - d_{cp'})(1 - M_{cp'})\, PCI_{p'}$$

Interestingly, the Economic Complexity Index and Opportunity Value do a remarkably good job at explaining growth over the past three decades, as shown in Table 1. In fact, few other traditional controls add much explanatory power, after controlling for these two variables. These poorly performing indicators include finance, education or governance, three of the most popular explanations of growth in the literature (Tables A1 to A3 in the appendix).

Table 1 – Economic Complexity and GDP per capita Growth

Dependent variable: Annualized 10 real GDP growth per capita (%)

VARIABLES	(1)	(2)	(3)	(4)
Initial GDP per capita, logs	-0.009	-0.667***	-0.489***	-0.738***
	(0.125)	(0.163)	(0.142)	(0.145)
Increase in real NNRR exports pc	4.034***	3.794***	4.062***	3.905***
	(0.830)	(0.919)	(0.967)	(0.979)
Initial Economic Complexity Index		1.393***		0.859***
		(0.228)		(0.197)
Initial Opportunity value Index			1.235***	0.832***
			(0.226)	(0.215)
Constant	1.326	6.267***	4.894***	6.776***
	(1.097)	(1.323)	(1.173)	(1.177)
Observations	294	294	294	294
R-squared	0.269	0.390	0.399	0.431
Year FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

We can use this framework to understand the Great Divergence and the recent beginning of a Great Convergence, where middle-income countries are narrowing the income gap with the rich countries, after two centuries characterized by divergence.

Figures 9a-d show the relationship between Opportunity value and the Economic Complexity Index for 1978, 1988, 1998 and 2008. They show how the relationship has changed from a highly concave pattern in the early years –characteristic of the quiescence trap – into an inverted U-shaped curve. This indicates that while it used to be the case that the most complex countries had an easier time at diffusing in the product space, it is now middle complexity countries that are best placed to make progress.

Figure 9a – 1978

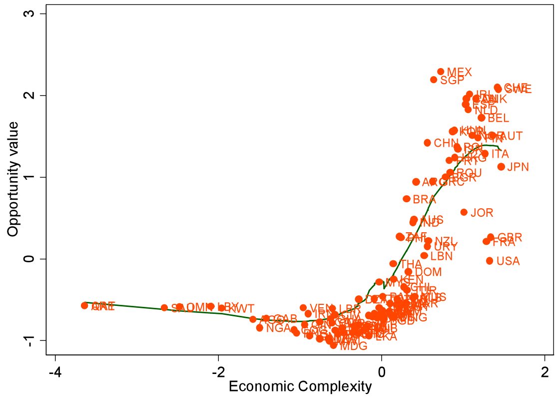


Figure 9b – 1988

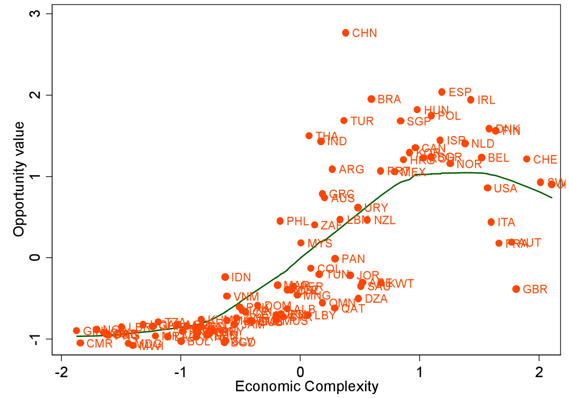


Figure 9c - 1998

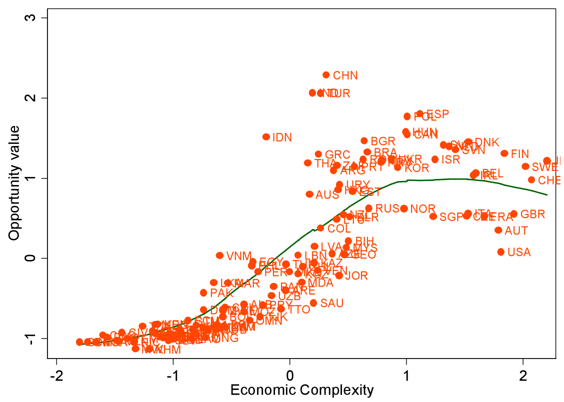
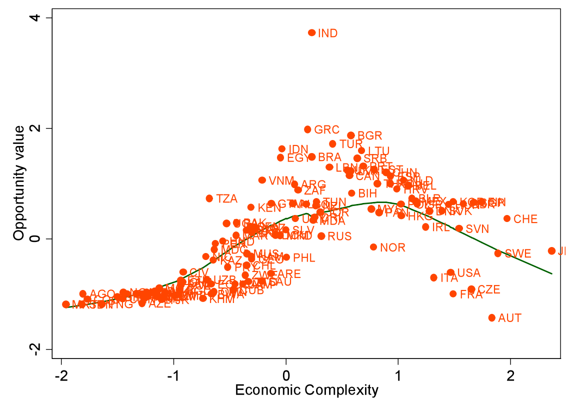
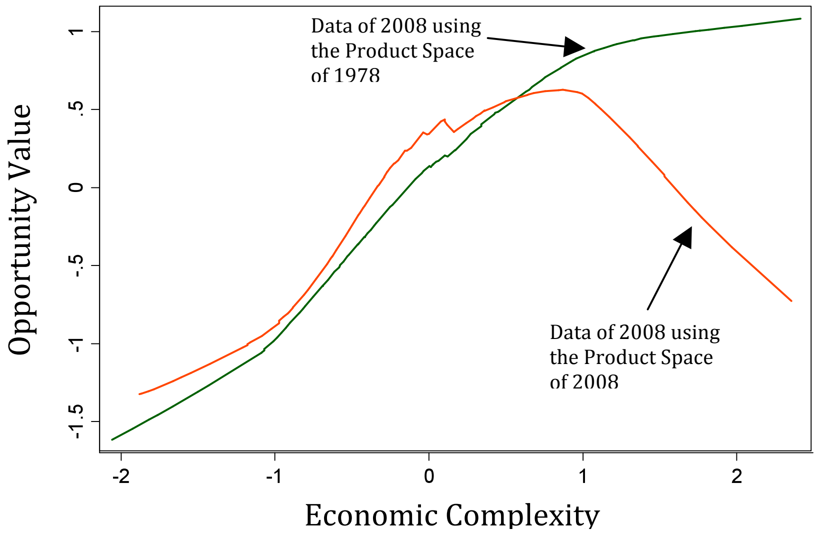


Figure 9d - 2008



We can further explore what caused this shift by decomposing the changes that have taken place in Opportunity Value since 1978 into two elements: changes in the nature of the product space, as captured by $(1 - d_{c,p'}.)PCI_{p'}$ and the increased diversification of the country, as captured by $M_{c,p}$. This decomposition is presented in Figure 10. It shows that the change from a concave relationship to an inverted u-shaped curve was not driven by the increased diversification of countries in the intervening period but by changes in the product space.

Figure 10



The conjecture I would like to propose is that globalization has split up the value chain. To be able to participate in it, countries no longer need to be able to write words, but just syllables. Words can be assembled out of syllables made in several different places. This has made it easier for countries with fewer locally available letters to enter production, making the quiescence trap less serious. Once in production, the accumulation of more letters and their expression in more syllables has become less challenging, causing an improved opportunity value at lower levels of complexity and accelerated convergence. This unfortunately is only true for countries at intermediate levels of complexity. Poor countries still face lousy prospects.

Concluding remarks

We have been able to make these findings without saying much about what these unobserved "letters" or productive capabilities actually are. Here, I will speculate about what is behind them.

The central idea is that things are made with productive knowledge or know-how. To make things you need to know how to make them. This knowledge can be embedded in products or in machines, but then somebody must be able to make the product or the machine that makes the product. Soap requires knowledge about the properties of lipid salts and the synthetic paths that lead to making them. Somebody needs to know how to make soap, but to wash your hands you only need to buy soap at the local grocery store. But if soap is to be made somewhere, somebody there must have the relevant knowledge. Countries can only make the products that they know how to make.

In this interpretation, the rise in incomes over the past two centuries – a factor of 25 in the United States since 1820 –has been caused by an explosion in productive knowledge. We learned how to make more things and how to make things more efficiently. But as individuals we have not become 25 times smarter. In spite of the expansion of knowledge over the past two centuries, a college degree still lasts four years and a PhD between 4 and 6. Instead, we have had to rely on an increased division of knowledge: the expansion of productive knowledge has had to go, hand in hand, with the reduction in the share of total knowledge each one of us holds. Each member of a hunters and gatherer tribe knew pretty much all that the group knew how to make. Nowadays, not even Apple knows how to make a computer. It must rely on a long chain of suppliers who know how to make the many parts and software that go into a computer.

If we think that individuals have a limit on how much productive knowledge they can hold, call it 1 personbyte, then societies can be characterized by the number of different personbytes they hold. Products can be characterized by the combination of personbytes they require. More complex products require more personbytes and hence larger aggregations of individual know-how. Production requires networks of individuals that have complementary copies of the requisite personbytes. Firms are networks of individuals and value chains are networks of firms. They are the mechanisms through which societies mobilize the personbytes they have to make useful things with them.

Institutions, finance and education may be important catalysts of this process. Institutions allow personbytes to interact through contracts. Finance allows some to hire others, so as to put the network together. Schooling may allow people to gain access to personbytes that would be inaccessible without some more basic skills: you cannot be a sound engineer unless you finish high school. But a catalyst facilitates a reaction: it is not the reaction itself.

By aggregating production into GDP we have disregarded much of the information that is contained in the structure of production and the knowledge it embeds. Was there some wisdom behind the obsession with manufacturing that Alexander Hamilton, Albert Hirschman and Raul Prebisch had? Can this obsession be more easily understood by noticing how differently positioned in the product space manufacturing is vis a vis other activities such as mining or agriculture? Can the so-called natural resource curse be more readily understood by noticing how peripheral it is in the product space, suggesting that the type of personbytes that they lead countries to accumulate have few alternative uses? Can an understanding of the reactants and not just the catalysts lead to a better understanding of what productive development policies might be about?

That is the challenge, and maybe the promise of the approach I have delineated. The proof of the pudding….

Bibliography

Hamilton, Alexander (1791) *Report on Manufactures,* available in http://press-pubs.uchicago.edu/founders/documents/v1ch4s31.html

Hausmann, R. & Hidalgo, C. The network structure of economic output *Journal of Economic Growth, Springer,* **2011**, 1-34

Hausmann, Ricardo and Bailey Klinger (2006) "Structural Transformation and Patterns of Comparative Advantage in the Product Space", *Kennedy School of Government, Harvard University,*

Hidalgo, C. & Hausmann, R. The Building Blocks of Economic Complexity *Proc. Natl. Acad. Sci.,* **2009***, 106*, 10570-10575

Hidalgo, C. Hausmann, R. Bustos, S. Coscia M, Chung, S., Jimenez, J., Simoes, A. Yildirim, M. The Atlas of Economic Complexity – Mapping paths to prosperity.

Hirschman, Albert (1958) *The Strategy of Economic Development*, Yale University Press.

Kahneman, Daniel (2011) *Thinking Fast and Slow, New York:* Farrar, Straus and Giroux.

Krugman, Paul "The Fall and Rise of Development Economics" http://web.mit.edu/krugman/www/dishpan.html

Raúl Prebisch (1950), *The Economic Development of Latin America and Its Principal Problems* (New York: United Nations)

Simon, Herbert (1996) *The Sciences of the Artificial,* MIT Press.

Appendix

Table A1 – Education, Economic Complexity and GDP per capita Growth

Dependent variable: Annualized 10 real GDP growth per capita (%)

VARIABLES	(1)	(2)	(3)	(4)	(5)
Initial GDP per capita, logs	-0.762***	-0.954***	-0.884***	-0.776***	-0.921***
	(0.144)	(0.170)	(0.154)	(0.149)	(0.162)
Increase in real NNRR exports pc	6.604***	6.146***	6.218***	6.593***	6.001***
	(1.257)	(1.303)	(1.331)	(1.262)	(1.339)
Initial Economic Complexity Index	0.958***	0.766***	0.774***	0.947***	0.744***
	(0.184)	(0.174)	(0.187)	(0.184)	(0.181)
Initial Opportunity value Index	0.851***	0.805***	0.869***	0.850***	0.827***
	(0.218)	(0.219)	(0.210)	(0.220)	(0.212)
Initial average years of schooling		0.211***			0.209*
		(0.066)			(0.113)
Initial percentage of Secondary Complete			0.045***		0.027
			(0.015)		(0.021)
Initial percentage of Tertiary Complete				0.009	-0.069*
				(0.029)	(0.038)
Constant	6.996***	8.383***	8.452***	8.376***	8.136***
	(1.187)	(1.235)	(1.175)	(1.207)	(1.184)
Observations	261	261	261	261	261
R-squared	0.386	0.406	0.407	0.386	0.417
Year FE	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A2 – Finance, Economic Complexity and GDP per capita Growth

Dependent variable: Annualized 10 real GDP growth per capita (%)

VARIABLES	(1)	(2)	(3)	(4)
Initial GDP per capita, logs	-0.756***	-0.712***	-0.720***	-0.781***
	(0.151)	(0.159)	(0.178)	(0.178)
Increase in real NNRR exports pc	3.927***	3.757***	3.889***	3.728***
	(1.008)	(0.966)	(1.007)	(0.948)
Initial Economic Complexity Index	0.939***	1.091***	0.985***	1.079***
	(0.213)	(0.243)	(0.227)	(0.240)
Initial Opportunity value Index	0.859***	0.831***	0.855***	0.821***
	(0.218)	(0.229)	(0.228)	(0.212)
Initial Domestic credit provided by banking sector (% of GDP)		-0.007**		-0.013**
		(0.003)		(0.005)
Initial Domestic credit to private sector (% of GDP)			-0.003	0.010
			(0.005)	(0.008)
Constant	6.040***	6.058***	5.900***	6.485***
	(1.162)	(1.177)	(1.272)	(1.288)
Observations	273	273	273	273
R-squared	0.474	0.482	0.475	0.485
Year FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1