Business and Politics 2017; 19(3): 397–423

Steven Weber*

Data, development, and growth

Abstract: Economic development and growth theory have long grappled with the consequences of cross-border flows of goods, services, ideas, and people. But the most significant growth in cross-border flows now comes in the form of data. Like other flows, data flows can demonstrate imbalances among exports and imports. Some of these flows represent 'raw' data while others represent high-value-added data products. Does any of this make a difference in national economic development trajectories? This paper argues that the answer is yes. After reviewing the core logic of 'high development theories' from the twentieth century, I analyze the sometimes implicit applications of these arguments to data as they are evolving in the existing literature. I then put forward a different argument which takes better account of unique characteristics of the political economy that emerges at the intersection of data, machine learning, and the platform firms that use them. I explore the implications of this new argument for some policy choices that governments face with regard to data localization, import substitution, and other decisions relevant to growth in both advanced and emerging economies.

doi:10.1017/bap.2017.3
Previously published online April 17, 2017

Introduction

This paper sets out some elements of a high development theory[1] for the era of data. High development theory is a set of "big ideas" about how the world economy functions and what countries should do to advance their growth, competitiveness, productivity, and wealth interests within it.[2] The phrase 'era of data'

Funding for this work was provided by the Hewlett Foundation via the Center for Long Term Cybersecurity at UC Berkeley; and by the Carnegie Corporation via the 'Bridging the Gap' grant program.

Acknowledgments: Thank you to Jesse Goldhammer, Stephane Grumbach, Phil Nolan, AnnaLee Saxenian, Naazneen Barma, Elliot Posner, two anonymous reviewers, and the Editors of *Business and Politics* for comments and critiques.

1 Krugman (1992), 15–38.

2 Lindauer and Prichett (2002), 1–39.

***Corresponding author: Steven Weber**, Professor, School of Information and Department of Political Science University of California Berkeley steve_weber@berkeley.edu

© V.K. Aggarwal 2017 and published under exclusive license to Cambridge University Press

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

captures the proposition that the world economy is transitioning from a phase of container shipping to one of packet switching, where the largest and most important cross-border flows are data not physical goods.

Cross-border data flows have become interesting and controversial in the last several years for a number of reasons. Privacy and security issues have garnered the most attention and (at least in the short term) with good justification.[3] This paper explicitly and intentionally does not deal with those concerns, other than tangentially. But I am *not* arguing that those issues are anything other than substantial, real, and important. This paper places them in the background because they are dealt with elsewhere. And—as important as they are—because they obscure attention to something I think will be even more important in the long run.

It is my goal here to open up a new dimension in the debate around data flows. I believe the predominant longer-term question about this new era will concern economic growth. Put simply, *do data flow imbalances make a difference in national economic trajectories?* If a country exports more data than it imports (or the opposite) should anyone care? Does it matter what lies inside those exports and imports—for example, 'raw' unprocessed data as compared to sophisticated high value add data products?

Consider a thought experiment. Country X passes a "data localization" law, requiring that data from X's citizens be stored in data centers on X's territory (ignore for the moment the various motivations that might lie behind this law). Now, a data-intensive transnational firm (say Company G) has to build a data center in Country X in order to do business there. The first-order economic effects are relatively easy to specify: Country X will probably benefit a bit from construction and maintenance jobs that are connected to the local data center, while Company G will probably suffer a bit from the loss of economies of scale it would otherwise have been able to enjoy.

It is the second and third order effects that need greater understanding. Imagine that the national statistics authority of Country X develops and publishes a "data current account balance" metric which shows that cross-border data flows two years later have declined in relative terms. Now the critical question: Is this a good or bad thing for X? Do nationally based companies that want to build value-add data products inside country X see benefits or harms? And does any of this matter to the longer-term trajectory of X's economic development? These are

3 For obvious reasons, most notably the controversy over the NSA documents released by Edward Snowden, but also the re-working of the trans-Atlantic Safe Harbor Agreement, and the general increase in cyberattacks and the awareness of cyberattacks. Note that these two categories are often conflated in popular and even scholarly discourse ("security and privacy"). They are related in some instances, but are quite different issues.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

the types of issues that I believe high development theory will need to address for the era of data.

Data, development, and inequality

Why do we need a new high development theory at all? The obvious reason is because "economic development" is not confined to what are colloquially "developing" countries but is relevant—in different ways of course—to countries at all levels of GDP. This is most powerfully true during periods of profound economic transition, when basic foundational arguments and beliefs about economic growth and development are deeply contested (as is true of the current moment).

A little bit of history situates the need further. The modern post World War II industrialized era has seen the cyclical rise and fall of two or three such big ideas, depending on whom you ask. The first big idea—import substitution industrialization (ISI)—lasted roughly from 1945 to 1982. The second big idea—Washington Consensus—lasted from about 1982 to 2002. Some observers believe that the success of the Chinese state-capitalism development model represents the foundations or empirical representation of a third big idea. Still others think the search for big ideas is an intellectual and policy diversion.[4]

I take the position that nature abhors a vacuum in this respect, and a (fourth) big idea for the era of data-enabled growth is wanted. I don't claim here to have landed on the 'right' one. My goal instead is to start the discussion in as constructive a fashion as possible by putting forward a few key principles that I believe the big idea, if and when it emerges, will have to confront and address.

Consider to start a high level rendition of the 'old' big ideas, in order to see clearly what principles those ideas were grappling with and what they were trying to achieve. The core argument of ISI was simply that decent national growth required having essentially the complete supply chain of an industry located physically at home, within national borders. The underlying mechanisms that would generate self-sustaining growth included "learning by doing," coordination of a large number of interdependent processes, the lumpiness of tacit knowledge, and the like. Either a country had a deep industrial base, or it did not. A country that did not would be stuck in a low productivity box and suffer from detrimental terms of trade that together would impede development, possibly indefinitely. The policy question that followed from this, was how to mobilize a complete supply chain within a country. The policy answer was (usually) through a

4 Rodrik (2005), chapter 14; Elsevier, 967–1014.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

combination of tariffs and other restrictions on imports along with subsidies and other inducements to jump-start domestic import substitution.

The core argument of the Washington Consensus was almost 180 degrees reversed. It built on the proposition that reductions in transportation and communications costs made it possible to unbundle (as Baldwin put it) the supply chain. Now the mechanism of growth rested on moving parts of the supply chain around the world and beyond national borders, and organizing the pieces.[5] Global growth became a story about combining high technology with low cost labor—most likely from a different country—and coordinating the package through information and communications technologies (ICT). Note that in the Washington consensus big idea, ICT was mainly about organizing a global supply chain of manufacturing processes, not about the value of data in and of itself or discrete 'data products'.

Getting macro-economic policy "right"—the most visible and consensual policy aspect of the second big idea—was a necessary condition to a country joining these increasingly globalized value chains which were now not only possible, but necessary from a competitive perspective given the vast scale economies and cost reductions they made possible. Failing on macro-economic policy would leave a country isolated from global supply chains and stuck without a ladder on which to climb toward self-sustaining and productivity-enhancing growth. Put differently, it would leave that country poor.

Other policy questions addressed the same core logic from different directions: what supply chains are most promising; how do you manage intellectual property and trade secrets; what's a reasonable trajectory for low-wage labor that starts the growth system rolling and gets your country moving up the ladder? This was the era of the cross-national production network that became the iconic image of 1990s globalization.[6]

In the second half of the 2010s the allure of big idea 2 is considerably reduced. One simple data point demonstrates the setback: Since the global financial crisis began in 2007–8, global flows of goods, services, and finance together have fallen substantially as per cent of GDP.[7] If 2017 represents recovery after nearly a decade, it is a tepid one at best for these global flows, which are still substantially below their level as percent of GDP in 2007.

5 Baldwin (2011).

6 For "cross national production network" see Zysman, Doherty, and Schwartz (1996).

7 Manyika, Lund, Bughin, Woetzel, Stamenov, and Dhingra (2016), page 3. Disaggregated, goods trade is flat; financial flows are sharply lower; services trade is modestly higher. Somewhere around 2014, the aggregate flows regained nominal pre-recession levels in terms of dollar value (eight years post crisis) but as percent of GDP that represents roughly a 14 percent decline (from 53 percent in 2007)

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

The contrast is particularly strong in goods trade, which grew nearly twice as fast as global GDP for the twenty years leading up to the Global Financial Crisis of 2008 (GFC), underwent a sharp decline and some rebound, but is now growing at a rate below GDP growth (which itself has slowed). While some of this decline may be cyclical (and some a statistical quirk of the depression in commodity prices), a good proportion is almost certainly structural and a consequence of changes in technology, policy, and beliefs about the viability and competitive advantages of global vs. local supply chains (more on this later.) It is also likely in part a consequence of perceived zero-sum competition for jobs in a political economy environment where employment is a key linchpin for political stability among democratic and non-democratic governments alike.

The Data revolution emerges

While the world has been looking for a big idea to re-ignite growth post-GFC, something potentially much more important has been happening in the world of technology. The labels for that phenomenon include "big data," "data science," "data-native companies," and a few others.[8] For the purposes of this paper, I will use the phrase 'data revolution' — with some caution about overhype, but with conviction that the increasing importance of data as a factor of production does and will deserve the term revolution.

What does the data revolution portend for high development theory and the next big idea?

I have to start by contending (lightly for now, and in more detail later) with the intuition and sometimes reflex response that "data is different." Of course it is—just as oil was different from manufactured goods, and intellectual property was different from both. That is to say, a generation of economic development big ideas will and must take account of the basic nature of the main 'fuel' that drives growth. The question really is, is data so radically different from previous growth drivers that basic analytic categories and variables used in previous generations don't apply? I don't believe that to be true.

One intuition is that data is extremely cheap to generate, once the basic infrastructure to produce and collect it is in place. But so was oil in the early days, when it bubbled up out of the Pennsylvania fields and the sand dunes of the Eastern provinces of Saudi Arabia.

8 See for example Chen, Mao, and Liu (2014) 171–209; Fosso Wambe et. al. (2015), 234–246; Chen et. al. (2016), 1–42.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

Another intuition is that "raw" data is non-rival, because it can be copied an infinite number of times for essentially no cost. But the same is true of much "raw" intellectual property. It's accurate to say that most of the time, other inputs—many of which are rival and/or expensive—have to be combined with intellectual property in order to create real value. The same is true of most data.

A third intuition: Data is everywhere, and the challenge is simply to collect it and figure out what to do with it. But the same is (almost) true of oil and natural gas. It is fully true of bacteria and viruses that represent the "raw materials" of the pharmaceutical sector.

The point, again, is not to say that there are no meaningful differences between data and previous growth drivers. It is to say that basic variables and categories used in existing growth theory, are legitimate starting points for an analysis of the data era.

Start, then, with a baseline view that trade in data exhibits the same basic properties as any other kind of trade exhibits in classical theory—specifically, that trade generates increased productivity through competitive advantage and the creation of more efficient markets with global scale. It could be the case then, that data flows simply raise productivity across the board and act as a tide that lifts all boats. The McKinsey Global Institute (MGI) has put forward a very clear articulation of this position, arguing that directionality and content is irrelevant because data flows "circulate ideas, research, technologies, talent, and best practices around the world."[9]

This tracks with one kind of intuition, widely held, that innovation flows in all directions along with data. If accurate, then data trade might very well be a positive sum game in most instances. And that in turn leads to the policy proposition that countries should aim most importantly to increase the centrality of their position in what is still a highly uneven network. The nascent high development idea embedded here, is that growth and development will accelerate if and when data flows accelerate further, and if and when countries that are 'behind' in the race to engage with data flows "catch up" to the leaders. Paying attention to the magnitude of imports vs, exports, or the content of those imports and exports, is a distraction at best and self-defeating at worst. Putting up constraints on data flows for other reasons (privacy concerns, for example) may serve other interests and values that are important, but come at a cost to economic growth and development.

It's crucial to keep in mind that these are theories or in some cases just intuitions, not laws of nature or empirical findings. Analogous theories about goods and services trade have been the subject of research and controversy for decades (in some cases for centuries). It is absolutely accurate to observe that

9 MGI (2016a), 98. See also the previous report from MGI, "*Global Flows in a Digital Age*" 2014.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

those controversies are almost never purely theoretical but almost always become politicized; that's why we use the term "political economy of trade." The same will almost certainly be true for data trade. But we need to start somewhere, if for no other reason than to provide some guidance and set some boundaries on the terms of political debate.

The best place to start is with a deeper examination of what some broad contemporary perspectives suggest about data flows and high development ideas. Following that, I dig deeper by assessing core theoretical perspectives on trade that ground the contemporary debates about data. I develop several propositions about the strategic position of a second-tier country in the data economy, and use a thought experiment to explicate and assess specific choices. The paper concludes with an argument about the new logic of vertical integration that might emerge, and a pessimistic evaluation of the possibility that less developed countries might "leapfrog" technology generations and catch up or get out ahead of today's leaders.

Data Imbalances and Economic Development—Perspectives from Political Economy

Do data imbalances make a difference in the trajectory of national development? How countries answer this question and use that answer to shape policies may be the single most important economic development decision made in the next decade. And there is a major gap in explicit theory to guide that decision making. That gap is being filled in by naive absolute gain arguments, loose extrapolations, big unexamined assumptions, rough-and-ready analogies, and in some cases reflexive nationalist impulses. An overly harsh characterization? Possibly, but the history of trade policy and trade theory in international political economy through previous decades and centuries is in fact a harsh reminder of just how often and how easily instincts and reflexes overwhelm decision making and can lead to dysfunctional and sometimes dangerous outcomes (that's why the political economy of data, not just the economics of data, are the right focus).

Contemporary perspectives on data imbalance fall roughly into three packages. An "absolute gains" perspective suggests that being engaged in global data flows is a critically important development objective for countries, regardless of the content or directionality of those flows. A related but slightly different "open is best" perspective suggests that restrictions on the free flow of data across borders will reduce efficiency and wealth creation overall, with negative growth implications generally. A third perspective I will call "data nationalism" to

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

capture the idea that (unless there is a definitive reason to believe and act otherwise) it will be better for countries to internalize significant parts of the data economy within their own borders. A short review of the logic of each package helps to situate these perspectives both historically and in the context of recent events.

Take the absolute gains perspective to start. The influential McKinsey Global Institute reports of 2014 and 2016 (cited above) are the most explicit representation of this point of view. Tracking a measure of "used cross border bandwidth" as a proxy for Internet traffic, MGI estimates that flows of data moving across national borders have grown by forty-five times in the nine years between 2005 and 2014.[10] It's a stunning number but hardly implausible, if you stop and think for a moment about the number of globally-active data-native companies that have been founded and grown massively during that period. (YouTube, for example, was founded in February of 2005; Facebook in February 2004.)

In the absolute gains perspective, the massive expansion in trans-border data flows is in a real sense the other side of the relative decline in more traditional (i.e., goods) trans-border flows. Importantly, the MGI model seeks to assess the impact of these flows on economic growth via improvements in productivity. Consistent with the fundamental insights of basic trade theory, greater trans-border flows should generate higher productivity and boost growth. The "absolute gains" argument follows from that, as their model examines how a country's overall position in the network of flows impacts its growth.

Put simply, the model posits "the more flows you experience, the greater the positive impact on your productivity." Countries are distinguished by being either "central" or "peripheral" in the overall global flow pattern, and the countries at the center (mainly the United States and European countries along with a few special cases like Singapore) benefit the most. Peripheral countries are catching up slowly in goods trade but not yet in cross-border data flows. The proposition is that countries in the "data periphery" stand to gain even more than those at the center by increasing their level of connectedness.[11]

This is an absolute gains perspective because it depends on the *depth of connectivity* and the *magnitude of flows* alone. It does not distinguish among *directionality* or *content* of data flows. In this model, data imports and data exports are equivalent. "Raw" data and highly processed data products are equivalent. What matters is simply the overall magnitude of flows. An analogy to goods trade would be this: Imagine a model in which we count the number of shipping containers that transit a country's borders. It does not matter in which direction the containers are

10 MGI (2016b), 4.

11 MGI (2016a), 10.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

moving (in or out) and it does not matter what is inside those containers (raw materials, high value added industrial products, intermediate products that enter another part of the supply chain in another country, and so on). What matters for productivity enhancement and growth is first and foremost the number of containers; or in this case the level of flows.

Now consider the second package of arguments, "open is best." This package focuses less on data flows per se, and more on the broader international political economy of Internet platform businesses like Uber and Airbnb (more on the precise definitions of "platforms" later). The International Technology and Innovation Foundation (ITIF) articulates this perspective clearly and strongly in a number of papers, most vividly "Why Internet Platforms Don't Need Special Regulation."[12]

Responding to growing anxieties in some European countries about the intrusion of *non-domestic* platform businesses (mostly based in the United States) into domestic economies, ITIF argues vehemently against regulation that would constrain access to markets. The case is made principally on the basis of market power analyses from antitrust theory: while it is true that many platform businesses benefit from positive network effects and economies of scale, their market power remains limited by a number of factors. These include "multi-homing" (consumers can easily participate in competing platforms with a couple of clicks or an app download); low barriers to entry (cheap and scalable cloud computing resources mean that new entrants to platform businesses are inevitably coming); the demand for continued innovation (which forces platform business to continually invest in new products and features); and the demonstrated history of platform businesses losing market dominance when they fail to innovate (the decline of MySpace and Friendster).

Because there is no systematic reason to expect that platform businesses are any more predisposed to accumulate exploitable market power than any other business—and possibly less reason to expect so—there is no prima facie case for a priori regulation to constrain their growth and market access. Countries should remain open to platform businesses regardless of where those businesses are domiciled; regulators have all the power they need to deal with any anti-trust abuses that might occur using standard tools on a case-by-case basis.

The "open is best" perspective acknowledges that the growth of platforms has caused other political economy anxieties beyond concentration of market power. These include the rights of contract labor and shifts in the baseline employment relationship characteristic of the independent contractor (or "gig") economy, and of course the resistance of incumbents (like traditional taxi companies) to

12 Kennedy (2015).

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

"disruptive" competitors. There is also a recognition of non-specific concerns about data security and privacy, and an awareness that general fears of declining national competitiveness can sometimes be layered on top of debates about non-domestic platforms.

But the 'open is best' perspective downgrades these issues and treats them as the conventional rumblings of companies and countries that have fallen behind *for the moment* in competitive markets. The prescription (familiar and consistent) is not to slow down change or regulate disruption or raise barriers to entry or try to rebalance the wealth through taxes or other re-allocation means. The correct response is to compete vigorously on the open playing field. Illustrating that, the ITIF lists among its "Ten Worst Innovation Mercantilist Policies of 2015" that are said to impede national growth, four cases of countries imposing some form of local data storage requirements.[13] Within the logic of the "open is best" perspective, there is nothing special worth arguing about data flows, data imbalances, or raw data/data products imports vs. exports.

The third package, data nationalism, is the most primordial in some respects. Data nationalism isn't simple old-fashioned mercantilism or even "enlightened mercantilism" (GATT-think as Krugman called it in a still-compelling 1991 paper).[14] It is an almost reflexive response to the emergence of a new economic resource (data) that appears to power a leading-edge sector of modern economies. Put simply, if this new resource emerges and no one yet understands precisely how, when, and why it will be valuable, the "hoarding" instinct tends to kick in as a default. Barring definitive reason to believe otherwise, shouldn't countries seek to have their own data value-add companies "at home" to build their domestic data economies? And if data is the fuel that drives those companies, why would a country allow that fuel to travel across borders and power growth elsewhere?

The argument that data is a non-rival "fuel" is accurate, but largely irrelevant within this perspective. It is true that "sending" data abroad is not like exporting a barrel of oil or even a semiconductor chip in the sense that a piece of data can be copied and kept inside a domestic economy as well as being exported, at zero cost. The point is that the value of that data depends on its ability to combine with other pieces of data, and that this positive network effect will create benefits at an increasing rate in places that are the landing points for broad swathes of data.

In previous work I referred to this characteristic as "anti-rivalness"; it matters here because accumulations of data become disproportionately more valuable as they grow larger.[15] This in turn enables a virtuous circle of growth in and through

13 Cory (2016).
14 Krugman (1991).
15 Weber (2005).

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

the production and sale of data products. *The more data you have, the better the data products you can develop; and the better the data products you develop and sell, the more data you receive as those products get used more frequently and by larger populations.*

A country that sees itself as falling behind in this dynamic might want to hoard data in order to subsidize and protect its own companies as "national data champions" in the same way that previous generations of developing economies sometimes provided support for national industrial champions. Hoarding would also deprive other country's companies of data, which might slow down the leader enough to make catch-up plausible.

Of course, data nationalism has also grown in the last few years because events like the Snowden revelations (mostly exogenous to the broader debate about data and economic growth) created overlaps with concerns and anxieties around espionage and privacy. There are other drivers behind the data nationalism impulse as well. Some countries emphasize consumer protection in advance of demonstrated harms and with more subtle arguments about digital "fairness."[16] Others highlight national differences in boundaries between what constitutes legitimate speech or illegitimate and objectionable content.

Data Imbalances and Economic Development—Theory

These perspectives on the political economy of data-intensive growth draw from foundational theories of trade, imperfectly, and selectively. It's possible to gain a better understanding of what's really at stake by taking a step back to delve a little bit more fully into the underlying theoretical propositions that ground the perspectives.

Start with the most basic metric. A current account balance is simply the difference between the value of exports (goods and services, traditionally) and the value of imports that transit a country's borders.[17] It can also be expressed as an accounting identity, the difference between national savings and national investment. When countries argue with each other about the current account, it's generally the case that the arguments circle around what is *causing* the imbalance and whether someone—a government that enacts certain policies or the business

16 A proposed French law cites the principle of *loyaute des plateformes* roughly translated as "platform fairness."

17 A small additional point is that the current account includes along with this trade balance, net income, and transfers from abroad, but these are typically small relative to trade.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

practices of firms, generally—is to "blame." In fact this has been a consistent theme in international political economy debates in the public realm, most visibly (for Americans at least) in the Japan-U.S. relationship during the 1980s and in the Sino-American relationship today. These arguments tend to mount when the magnitude of a current account imbalance becomes, in the estimation of some important actors, "too large."

Now consider in this context a thought experiment, of a "current account in data." Countries at present do in fact complain over some form of data imbalance and argue whether the dominance of U.S. platform businesses ought somehow to be "blamed."[18]

The modal argument goes something like this: a small number of very large "intermediation platform firms," most of which are based in the United States, increasingly sit astride some of the most important and fast-growing markets around the world. These markets are driven by data products, and the new firms use data products to "disrupt" existing businesses and relationships without regard to domestic effects (that can be economic, social, political, and cultural). The term intermediation platform captures the essential nature of the two-sided markets that these firms organize.[19] They collect data from users (on all sides of the two-sided or many-sided market) at every interaction; bring that data "home" into vast repositories which are then used to build algorithms that process "raw" data into valuable data products; and use those data products to create new and yet more valuable products and lines of business. The platform businesses grow more powerful and richer. The users in other countries *get to consume the products but are shut out of the value-add production side of the data economy*.

That's a data imbalance in the sense that one country is being "locked in" to the role of raw data supplier and consumer of imported value-added data products; on the other side, the home country of the platform business imports raw data and adds value to create products that it then exports. The question is, so what? Does it matter for either country?

It might be easier to answer that question if the analogous question in traditional forms of trade (goods and services) was fully settled; it isn't. But the ways in

18 See for example Tom Fairless, 23 April 2015 "Europe Looks to Tame Web's Economic Risks," *The Wall Street Journal*. http://www.wsj.com/articles/eu-considers-creating-powerful-regulator-to-oversee-web-platforms-1429795918; Tom Fairless, 14 April 2015 "EU Digital Chief Urges Regulation to Nurture European Internet Platforms" *The Wall Street Journal,* http://www.wsj.com/articles/eu-digital-chief-urges-regulation-to-nurture-european-internet-platforms-1429009201.

19 The literature on the economics of two-sided markets is now extensive. A good summary of the basic issues around market creation, competition and pricing is Rochet and Tirole (2006), 645–667.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

which trade theory has tried to grapple with the question can provide rough templates for how to think about the issue.

A basic IMF publication summarizes the mainstream consensus view on current account in these terms:

> Does it matter how long a country runs a current account deficit? When a country runs a current account deficit, it is building up liabilities to the rest of the world that are financed by flows in the financial account. Eventually, these need to be paid back. Common sense suggests that if a country fritters away its borrowed foreign funds in spending that yields no long-term productive gains, then its ability to repay—its basic solvency—might come into question. This is because solvency requires that the country be willing and able to (eventually) generate sufficient current account surpluses to repay what it has borrowed. Therefore, whether a country should run a current account deficit (borrow more) depends on the extent of its foreign liabilities (its external debt) and on whether the borrowing will be financing investment that has a higher marginal product than the interest rate (or rate of return) the country has to pay on its foreign liabilities.

This puts a number of issues on the table that are relevant to data. First is the issue of time frames and sequencing. The question is not whether countries can afford to run current account deficits, because they obviously can. The question is for how long, at what magnitude, and for what purpose is the deficit being used?

In traditional flows that translate into conventional metrics (current account deficit as percent of GDP, for example) there is, empirically, a short term issue of concern: many countries have experienced rapid and sharp reversals of external financing for their current accounts, which spawn financial crises (the 1997 Asian Financial Crisis is an iconic example).[20] That wouldn't seem a risk in data flows per se (except and unless the data imbalance was profound enough and monetized enough that it manifested as a traditional current account deficit crisis).

The more important concern with regard to data is over the long term economic development effects of a persistent imbalance. The logic would proceed along these lines. Country X exports much of its "raw" data to the United States, where the data serves as input to the business models of intermediation platform businesses. Intermediation platform businesses domiciled in the United States use the "imported" data as inputs along with other data (domestic, and other imports from Countries Y and Z) to create value-added data products. These might be algorithms that tell farmers precisely when and where to plant a crop for top efficiency; business process re-engineering ideas; health care protocols; annotated maps; consumer predictive analytics; insights about how a government policy actually affects behavior of firms or individuals (these are just the beginning of what is

20 Ghosh, Ostry, and Qureshi, 2016.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

possible). These value-added data products are then exported from United States platform businesses back to Country X.

Because the value-add in these data products is high, so are the prices (relative to the prices of raw data). Because there is no domestic competition in Country X that can create equivalent products, there's little competition. Because many of these data products are going to be deeply desired by customers in Country X, there's a ready constituency within Country X to lobby against "import restrictions" or "tariffs." And unless there's a compelling path by which Country X can kick-start and/or accelerate the development of its own domestic competitors to U.S. platform businesses, there may seem little point to doing anything about this imbalance.

Here's a concrete example of how this might manifest in practice. Imagine that a large number of Parisians use Uber on a regular basis to find their way around the city.[21] Each passenger pays Uber a fee for her ride. Most of that *money* goes to the Uber driver in Paris. Uber itself takes a cut, but it's not the money flow that's under consideration here. Focus instead on the data flow that Uber receives from all its Parisian "customers" (best thought of here as including both "sides" of the two-sided market; that is, Uber drivers and passengers are both customers in this simple model).

Each Uber ride in Paris produces a quanta of raw data—for example about traffic patterns, or about where people are going at what times of day—which Uber collects. This mass of raw data, over time and across geographies, is an input to and feeds the further development of Uber's algorithms. These in turn are more than just a support for a better Uber business model (though that effect in and of itself matters because it enhances and accelerates Uber's competitive advantage vis-a-vis traditional taxi companies). Other, more ambitious data products will reveal highly valuable insights about transportation, commerce, life in the city, and potentially much more (what is possible stretches the imagination). Here's a relatively modest consequence: if the Mayor of Paris in 2025 decides that she needs to launch a major re-configuration of public transit in the city to take account of changing travel patterns, who will have the data she'll need to make good decisions? The answer is Uber, and the price for data products that could immediately help determine the optimal Parisian public transit investments would be (justifiably) high.

Stories like these could matter greatly for longer term economic development prospects, particularly if there is a positive feedback loop that creates a tendency toward natural monopolies in data platform businesses. It's easy to see how this

21 Credit for this example belongs to Jesse Goldhammer, who developed it in a conference discussion at Chamonix during December 2015.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

could happen, and hard to see precisely why the process would slow down or reverse at any point. The more data U.S. firms absorb, the faster the improvement in the algorithms that transform raw materials into value-add data products. The better the data products, the higher the penetration of those products into markets around the world. And since data products generate more data as they are used, the greater the character of data imbalance would become over time. More raw data moves from Country X to the US, and more data products move from the United States back to Country X, in a positive feedback loop.

This simple logic doesn't yet take account of the additional complementary growth effects that would further enable and likely accelerate the loop. Probably the most important is human capital. If the most sophisticated data products are being built within U.S. firms, then it becomes much easier to attract the best data scientists and machine learning experts to those companies, where their skills would then accelerate further ahead of would-be competitors in the rest of the world. Other complements (including basic research, venture capital, and other elements of the technology cluster ecosystem) would follow as well. The algorithm economy is almost the epitome of a 'learn by doing' system with spillovers and other cluster economy effects.

No positive feedback loop like this goes on forever. But without a clear argument as to why, when, and how it would diminish or reverse, there's justification for concern about natural monopolies, with real consequences. The potential winners in that game—mostly American-based intermediation platform firms at present—have clear incentives to talk about their business models as if that were not the case, and they do tend to emphasize the reasons why their ability to accumulate sustainable market power would be limited (as discussed earlier, arguments such as multi-homing, low barriers to entry, demand for continued innovation, and the recent evidence of platform businesses losing market dominance very quickly when they fail to innovate). For the potential losers in that game, those counter-arguments are mostly abstract and theoretical, while the tendency to natural monopolies seems real, based in evidence, and current.

That anxious reflex would probably be weaker if it were possible to point to specific compensatory mechanisms outside the business models of the platform firms—"natural" reactions that counterbalance the positive feedback loop of an advanced data economy. But the "normal" compensatory mechanisms that are part and parcel of normal current account imbalances don't translate clearly into the data world. Put differently, there's no natural capital account response that funds the data imbalance per se. And there's no natural currency adjustment mechanism. (In standard current account thinking, a country with significant surplus will see its currency appreciate over time, making imports cheaper and exports more expensive, thus tending to move the overall system at least partially

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

toward a dynamic balance over time). A data imbalance by itself would have neither of these effects—except in so far as the data imbalance causes a traditional current account imbalance, which would then drive a capital account and currency adjustment response in financial flows, not in data flows per se.

Regardless of whether these compensatory mechanisms can themselves bring about sufficient *financial* re-balancing, they don't have any obvious consequences for the *data* imbalance dynamic and the longer-term economic development consequences it would create. It's possible to imagine at the limit a vast preponderance of data-intensive business being concentrated in one or a very few countries. These countries would then own the upside of data-enabled endogenous growth models. They would combine investments in human capital, innovation, and data-derived knowledge to create higher rates of economic growth, along with positive spill-over effects into other sectors.[22] In Paul Romer's parlance, these countries would be advantaged in both making and using ideas.

And they would almost certainly enjoy an even greater and more significant advantage in what Romer called "meta-ideas," which are ideas about how to support the production and transmission of other ideas. What is the best means of managing intellectual property like algorithms and software code? What are the most effective labor market institutions that can support the growth of algorithm-driven labor demand? These are the meta-ideas that can keep the positive feedback loop going, and they are more likely to emerge in countries and societies that are already ahead in the data economy.

The lot of the (semi)-periphery

From the perspective of societies that aren't leaders in the data economy, this argument adds up to a troubling story of persistent development disadvantage. It has echoes of dependencia theory, where the words "core" or "metropole" and "periphery" carry deep ideological as well as analytic significance.[23] The implied causal mechanisms are eerily parallel: raw data flows from a data periphery to a data core; the data 'core' becomes wealthier and smarter at the expense of the data periphery; and the periphery becomes trapped at the bottom rung of the international division of labor. The most important part of the argument is its

22 Romer (1993), 63–91.

23 See Cardoso and Faletto (1979). Dependencia theory also carried with it a set of claims about distorted politics and economy in the periphery that were the result of this process; an assessment of these dynamics for the era of data is beyond the scope of this paper but is an intriguing subject for future work.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

persistence over time, as the periphery becomes less capable of developing an autonomous, dynamic process of technological innovation.[24]

At a more pragmatic level and from the perspective of governments, this argument also suggests the kind of market failure on an international playing field that is believed by many to justify policy intervention, for example to subsidize and protect "infant" domestic competitors. The simplest and least ideologically charged theoretical grounding for such action is "strategic trade theory." This has its contemporary roots in the U.S.-Japan debates of the 1980s, but the basic insight goes back at least to Alexander Hamilton's concern that a raw-material and cash-crop exporting American colonial economy would become stuck in that role by importing finished industrial goods from Britain.[25] Paul Krugman formalized many of these ideas in his work on increasing returns to scale and network effects starting in the late 1970s, under the label "new trade theory."[26]

The policy-relevant proposition here is that the "dependency" relationship need not be permanent or semi-permanent, but can be acted upon by governments. Simply put, path dependent industrial concentrations in the "periphery" can be jump-started by policies that protect and subsidize infant industries in a manner that allows them to achieve self-sustaining growth. The direct cost is a short-term harm to consumers (who suffer from protection as they are unable to access the 'better' imported goods at market prices). But that cost may be worth paying if the long-term benefits are meaningfully more robust local economic development. The principal caveat is the risk of political and bureaucratic dysfunction, where protection becomes a cover for rent-seeking rather than a spur to competitive upgrading. In sum, strategic trade theory allows for interventions that compensate for market failure, but only if those interventions can be properly implemented and themselves protected from "political" failure.

The tone of these arguments should sound familiar to even a casual observer of contemporary debates about intermediation platforms, data localization, and data nationalism. In the popular media, a June 2016 *New York Times* article is broadly representative, arguing that the "frightful five companies—Apple, Amazon, Facebook, Microsoft, and Alphabet, Google's Parent—have created a set of inescapable tech platforms… expansive in their business aims and invincible

24 This is in contrast to the more familiar "catch-up" or "convergence" arguments, which rest on the proposition that it becomes increasingly difficult and expensive for the leading country to advance at the horizon of technological leadership, while it becomes cheaper and easier to 'fast-follow' and converge (at least partially) for countries that are behind in the race.

25 Hamilton (1791).

26 A good overview is in Krugman (1986; 1990).

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

to just about any competition."[27] The article claims that for Americans at least there is one "saving grace: The companies are American." Beyond the non-specific post-Snowden fears of surveillance and spying outside the United States lie more profound anxieties about American hegemony, in economic, "values," and cultural spheres.[28]

In the scholarly literature, Faravelon, Frenot, and Grumbach (FFG) use the phrase "dependency of most countries on foreign platforms" to describe what they see. Their analysis uses data from alexa.com to rank the most visited websites for each country, coding the country from which the firm that sits behind that website is domiciled.[29] The results aren't generally surprising but they are notable in this context for what they suggest about policy intervention.

The largest countries (the United States, China, Russia) have a substantial "home bias" in data traffic much as they do in trade, in part simply because of their size. Smaller countries (France and Britain for example) engage at a higher percentage level with external platform businesses—roughly 70 percent of intermediation platform visits from both countries land at firms outside France and Britain. One Chinese platform (Baidu) is among the top five in the world, although its influence is limited to a small number of other countries. (Simply by virtue of China's size, the biggest Chinese platform(s) will rank highly in global measures even if its use is heavily concentrated in China—and in Baidu's case, among a small number of additional countries.)

The most important findings point to an outsized concentration of intermediation platforms in the United States. Only eleven countries host an influential platform business. The United States hosts thirty-two; China hosts five; a few other countries have just one. The distribution of influence is notable: U.S. platforms have a nearly global reach; Chinese platforms are big players in a small number of countries (below ten); Brazilian platforms are big players in an even smaller number of countries (around five); and the numbers go down from there.

If this is a measure of power in the data economy and traditional language is warranted, it would be fair to say the the United States is the only real global power. China

27 *The New York Times,* 1 June 2016, Farhad Manjoo, "Why the World is Drawing Battle Lines Against American Tech Giants."

28 Ibid. Some observers go further and stretch the concept of "American" hegemony toward "corporate" hegemony, where private firms become the definitive and de facto authoritative source of rules that are traditionally the province of governments.

29 Faravelon et al. (2016). Data from alexa.com is widely used, but it is important to recognize that it is limited and may be subject to non-random bias. Jeanie Oh and I are currently comparing alternative data sources to seek a systematic assessment of bias and a means to improve the metrics.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

is a major regional power along with a few other regional powers like Brazil.[30] And everyone else is influenced, weak, or dependent. For France, as an example, about 22 percent of traffic is with French platform businesses; the remaining 78 percent goes to foreign platform businesses almost all of which are in the United States.

FFG argue that "the dependency of most countries on foreign platforms raises many issues related to trade, sovereignty, security, or even values."[31] These represent a complicated mix of concerns that rise to the fore, individually or collectively, subject to particularly salient events (such as the Snowden revelations), as well as gradually mounting, long-standing concerns about somewhat abstract notions of cultural autonomy (recent Netflix and other media restrictions) and the like. These are legitimate subjects for discussion and possibly for political action. My focus here is, to repeat, more narrow—it is on the medium and longer term economic growth and development trajectories.

So let's engage in a policy thought experiment, first from the perspective of a rich, developed E.U. member state that finds itself in a position like FFG describes for France. If this country (call it Country F) decides that its position in the data economy does indeed place it in a dependent relationship with U.S. platform businesses, and that the risks of a self-reinforcing dependency that traps F in a data periphery role as a low value-add raw material exporter and high-value add data product importer are real, what options present themselves to a policy maker in the capital of F struggling with longer term economic growth prospects?

Options for a developed but data-dependent country

In the broadest sense, Country F has four choices with regard to how it relates to transnational data value chains. It could

1. Join the predominant global value chain that is led by American platforms, and seek to maximize leverage and growth prospects within it, to enable some degree of catch-up.
2. Join a competing value chain, possibly grounded in large Chinese intermediation platform businesses, where catch-up might be easier.

30 An alternative metric, also imperfect but interesting, is to count the share of profits of the fifty biggest listed platforms by country. U.S. firms have more than 80 percent while European firms have only about 5 percent. *The Economist* 28 May 2016, page 57, "Taming the Beasts."

31 Faravelon (2016), page 2.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

3. Diversify the bets and leverage each for better terms against the other, by combining elements of 1 and 2 above.
4. Insulate or disconnect to a meaningful degree from those value chains, and work to create an independent data value chain within F; or perhaps regionally within the European Union of which F is a part.

The analysis of the emerging political economy of data that this paper develops suggests a few important points about how F's policy makers should think about and weigh these choices. I offer these not as comprehensively exhaustive or mutually exclusive strategic options. They do not address tactical moves—as in, how would a country actually go about enacting one of these strategies in practice, what specific policies and decisions would it need to enact.[32] And to repeat once more the limitations of this argument, they do not take real account of other concerns and objectives that the data economy raises with regard to privacy, surveillance, and the like.

The first three strategic options are really variants on one big choice: does joining existing global data value chains point toward an economic and technologically advantageous future? This depends, as I've argued up to now, on whether data platform intermediation offers a development "ladder" to (initially) less developed countries that start out as raw data providers to the platforms with global reach.

The analysis here suggests a healthy dose of skepticism about that prospect. The fundamental argument rests on positive feedback loops that connect raw data "imports," algorithm and other production process improvements in machine learning, and high value data-add exports in a virtuous circle. Virtuous, that is, for the leading economy that accelerates away from the rest in its ability to add value to data and thus in growth. The lack of an equally compelling argument about catch-up mechanisms should be a deep cause of concern.

Can such an argument be plausibly constructed for Country F? Let's try. The data economy catch-up argument would in principle depend on F climbing a development ladder that starts with outsourced lower-value add tasks in the data economy, and climbing it at a faster rate than the leading economies climb from their starting (higher) position.

More concretely, imagine that the leading edge of the data economy sits in the San Francisco Bay Area where the costs of doing business are extremely high. There are discrete, somewhat standardized and lower skilled tasks in the data value chain—cleaning of data sets, for example—that could be outsourced to lower cost locations. Firm T in San Francisco contracts with Firm Y in Country F to provide data cleaning services that prepare large data sets for use in T's models.

32 Daniel Griffin and I are currently working on this next set of arguments for a separate paper.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

Firm Y then acts as a draw for human capital as well as training and investment in certain data skills in its home geography (in F).

The critical question is whether this represents a realistic rung on a climb-able ladder, or just an outsourced location for relatively low value-add and low wage data jobs. Are there opportunities for Firms like Y to move up the value chain? Perhaps there will be for small, niche data products that are of local interest … but for larger scale data products that address global markets the case is much harder to see. Firm Y will be at a huge disadvantage as it lacks access to all the data raw materials that would enable that kind of product development. That is because Firm T back in San Francisco is likely to distribute the outsourced work across multiple geographies, simply to get better terms on the work in the short term and prevent single-supplier hold-up. If Firm T is thinking long term, it would distribute the work in a strategic way, to intentionally limit the ability of anyone in the periphery to attain a critical mass of data that would facilitate catch-up competitors entering T's potential markets.

The government of F might try to push back by passing a law that requires more value-add data processing to take place in-country (inside F). Firm T in San Francisco would most likely respond by moving its data cleaning operations elsewhere, outside of country F. This is an attractive arbitrage play in data more so then ever, because investments in fixed capital for T's outsourcing operations are minimal to zero. And so because there's almost no reason not to move the work elsewhere, F's government is recursively stopped from changing the rules in the first place (if rational expectations hold). (Another arbitrage option would be for Firm T to invest in automation and "re-shore" the work in the form of automated systems which most likely would be located back in the Bay Area.) F, the host government of Firm Y, has very little leverage in this game.

The best fitting analogy here is not the industrial catch-up that took place in "factory Asia" during the 1970s.[33] It is instead to call centers for a company like United Airlines, that are located in a developing economy like India. There's almost no spillover or ladder climbing between an airline call center and a globally competitive national airline; and call center functions can be re-located at low cost and in very short periods of time if the terms of trade change. The burden of proof has to be on those who believe that these kinds of outsourcing arrangements can create plausible catch-up development paths, rather than simply perpetuate and extend the advantage of the leading economies.

33 The clearest and most elegant articulation of this development story (and many others) in relation to global value chains during the twentieth century is in Baldwin (2016).

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

A return to vertical integration?

So where does this leave us conceptually in terms of strategy for development in the new data economy? One answer, I believe, lies in a return to vertical integration tied to a new form of ISI, important substitution "industrialization" for the data world. Counterintuitive as that might seem to start, it's worth noting that that the notion of a "full-stack" company is now back in vogue in a number of conversations about corporate strategy.[34] Does this scale up to national development strategies? There are potentially many reasons for this revival of vertical integration thinking.[35] Data flow and the externalities that accompany it are the most important.

One way to see clearly the logic behind this, is to engage in another simple thought experiment: The political economy of a value chain that leads from a dirty shirt to a clean shirt, with two critical nodes—the laundry detergent and the washing machine.[36]

Just a few years ago, the conventional view of this value chain would label the typical washing machine as a "white good," a mostly undifferentiated product where price is the basis of competition. (A trip to Best Buy or any other home appliance outlet confirms that intuition: GE, Samsung, LG, and a few other manufacturers offer home washing machines that are essentially alike, at nearly identical prices.) The differentiating part of the value chain was the laundry detergent. The chemistry of cleaning a shirt without destroying the fabric is a complicated one and so the intellectual property that goes into the chemical formulations of

34 See "The Full Stack Startup," (Andreesen Horowitz blog) at http://a16z.com/2015/01/22/the-full-stack-startup/; *Forbes,* 26 May 2015, Ryan Craig, "The Full Stack Education Company," at http://www.forbes.com/sites/ryancraig/2015/05/26/the-full-stack-higher-education-company/#24a13126459d; *The Economist,* 13 April 2016, "Keeping It Under Your Hat," at http://www.economist.com/news/business-and-finance/21696911-tech-fashion-old-management-idea-back-vogue-vertical-integration-gets-new.

35 These include but are not limited to: the power of the Apple example; fast moving change at the technological frontier brings back issues of rapid coordination that can be done best inside a firm; the broad notion of owning the "relationship to the customer" in fast changing technology markets; new geopolitical uncertainty/boundaries that could disrupt cross-border value chains. I detail these arguments in my forthcoming book *How To Organize a Global Enterprise,* Harvard University Press 2018.

36 Two nodes for now. There could and likely will be more discrete steps, as automation makes it possible to reduce the human labor still engaged in the clean shirt value chain. I would welcome a robot that put the shirt in the laundry, removed it from the dryer, folded it, and hung it in my closet. Many of the same arguments above about data flow through that part of the value chain would still apply.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

a laundry detergent (which now needs to be environmentally acceptable, effective in cold water, and the like) was the important asset.

Now update this clean shirt value chain to the present or near future, and consider the data flow that is embedded within it. The detergent doesn't produce much data by itself (at least not beyond the sales data created when the detergent is sold off a grocery store shelf). Its intellectual property is static. The interesting action from a data perspective is now *inside* the washing machine—where the detergent, the shirt, the water, and agitation, etc. meet in the act of cleaning, and the intellectual property is placed "into motion."

And that is where the most relevant and valuable data is created. Now it is the sensors in the washing machine collecting that data that become the key node in the value chain. And the data flow off those sensors, becomes the most important asset to "own" if a firm seeks to create new value (a more effective detergent, a better washing machine system, a warning about zippers that are about to fail, almost any other set of value-add services that could be offered to the customer).

At a minimum. this thought experiment suggests a migration of critical value from detergent to washing machine, tracking the migration from traditional intellectual property (detergent chemistry) to data flows that come out of the machine. The logic of competition is now going to be mixed, with value-add happening in both domains. Recognizing that, a firm that wants to dominate might choose to own both nodes so that the bet is robust.

But *even more substantial value might be created in a vertical re-integration due to externalities that link the domains together*. Extensive data from washing machines in the wild will inform the chemical development process for the next detergent formulation and vice versa. Samsung (a washing machine manufacturer) might merge with a laundry detergent manufacturer to internalize, accelerate, and make bi-directional the information flow between the two. Or Proctor and Gamble (the detergent manufacturer) might buy a washing machine company.

Regardless of who leads in strategic vertical re-integration, the logic is the same. If the promise of internalizing the data economy within one firm exceeds the costs of placing these two rather different functions underneath a single firm's administrative structures (and out of "the market") then the move to data also implies a move back to vertical integration, at least for some parts of the value chain.[37]

How might this scale up from a firm's strategy to a national development strategy? Obviously, one way to scale is simply through a concatenation of the strategies of many individual firms. There is another mechanism that relies on a new kind of

37 This is basic transaction cost economics reasoning along the lines developed by Williamson. See for example Williamson (1981), 548–577.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

"ideas" dynamic, an earlier version of which was best explained by Romer.[38] In that prior generation of argument there is consensus on the view that governments should subsidize education in order to improve human capital, because human capital is the key ingredient for the creation and use of ideas.

In the data economy, though, the most important "ideas" depend not just on smart people but also "smart" algorithms. The ideas that are embedded in algorithms sometimes are formalizations of ideas in the smart peoples' heads. But increasingly, the "ideas" in the algorithm are the outcomes of machine learning processes that depend on access to data sets.

In that context, the case for governments to support the creation of those algorithmic ideas is as compelling as the case for government to support education. And that means governments might want to support and subsidize the creation and retention of data, for use at home, in exactly the same way that governments seek to train and retain smart people.[39]

Data science and particularly its subset machine learning can be thought of simply as a quicker, more systematic, and more efficient means of distinguishing useful or productive ideas from non-useful, non-productive ideas. Thus, the simplest development bet would be to say the way you are most likely to get to smart algorithms, is through a combination of smart people *and* lots of data. Why wouldn't a government then act to support the creation and retention of both at home?

A final thought on leapfrogs

This paper focuses mainly on the consequences of data imbalance for growth in developed countries, and in the last section in particular on those that are one or two "steps down" from the leaders of the data economy. What of the less developed countries that are several steps down? I leave a full analysis for further work but it is worth considering here some speculative possibilities. The core of an argument, it seems to me, would have to contend with the more hopeful stories about "leapfrog" development paths. Can low income countries bypass the 1970s–80s development ladder (rooted in low-cost manufacturing) and perhaps even the

38 Romer (1993), 63–91.

39 Governments have within their existing policy toolboxes ways to do this, including but not limited to data localization regulations, tariffs and non-tariff barriers to imported data products, public procurement rules, and so on. If some of these policy moves were to be defended publicly for other reasons (such as privacy) that makes them no less important as development strategies. Using other rationales would be nothing new: governments of course have long justified protectionism on other grounds when it suits them to do so.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

traditional "services" sector, to jump right in on the leading edge of the data economy?

In principle they would do so with the advantage of being unburdened by the fixed investments and political-economic institutions of an earlier growth paradigm. The somewhat trite but still evocative analogy is moving directly to mobile telephones without having to go through a copper wire landline "stage." In this next-phase hopeful story, the leapfrog takes one more jump "beyond" the mobile phone and the services it offers, to land quickly on the data economy where businesses grow out of the data exhaust from the phone as it delivers those first generation services.

This is a hopeful story in part because the manufacturing ladder seems now to be largely cut off for most emerging economies by the phenomenon of "premature deindustrialization."[40] The move to services as a new growth model has been the favored conceptual response, based in large part on the idea that many services (including IT and finance) are tradable, high productivity sectors (at least in principle) and could substitute for manufacturing in the growth story.

But keep in mind three constraints. These service industries are more skill-intensive than most manufacturing. They do not generally have the capacity to employ large numbers of rural to urban migrants.

Most important for this paper, these services are themselves being leapfrogged in value-add by the data economy that sits above them in the "stack." The disadvantages that second tier data economies bring to the competition here and that have been the major subject of this paper are likely going to be magnified in the case of less-developed third tier countries.

In concrete terms, who is most likely to build high value-add data products from the data exhaust coming off of mobile phones and other connected devices in less developed countries? If those data exhausts are collected primarily by U.S.-based companies, the answer is self-evident and over-determined. The pushback by the Indian government against Facebook's "free basics" model notwithstanding, it is an entirely rational move for U.S.-based platform businesses to structure their relationships with developing countries to support their own growth in this fashion.[41]

None of this is good news for the third tier developing countries trying to find routes to growth and economic convergence in the data era.

40 Rodrik (2015).

41 The Indian government's decision and justification is at http://trai.gov.in/WriteReadData/PressRealease/Document/Press_Release_No_13%20.pdf A good journalistic account of the key events is *The Guardian* 12 May 2016, Rahul Bhatia, "The Inside Story of Facebook's Biggest Setback," at https://www.theguardian.com/technology/2016/may/12/facebook-free-basics-india-zuckerberg.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

References

Baldwin, Richard. 2011. "Trade and Industrialization after Globalization's Second Unbundling: How Building and Joining a Supply Chain Are Different and Why It Matters." NBER Working Paper 17716. Available at http://www.nber.org/papers/w17716.

Baldwin, Richard. 2016. *The Great Convergence: Information Technology and the New Globalization*. Cambridge, MA: Belknap Press of Harvard University Press.

Cardoso, Fernando H. and Enzo Faletto. 1979. *Dependency and Development in Latin America*. Berkley, CA: University of California Press.

Chen, Min, Shiwen Mao, and Yunhao Liu 2014. "Big Data: A Survey." *Mobile Networks and Applications* 19 (2): 171–209.

Chen, Yong, *et al*. 2016. "Big Data Analytics and Big Data Science: A Survey." *Journal of Management Analytics* 3 (1): 1–42.

Cory, Nigel. 2016. "The Worst Innovation Mercantilist Policies of 2015." ITIF, 11 January 2016. Available at https://itif.org/publications/2016/01/11/worst-innovation-mercantilist-policies-2015.

Faravelon, Aurelien, Stephane Frenot and Stephane Grumbach 2016. "Chasing Data in the Intermediation Era: Economy and Security at Stake." *IEEE Security and Privacy Magazine, Economics of Cybersecurity, Part 2*.

Fosso Wambe, Samuel, *et al*. 2015. "How Big Data Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study." *International Journal of Production Economics* 165: 234–246.

Ghosh, Atish R., Jonathan D. Ostry, and Mahvash S. Qureshi. 2016. "When Do Capital Inflow Surges End in Tears?" *American Economic Review* 106 (5).

Hamilton, Alexander. 1791. Report on the Subject of Manufactures presented to the House of Representatives. 5 December 1791 by the Secretary of the Treasury.

Kennedy, Joe. 2015. "Why Internet Platforms Don't Need Special Regulation." The Information Technology and Innovation Foundation (ITIF). 19 October 2015. Available at https://itif.org/publications/2015/10/19/why-internet-platforms-don't-need-special-regulation.

Krugman, Paul. 1991, "The Move Toward Free Trade Zones." In *Policy Implications of Trade and Currency Zones*, symposium sponsored by The Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming. 22–24 August.

Krugman, Paul, ed. 1986. *Strategic Trade Policy and the New International Economics*. Cambridge, MA: Massachusetts Institute of Technology Press.

Krugman, Paul. 1990. *Rethinking International Trade*. Cambridge, MA: Massachusetts Institute of Technology Press.

Krugman, Paul. 1992 "Toward a Counter Counter-Revolution in Development Theory." *World Bank Economic Review* 6 (suppl 1): 15–38.

Lindauer, David L. and Lant Prichett 2002. "What's the Big Idea? Third Generation Policies for Economic Growth." *Economia* 3: 1–39.

Manyika, James, Susan Lund, Jacques Bughin, Jonathan Woetzel, Kalin Stamenov, and Dhruv Dhingra. 2016 "Digital Globalization: The New Era of Global Flows." McKinsey Global Institute, exhibit E1 page 3.

MGI. 2016a. *Globalization: The New Era of Global Flows*. San Francisco.

MGI. 2016b. *Globalization: The New Era of Global Flows. Exhibit E2*, San Francisco.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press

Rochet, Jean-Charles and Jean Tirole. 2006. "Two Sided Markets: A Progress Report." *Rand Journal of Economics* 37 (3): 645–667.

Rodrik, Dani. 2005. "Growth Strategies." In *Handbook of Economic Growth Volume 1*, edited by Philipe Again and Steven Durlauf, chapter 14.

Rodrik, Dani. 2015. "Premature Deindustrialziation." IAS School off Social Sciences Economic Working Papers Number 107, January.

Romer, Paul M. 1993. "Two Strategies for Economic Development: Using Ideas and Producing Ideas." Proceedings of the World Bank Annual Conference on Development Economics, published in supplement to the World Bank Economic Review, March 1993, 63–91.

Weber, Steven. 2005. *The Success of Open Source*. Cambridge, MA: Harvard University Press.

Williamson, Oliver E. 1981. "The Economics of Organization: The Transaction Cost Approach," *The American Journal of Sociology* 87 (3): 548–577.

Zysman, John, Eileen Doherty, and Andrew Schwartz. 1996. "Tales From the 'Global' Economy: Cross National Production Networks and the Re-organization of the European Economy." BRIE Working Paper 83 June.

https://doi.org/10.1017/bap.2017.3 Published online by Cambridge University Press