REGIONAL CONTEXT AND GLOBAL TRADE

Michael Storper
Professor, Sciences Po, LSE, UCLA
michael.storper@sciences-po.fr, m.storper@lse.ac.uk

This version: July 29, 2008

Forthcoming: Economic Geography, January 2009

Abstract

How should we think of the role of regions in relationship to the global economy? Theory has surprising gaps when it comes to building a unified vision of these two scales of development. Two contributions to such a vision are proposed in this paper. First, the relationship between geographical concentration and the regional economic specialization it underpins, and globalization, should be theorized as a dynamic process. Standard location and trade theory is not adequate for this task; instead, the dynamical relationship can be captured through growth theory. But this in turn requires correcting growth theory to separate its local and its global components, respectively Marshall-Arrow from Romer externalities. Second, we consider the missing element in all theories of geographical concentration and locally-specialized development, an element labeled “context” here. A theory of context in turn raises important new questions about the dynamic welfare and developmental effects of contemporary processes of fragmenting and re-locating production at a global scale.
JEL: D62, F12, F15, R12

---

An earlier version of this paper was delivered as the Roepke Memorial Lecture at the AAG Annual Meeting in Boston, March 2008, as well as at the DRUID Summer Conference in Copenhagen (June 2008). This version benefited from the comments made at both those meetings. I would particularly like to thank Gilles Duranton, Anders Malmberg, Jim Love and Stefano Breschi for acting as discussant on the paper, and Peter Maskell for including it on the DRUID program. Yuko Aoyama encouraged me to prepare this paper for publication, and the editors of Economic Geography provided many useful points for clarifying the argument. Markus Perkmann and Ewald Engelen provided me detailed criticisms that helped in revising the paper, and participants at the Society for Advancement of Socio-Economics in San José, Costa Rica (July 2008) offered a number of stimulating new points that also helped in restructuring the initial argument. Any errors are the sole responsibility of the author.
1. Regional variety and the gains from trade

The combined effects of the division of labor, regional specialization and gains from trade are widely agreed to be one of the two main forces behind world economic growth since its modern take-off around 1820, the other being technological innovation (Mokyr, 1990; North, 2005). Contemporary debates about the geographical reshuffling of output and employment through outsourcing and off-shoring ask whether we have crossed some kind of new threshold in the world division of labor. There are now increasingly fine geographical divisions of labor, not only by broad “functions,” but by “tasks,” that affect regional patterns of specialization in ways that may be unprecedented (Grossman and Rossi-Hansberg, 2006). Does such fragmentation fundamentally alter the process of economic development, so that it is no longer meaningful to speak of local and regional economies as functioning systems? Are we on the verge of replacing them with a radically new type of geographically-distributed system of organizing production and trade?

In this paper, we propose elements of how to think about these questions, asking both how they change the functional dimensions of the economy, as well as its welfare outcomes. Trade theory is confident about the potential welfare effects of a great transformation of the geography of the economy: there may be new kinds of adjustment costs from recomposition of local and national economies, but there will be gains to specialization, resulting from the process of fragmentation and trade, for the world economy as a whole. A dynamic extension of this notion can be found in political economy: that the “whip” of inter-place competition increases incentives to firms in different places to become more efficient over time (Wolf, 2001). These effects hold even
when a particular country or region has historically benefited from strong localization effects: according to trade theory, by definition, the geographically-distributed systems that replace local clusters will be more globally efficient and hence generate overall welfare gains, though certain localities may not be winners at all times. Formally, models of the gains to specialization and trade are based on optimizing locations so as to maximize aggregate productivity gains, with technology held constant. External shocks (technological change among them) can then change the optimal geographical pattern.

Development economists have nonetheless identified some highly-contested exceptions to the assumption that the overall level of geographical concentration and diffusion is always and everywhere optimal: economies of scale create advantages for bigger markets and penalize smaller ones when trade costs are high; timing of entry is important, so that early movers can “fill up” the available slots for agglomeration and specialization, especially when economies of scale are strongly positive; sequencing is important, in that late entrants or big entrants can suffer deteriorating terms of trade; and infant industries are generally not diffused to less developed places, but require local resources and originality (Bardhan, 1971; Bruton, 1998). Thus, there is not enough diffusion or delocalization, not too much. As a result, less developed countries and regions have insufficient opportunities to get into the world economy, and convergence through the globalization of production faces significant barriers (in addition to those from imperfections in capital markets and differences in institutions) (cf. Helpman, 2005). Notice that this critique comes from a dynamical view, rather than the optimization problem derived from standard trade and location theory.
But there are some dynamic perspectives on localization-globalization that have flaws as well. The most prominent of these is an idea that has influenced much regional and national development policy. It consists in claiming that the more the supply chain is localized, the more developmental benefits for a locality can be captured over time through expansion of the activity. This idea expresses a fundamental bias against long-distance linkages and commodity chains that are highly fragmented over different territorial jurisdictions. In simple accounting terms, it has some short-term empirical validity, in that the more localized the value chain, the more of it will be captured locally for a given increment of output increase in the activity in question. This way of thinking is fundamentally flawed with respect to economic dynamics, however (Puga and Venables, 1999). There is reciprocity in trade: so if everyone adopts policies to capture these chains, ultimately the mechanism of comparative advantage will be blocked. Protecting local content also frequently generates political economy problems, i.e. signaling through such policies to local firms that they can be lazy about quality and innovation.² The dynamic process of economic development has less to do with how much of an input-output chain is concentrated locally than with whether the ongoing re-composition of local output (through sectoral and activity “succession” and intra-sectoral innovation, quality improvements, and vertical differentiation), as certain activities are fragmented and relocated to other places, is sufficient to increase local productivity levels enough to permit increasing local real relative factor prices over the long-run.³ In other words, long term growth depends on local learning, innovation, and adjustment, not on

² This is the “whip of competition” point made above (Wolf, 2001). However, some of the research on East Asia’s successful economies suggests that this isn’t always the case, as it depends on political governance of firms and markets (Wade, 1990; Amsden, 2001; Rodrik, 2007)
³ That is, if the region were a sovereign currency area, its real equilibrium exchange rate would appreciate over the long run.
mercantilist “capture” of supply chains. But development economics, on one hand, and regional development studies in developing countries, on the other, show that there is nothing automatic about this process. The mere existence of the whip of long-distance competition does not call it forth. It depends on intricate socio-economic processes at the local level, and appropriate broader institutional enabling conditions (such as market rules and property rights, but also certain non-standard institutional innovations; cf Rodrik, 2007).

All in all, then, there remains considerable confusion about how we should conceptualize the relationship of regional economic development to the globalization process: should it be static or dynamic, and if the latter, then what kind of dynamical theory? In this paper, we seek to address the relationship of regional development to globalization – centering on the processes of geographical concentration, fragmentation, and trade – in two principal ways. First, we show that the standard approach to gains to trade should be complemented by a dynamic, evolutionary view, rooted in growth theory. To do so, however, the geographical dimension of endogenous growth theory needs to be sorted out, which is what we attempt to do in section 2. The second principal task is then to delve back down into the regional level and to ask which types of fragmentation, de-localization and trade are likely to generate dynamic welfare gains for the world economy. In this vein, our second main argument will be that the effects of decisions to fragment and relocate production are not just manifest in terms of outputs and productivity levels – they also involve the creation, loss and change of contexts. A theory of context is proposed in section 3. Contexts affect future development potential as well as current performance. Decisions about the organization and geography of production
may therefore have unintended long-term welfare effects. This is a particularly important contemporary question, if globalization is bringing about the replacement of traditional regional contexts by territorially-distributed “global” contexts for economic action.

2. Why we need geographical fragmentation and trade: there are no “MAR externalities”

As noted, the case for trade rests on productivity gains from comparative advantage and dynamic effects of competition due to greater openness. There is another way to see the positive effects of specialization and trade, however. It comes from growth theory rather than location theory, and centers on the geography of positive externalities and increasing returns. The Romer endogenous growth model establishes economy-wide increasing returns as the principal source of long-run economic growth under resource constraints (Romer, 1990). Knowledge and technology are non-rival and generally only weakly excludable over time; hence they can be infinitely re-used without loss and tend to spread their effects among communities of users (industries, for example) and geographical areas. These characteristics of knowledge liberate it from the constraint of diminishing returns. It can therefore become a source of limitless total factor productivity growth through change in technique. This point is reconciled with standard theory because the different specific activities to which innovations are applied tend to become perfectly competitive in the long-run (Romer, 1986, 1990, 1994). The link between these two seemingly contradictory processes is that monopoly rents to innovation are bid away through diffusion and entry into each sector, but at the economy-wide level the
recombination and re-use of technology (across firms, sectors, users) creates increasing returns.

Applied to the geography of economic growth, the Romer theory is frequently allied to earlier contributions from Marshall (1920) and Arrow (1962), respectively about technology spillovers at the regional scale (“the secrets of industry are in the air”) and “learning-by-doing.” Both capture key mechanisms of increasing returns: those of re-use, spillover and improvement. Many such Marshall-Arrow processes occur at definite territorial scales (local, regional, national innovation systems and spillovers) (Acs, 2002; Feldman, 1994). There is a lively ongoing debate over whether such externalities are located principally within sectors, due to specialization, or between different sectors in cognate or complementary knowledge fields, with the latter said to give the edge to diversity as a key to dynamic regional development and the former to high levels of specialization (Duranton and Puga, 2002; Jacobs, 1969; Glaeser et al 1992). Notwithstanding these differences, there is widespread adherence to the notion that “Marshall-Arrow-Romer externalities” at the regional scale are a principal source of growth-enhancing increasing return in the economy as a whole.

Yet there is a key contradiction between this claim and a central point of the Romer theory, i.e. that increasing returns can be reconciled with allocative efficiency through perfect competition. If the only source of increasing returns were technologies to which access were restricted to a (regional) community of producers, then in effect these insiders would price their outputs to capture all of the returns to their monopolized

---

4 Since these founding ideas, the literature on the socio-institutional sources of regional positive externalities has expanded to include the advantages of “thick labor markets,” local labor pooling, local rules and conventions, trust and social capital in economic transactions and technological spillovers, and regional systems of innovation.

5 The term “MAR externalities” shows 221,000 listings in Google on 1/25/2008.
innovations, in the form of localized technology or knowledge rents. We know that such rents exist when knowledge is sufficiently complex or un-codified that some kind of proximity (geographical, member of a network, often both) is necessary to gain access to the know-how required to use it properly. That such rents allow regional wages and profits to rise above the economy-wide norm, at least for a time, and hence contribute to (at least temporary) successes in regional development is widely admitted among scholars of regional development and development economics.\(^6\)

But if this were the end of the story, then there would be no way to account for the sustained economic growth the West has experienced since 1820: rather than increasing returns at the economy-wide level, there would simply be long-term accumulation of rents in certain lucky hands. Knowledge would be permanently geographically excludable. What would such a world look like? It depends on two imponderable factors. If knowledge didn’t diffuse “out,” perhaps entry “in” might temper the degree of local monopoly and rent capture. Conditions of entry can be institutionally determined (more or less open-ness), or economic (natural barriers to entry). Another dimension of a world of geographically-excludable technologies is whether competing but substitutable technologies, developed in other regions, bid away the rents of the monopoly region. In this case, however, if one set of rents succeeds the previous set, there would still be major limitations to the diffusion of economic benefits. The geography of rents would be unstable, but they would still establish strong limits to the

---

\(^6\) One major debate between more neoclassical models of economic development and more heterodox ones is how long the rents can be enjoyed by the region or country before they are bid away due to rising costs (and the appreciation of the exchange rate) or by technological change or some other force. Strict neoclassical work emphasizes the return to equilibrium, whereas work based on an evolutionary or “history matters” frame of reference emphasizes long-term perpetuation of advantage to leading areas (countries or regions). The two principal sources of circulation and cumulative advantage are agglomeration economies and successful ongoing learning/regeneration of leading position effects. These are the barriers to convergence that worry so many students of economic development.
diffusion of these benefits. Even in the most optimistic of these scenarios, the defining pillar of the theory of Romer externalities would be broken: the notion of limitless increasing returns through widespread long-term non-excludability of knowledge.

Economic historians reject this vision of geographical monopoly of the economic returns to innovation. Though the spread of technology and knowledge are hardly similar to the fluid process imagined by standard theories (to the chagrin of developing countries and regions), there is still ample evidence that much of the economic development process is driven by sharing and diffusion of technologies (Mokyr, 1990; Helpman, 2004). The reality is somewhere in-between durable monopolization of knowledge and rent-seeking, and easy, seamless sharing. Diffusion is temporally- and geographically-uneven, with rents accruing to some firms and places for some periods of time, and then slowly breaking down, as the potentially non-rival and non-excludable character of technology progressively becomes real. Economy-wide increasing returns are only realized through the geographical- and organizational process that leads from localization and monopoly rents to geographical (hence economic) diffusion of technology and the breakdown of those rents. This point was central to classical studies of the geography of innovation diffusion (Pred and Hagerstrand, 1967) but those scholars did not have the benefit of recent advances in growth theory or the economics of agglomeration.

A major consequence of this point is that, strictly speaking, there are no so-called “MAR externalities.” Rather, there are M-A sources of local technological externalities and possible local monopoly rents; but the true R-sources of economy-wide increasing returns cannot remain localized because of the problem identified above, i.e. that restricted entry would result in monopoly rents and limit the growth process. This does
not imply that localized externalities (henceforth M-A) are inexistent, but that they are part of a wider growth process. M-A and R externalities are linked through (a) the geography of where technologies and technological rents emerge; (b) processes of localized learning and sharing, with rent capture; (c) a process of codification, so that such innovations can be more widely used through entry and recombination, including use by other users in other places; leading to (d) the destruction of technological rents, in turn allowing increasing returns in the entire economy.

A reformulation of this sequence looks as follows. In t1, innovations emerge in certain, specific places and organizational settings (firms). The geography of this innovation often reflects M-A effects of proximity and localized learning and innovation. They earn monopoly rents to such innovations, for a certain amount of time. These monopoly rents last as long as there are barriers to imitation, which include both knowledge barriers and trade/communication costs associated with using the technology – it is effectively, if not juridically, excludable. This notion has been nicely captured by the literatures on the geography of patenting and learning (Henderson, 1999; Jaffe et al., 1993; Malmberg and Maskell, 2006, Feldman, 1994).

In t2, these innovations can diffuse to a wider set of places, firms and uses. The knowledge is more amenable to imitation and copying because as it becomes more widely used, it tends to be codified, allowing more people learn the codes for using it, and hence reducing the trade costs associated with deploying it to other uses and places.

---

7 In the literature on local sources of technological change, reference is also made to “Jacobs externalities” (Jacobs, 1969). The common interpretation of Marshall is that he called our attention to intra-sectoral technological spillovers, thus to the benefits of specialization, whereas Jacobs emphasized diversity and the learning that comes from encountering the unknown, hence the benefits of a non-specialized local economy. However, one finds no reference to Jacobs-Romer externalities, curiously. In any event, our point would hold for such a construction: if there are Jacobs-type sources of technological change, in order to contribute to economy-wide increasing returns they would have to undergo economic and geographical diffusion or de-localization, otherwise they, too, would be captured as localized technology rents.
On the ground, this comes about through geographical fragmentation of production and its de-localization. Schumpeterian competition is replaced by standard (or at least semi-standard) market structures (Aghion and Howitt, 1997). The locational dynamics of this are described by the spatial versions of the product cycle, even though the latter does not have an explicit growth theory dimension (Norton and Rees, 1979).

Simultaneously, geographical diffusion and increasingly non-exclusive application of the knowledge emerge, through its recombination into further rounds of innovation. These drive long-run growth as described by the Romer model.

The process repeats itself over an unlimited number of cycles. The precise parameters for technology-creation, the trade costs and barriers to imitation, and Romer-like diffusion and re-application of technology will determine such things as: the spatial hierarchy of incomes at any given moment; the amount of time it takes for a shift away from the rent-earning (first mover) part of the innovation cycle to economy-wide increasing returns; and, the amount of increasing returns in the economy as a whole. Overlain on all this is the geography of innovation: where additional innovation processes get started and new rounds of monopoly rents are earned, driving favorable local economic development processes, at least for a time.

This also provides elements of thinking about the dynamics of income convergence and divergence. Logically speaking, MA rents are a force for divergence, and the same process that leads to R externalities is a force for convergence, at least over the relevant spaces of R diffusion. Unlike the standard models of trade, however, we are closer to a Schumpeterian perspective which sees both diffusion and tendency toward convergence, and creation with tendency toward divergence, as necessary and inter-
dependent parts of single process (Aghion and Howitt, 1997). The former is not a steady state or point of arrival, but one part of an efficient dynamic which also generates temporary divergence. Inside developed countries, these two processes are in a dynamic and tolerable tension with one another. At a world scale, of course, their relationship is more problematic. The poorest areas of the world enjoy neither the rents from invention/innovation, nor the productivity effects of diffusion. Some developing areas get diffusion, but because they are not in a close enough relationship to “MA regions,” they do not climb up the technology ladder, and as a result their convergence effects with the wealthy parts of the world are minimal. This reframes the classical problem of the world’s technology/knowledge hierarchy in a way that takes greater account of its wider geographical context.

2.1 The “Strong Case” for Globalization

In any case, we have now made a different – and in some ways stronger -- case for the gains to geographical fragmentation and trade than the standard Ricardian/H-O accounts of allocative efficiencies. MA and R externalities are different, but interrelated, moments in the dynamic of long-term economic growth. This directs our attention to the developmental effects of both localization and de-localization, and the temporal- and spatial structure of how they inter-relate.

The heart of such a dynamic view is this link between creation and diffusion of the sources of growth. Once such sources are created, if they are not diffused, the economy-wide increasing returns that make long-term growth possible cannot occur. The process of creation occurs – at least up to this point in history – to a significant extent, at
definite territorial scales, where specific and differentiated socio-economic forms of organization and practice are key to creation. As noted, the mere existence of a “whip” of long-distance competition is insufficient to “call forth” these complex socio-economic systems, even though the process of global liberalization may indeed increase the incentives to local, regional and national innovation. The question then becomes: how well can we explain the sources of such geographically specific- and differentiated innovation and creation of the sources of growth? This question should apply to regions in many different positions relative to the global economic system: regions that are at the technological frontier in their areas of specialization; regions that are near the frontier, where important forms of application of technologies are created, so that frontier regions do not become monopolists and so that there are quality ladders and variety from each basic technology; and regions and nations far from the frontier in any activity, where their underlying specificities and differences might ultimately be harnessed in ways that they could move up the development ladder, and this in ways that simple adherence to global benchmarks and price-cost competition are unlikely to achieve. In other words, how well do we do at explaining these M-A externalities? In what follows, we shall argue that we have quite some distance to go.

3. The “Dark Matter” Problem: Difference in the Economy

M-A externalities exist in places where knowledge is created. They are sources of geographical economic specificity and difference. Theory has struggled with how to explain economic difference. Anthropologists have long asked whether differences in economic organization between societies were simply reflections of great differences in
resources (endowments), which then lead to different power structures or rules (incentives), or whether it is the actors from one place that have different processes of rationality, goal-seeking, learning and cognition (Sahlins, 1995). The anthropology of “primitive” economies uses their relative isolation as an experimental control for this question (“weak contamination” of practices and resources through contact), but has not arrived at consensus about it.

Mainstream economics tends to claim that actors and decision-making processes are the same everywhere, but that preferences, endowments and factor costs are different. Contemporary institutional economics goes a step further. Differences in endowments and initial conditions create different scarcities and collective action dilemmas, leading to construction of different rules and market structures (Acemoglu, Johnson, Robinson, 2004). From this, “history matters,” and can lead to durably different outcomes for economic organization and development.

A third perspective simply acknowledges institutional differences as a fact. Economies are such complex aggregates of microeconomic phenomena that in any given place, even with common overall rules and laws, important differences of practice and history will emerge. These will affect choices and perceptions, and reproduce themselves as ‘varieties of capitalism’ (Hall and Soskice, 2001) ‘embedded firms and production systems,’ and ‘regional worlds’ (Storper, 1997; Storper and Salais, 1993). A now vast literature on localized processes and systems of innovation, localized technological spillovers, firm practices, and local implicit and explicit institutions of industrial policy, has identified other potential sources of such differences (for overviews, see Acs, 2002; Breschi and Malerba, 2005; Polenske, 2007).
A specter nonetheless haunts the literature on regional difference and specificity. Those who ardently defend the vision of a world economy whose innovations come from regional differences, often find that the toolkit referred to just above is somehow not up to the job of fully explaining such differences. There is something that escapes the extensive apparatus of explanations. Laypersons like to call this “culture.” What is it about the Japanese that enables them to do such effective quality control and still to be able to change car models so quickly? What is it about Germans that makes their cars have a distinctive type of engineering? Why can Italians design with such flair? The possible examples are almost infinite. Even when we take into account the enormous progress in economic geography and regional development studies of the past quarter century, we still have great difficulty in analytically deciphering such differences, much less instructing policy-makers how to imitate them in practice. Something must be missing.

From an entirely opposite standpoint, it is held that the problem is not that we have not explained enough about difference, but that we have made too much of it. Arrayed against all of the above is the notion that the conditions for diversity are being decisively eroded in a world of intricate long-distance divisions of labor, trade and interpenetrating markets. In this view, global interaction pipelines make it such that there is no longer enough separation of contexts to allow such differences to persist, except as residues of institutional rigidity, and that they are steadily being benchmarked out of existence (Friedman, 2005).

The first difficulty suggests that there are un-theorized and therefore un-measured sources of territorial-economic difference. If this is the case, then the position of the
second group is likely to prove analytically simplistic (cf. Rodriguez-Pose, 2007; Leamer, 2007), but also questionable on the grounds of its assumption about the welfare effects of all globalization. My proposition – and this is the second main point of the present paper – is that this is the “dark matter” of regional economies and that it should command much greater attention than it does.

3.1 A Story of Context

Let’s consider a light-hearted, but analytically-serious example of the dark matter problem. My franco-american godson has been passionate about food and cooking since he was about ten years old (he is now 22). In the last several years, he has worked in high-level restaurants in Los Angeles, Chicago and New York, and has made good friends with an excellent restaurateur in Paris, and some food critics. We’ve had an ongoing conversation about food quality, cooking, and agriculture. He keeps returning to the conclusion that, though in the best American kitchens there is a great deal of innovation and excellence, it is nonetheless more difficult to get an excellent meal in even the most gastronomic cities in the US than it is in France. In order to get such a meal in the USA, you generally have to go all the way to the top of the food chain, to the “big deal” restaurants; by contrast, there are a lot of excellent but unassuming restaurants in France, and generally at more modest prices than in the US.\footnote{Before the 2007-08 decline in the dollar-euro exchange rate at least.} This point does not challenge the statistics on starred restaurants in the restaurant guides of Paris versus New York, showing that the latter has many world-class restaurants; rather, it is that just below this very top level, there remains a huge difference in the French and American supply
curves for good cooking. A fair guess is that the same would be said of Japan versus the USA (Tokyo has more than double the number of starred restaurants of Paris).

My godson has long thought about trying to open a simple, excellent restaurant in the USA, but has concluded that it would be much more difficult than in France, and perhaps impossible to make it economically viable. What are the differences? First, there is the supply chain for ingredients. This begins with the massive market at Rungis near Orly Airport, for which there is no equivalent in the USA. Perhaps more importantly, that market is complemented by numerous alternative ways to get very high quality ingredients, many of them just by going to the right vendors at the Paris food markets. Though in some US regions, markets have been developing in recent years (such as northern California, the Hudson River Valley), the best ingredients are the province of only the very high-end restaurants, because relative prices for such ingredients are much higher in the US than in France. This leads to the second difference. Because non-star chefs get hold of such ingredients relatively easily in France, there is an inherently lower barrier to entry to excellent cooking than in the USA. Thus, and this is the third difference, to use these ingredients profitably, American restaurants generally compensate by operating at higher scale than their French counterparts.⁹ They are also pushed to higher scale by the different “granularity” of urban structures in the USA, where location and bigger size are essential to capturing a sufficiently big market to offset the higher fixed cost structure they face. Fourth, there tends to be a different production process in the American kitchen. In the generally larger American restaurant,

⁹ Thus, the Paris daily free newspaper, Métro, notes, on December 4, 2007, that the prize for “Fooding” in 2007 goes to Pierre Jancou, with his restaurant Racines, which has only 36 seats, 10 of them outside. His secret: “I search for artisans for vegetables, meat and cheeses, and here we just showcase the products, without chi-chi.”
a greater proportion of the staff will be carrying out orders in a steeper hierarchy than in the excellent, but not-big-deal restaurant in France.\textsuperscript{10}

The point is that a relatively simple restaurant, which operates at modest scale, and offers excellent food, in turn relies on easy access to external inputs, moderate scale of operations, and moderate hierarchy in the kitchen. These features enable it to keep its offerings focused and simple and hence control costs. In the USA, the incentives work in the opposite direction, so this type of restaurant is rare, pushing producers and consumers toward the big, “star quality” restaurant model.

The fifth consequence is that the French restaurant production system has a much larger proportion of excellent, but not “celebrity” restaurants than the USA. This in turn leads to a recursive feedback effect in both countries. In France, there is a bigger, more open and competitive milieu for training cooking talent, and this is an ongoing “nursery” of talent for both the “normally good” and the higher-end restaurants. In the US, this type of training milieu is much thinner. The exception to the US pattern has been perfected by Italians in the US, and it is a telling exception. The average size of their successful restaurants is still higher than their versions at home in Italy, but they are not huge affairs like the multi-starred restaurants. The Italian success comes from relying on the clan method: bringing in virtually the entire staff (often with the family at its core, plus investors from the home town) from Italy, sometimes (but more rarely) down to the waiters, but often including the architects and business managers.\textsuperscript{11} In essence, rather than drawing from a local context in the USA, they import it, which the French generally don’t do.

\textsuperscript{10} In the highly prestigious restaurants in France, of course, there’s a very formal division of labor in the kitchen, very rigidly codified by education and practice.
\textsuperscript{11} Thanks to Monica Viarengo of Farina in San Francisco for explaining the Italian “method” to me.
The sixth consequence has to do with the type of innovation carried out in the two places. According to my godson, while the top American restaurants innovate in the creation of new dishes, the innovations in fundamentals of technique, ingredients, and focused sauces and flavors, still mostly come from France, Italy and Spain. This is because the basic innovations are about artistry, and in this, the division of labor between conception and execution found in top American restaurants is a disadvantage. Both systems innovate, but they do so along different trajectories.

That such differences should persist is curious, given strong international knowledge spillovers in cooking and the restaurant industry. In recent years, a worldwide system and culture of gastronomic restaurants has come into being, involving the mobility of labor (celebrity chefs); knowledge (cooking schools, books, magazines); and capital (investors and owners). A sophisticated clientele travels and eats and compares. There are long-distance networks and collaboration, as well as a few truly multinational chefs/restaurateurs. In other words, the story does not seem to be one of knowledge that is so un-codified that it cannot travel. Increasingly, the knowledge does travel via more and more institutionalized channels that connect far-flung areas. That knowledge is then used differently when it arrives in different places. The causes of such differences cannot be reduced to “history, culture and ways of life,” either. The processes of learning, information exchange, and factor mobility are able to alter history, culture and ways of life. In this sense, the cultures of a certain population of gastronomes in both France and the USA have already been mutually influential to a high degree, but production systems have converged to a much lesser extent.
This probably seems like a quaint story, but the same type of account could apply far and wide across the contemporary economy. Even in markets where products aim to be substitutable, German cars remain resolutely different from their American counterparts; French service firms deliver mass-produced consumer and infrastructure services systematically differently from the way American firms do so. In other markets, there is a deep and recursive relationship between product qualities and production systems, as in our restaurant example: Hollywood and Bollywood, Milan fashion versus New York fashion; Danish food processing versus French or American food processing.

In what follows, we will theoretically re-interpret the issues exemplified by the restaurant industry case. First, we will argue that the various differences in the two restaurant production systems described above can only be captured fully by introducing the theoretical notion of “context,” which is a theoretical concept proposed to address the “dark matter” problem. Second, we will ask whether new forms of geographically-distributed production systems represent a “great transformation” of the role for context in the economy. In this perspective, are the Paris and Tokyo restaurant economies mere vestiges of the past, about to be replaced by a global model based on the New York system, which can be duplicated via benchmarking in Chicago, Singapore, and Dubai? Third, we will return to place these ideas about context in the wider framework of the welfare effects of the local in relationship to the global process of delocalization and integration.
3.2 Defining Context

Psychologists who work on the economy provide analytics that enable us to see difference neither as a simple aggregation phenomenon due to different endowments, preferences or transaction costs, nor as simply “cultural.” They start by challenging the notion that actors have universal, powerful and rational cognition (Ariely, 2008). Experimental evidence confirms that choices are typically made with very imperfect information and on the basis of a set of limited criteria, notably “local” points of reference, or “anchors.” Typical worlds of economic decision-making are not those of ordered preferences, obtaining complete information and rationally weighing alternatives, but rather where participants have widespread and persistent cognitive errors and biases.

The unifying theme in behavioral economics is, thus, *situationalism* – the idea that decisions are always strongly conditioned by anchors, cues, frameworks, and limited information (Glaeser, 2003, from Ross and Nisbet, 1991). Some psychologists argue that these features of human behavior have strong evolutionary roots (Heselton et al, 2005).

Five dimensions of the situation of the actor can be highlighted. First, valuable specialized information is not uniformly available. There are costs to obtaining it, and there are also barriers to access – one’s social and economic position define whether and under what conditions, and sometimes at what cost, one can get such information. In productive activity, information needs and access are defined by place in the division of labor. Divisions of labor are dizzyingly complex in the 21st century. Most people don’t have an overall vision of the productive chain in which they work, not to mention the millions of other specialized systems that affect their lives as workers, consumers and citizens. This may be the beauty of the division of labor, i.e. that we don’t need to know
these things in order to benefit from the productive power unleashed by such complex structures.\textsuperscript{12} But they also mean that what we know is partial. Hence, the situation that we define for ourselves is a “local” – in the sense of partial -- one. Many economists would reply that this isn’t a problem, since we can depend on prices to reveal – even if only after-the-fact – the value of what we do not know and hence to correct the problem. But the assumption that we all know the precise prices for each type of potentially relevant information, and can separate them from all the prices for everything else, is unreasonable. Moreover, this type of thinking is challenged by prospect theory, which holds that people engage in very limited search behavior, and put enormous weight on limited reference points, even when these are quite arbitrary and ephemeral (Kahneman and Tversky, 1979). This is the second dimension of the actor’s situation.

Social and economic networks also underpin access to information and knowing what information to look for (Powell, 1990). Networks have been shown to affect opportunities afforded to individuals in ways that are only loosely related to underlying capacities or talents, so certain types of markets don’t clear optimally. A different point can be inferred from this established finding: through the way that they structure our string of opportunities and roles, networks affect who we become and hence how we frame preferences and choices. Thus, networks are a third dimension of the situation of most actors.

A fourth dimension of situationalism is how actors form their goals. In the place of fully subjective preferences, on the one hand, or rationally-constructed ones on the other, an enormous body of evidence shows that goals are strongly influenced by comparison and emulation. Status comparison is a big element of this, but it affects

\textsuperscript{12} Known in economics, of course, as the “first welfare theorem.”
virtually every dimension of preference formation. Since information is limited and prospecting is the norm, what we compare to and emulate is not an exercise in global maximizing, but in many ways is highly dependent on our social and geographical positions in networks, and on our role in the division of labor (Frank, 2001; Fine, 2006). Detailed studies show that even well-travelled economic elites “think differently” and have different goals (Lamont, 1992).

Fifth, the information we do access is generally processed in a relatively narrow way. Mental accounting shows that people mostly make decisions by ignoring events and consequences outside of a particular narrow domain (Thaler, 1985), and most decisions are made using rules of thumb that are far away from the processes that would be necessary to maximize.\(^\text{13}\) Even such “tricks” as “reframing” a situation can induce big changes in choice behavior, as can the way that local stimuli trigger emotions (Romer, 2000; Ariely, 2008). People are inherently conservative in their decisions, as shown by the endowment effect, that they want to keep what they have over almost any other goal (Thaler, 1994). They are also averse to small-value risky gambles, and are often vengeful (Fehr and Gachter, 2000). None of this means that decisions are merely the unleashing of emotions, because economic actors may be able to learn to manage, interpret and manipulate their own emotions (Gul and Pesendorfer, 2001); but they cannot cut around them and construct a non-situational world for themselves (Gilovich, Griffin and Kahneman, 2002).

Theories of corporate bureaucracy vacillate between the idea that organizational structure can correct the influence of the situation, and another, that such bureaucracies

\(^{13}\) See: [www.rulesofthumb.com](http://www.rulesofthumb.com)
are themselves situations, and they do not produce standard cognitive rationality (Simon, 1986; Friedberg, 1993). Theories of entrepreneurship do the same: from those that argue that the entrepreneur rationally arbitrages unexploited opportunities to those that hold that s/he leaps into the void with courage and flair. In the end, it would seem that the very structures set up to control our worst cognitive biases and errors end up creating new situations.

We can now propose a definition of “context.” Its structural component is defined by the division of labor and the networks in which the actor finds herself, which has a decisive influence on the informational environment for the individual, hence her “input” structure of cues and reference points. In turn, the individual actor engages in her search behavior (prospecting) and goal formation (emulation/aspirations), leading to choice/evaluation behaviors (strongly influenced by rules of thumb and framing that unfold in both network and geographical environments).

The notion of “context” has affinity with our earlier notion of a “world,” which is a collective environment defined by the conventional ways actors coordinate with one another to reduce uncertainty. Those conventions, in effect, collectively stabilize many of the anchors and frameworks for individual action in a given type of environment (Storper and Salais, 1993). However, behavioral economics provides the micro-foundations that are missing in the literature on conventions.

4. Geographies of Contexts

Actors are not “programmed” by their situations; this would gravely underestimate the complexity of motivations, the ways that preferences combine and
interact, and the capacity of actors to reason. Piore (1995) argues, in this regard, that the meaning actors make of their situation depends on a process of interpretation, itself in part dependent on the interactions that occur in the situation. Time is a critical aspect of this – a string of interactions and interpretations, rather than a set of instantaneous and history-less discrete calculated choices. To that notion, we can add the importance of the geographical scales of such interactions, because the string of frames and cues, and processes of emulation, that emerge in specific geographical contexts (local and far-away), logically have an influence on the making of meaning and future choices. Up to this point in history, at least some contexts are unavoidably and importantly geographically local, though actors can mix stimuli from other places through travel and communications.

The economy’s intricate organizational and geographical structure shapes the actor’s situation along several dimensions: the immediate environment of her task in the detailed division of labor; her environment in the work unit; the environment between closely-related work units, possibly inside an organization (e.g. a firm); intended regular relationships to the environment (intra- or inter-organizational or market); and unintended or unplanned relations to the external world. Thus, some resulting environments are more specialized and homogeneous, while other pieces of the division of labor are located in places with diverse, more heterogeneous local activity mixes. All these relationships are shaped by geographical location, which is in turn strongly related to productive organization, i.e. the division of labor and the trade costs between parts of it.

A way to begin to see this is to extend the standard logic of scale and scope to the structure of the actor’s situation. The more organizationally internalized the actor’s
relationships, combined with the effects of the internal division of labor in an organization, the more an actor’s context is intra-organizational and, possibly, task-specialized. In turn, this will direct her communication within the organized chain and tend to simplify communications to the local, external environment. At the other extreme, shallow or artisanal divisions of labor and less “purified” task definitions will depend on more diverse, irregular, and uncertain external communications. Locational economics suggests the latter will tend to be located in more “diverse” and urban local economies, the former in more specialized and less urban ones, all other things being equal (Duranton and Puga, 2002).

This is just a simplified vision of two ideal-types at opposite ends of a spectrum. There is rarely a clean division between “locationally-fragmented, highly organized, specialized contexts” and “highly diverse, market-oriented, dense communication contexts,” but rather some fascinating mixes of them. Thus, in a local context such as a city, we can find firms involved in high levels of internal communication, using algorithms and rules, and highly-organized long-distance professional supply-chain relationships, as well as local, spontaneous, diverse interactions. The architecture of the actor’s context, in this case, is some mixture of the two, which is as yet rather poorly understood. Likewise, in the geographically-fragmented, highly organized supply chain environment, there is still likely to be some leakage of unplanned information from the local or long-distance environment. The core issue is how the information and signals that are the key inputs into the actor’s situation are defined and channeled, and then what actors do in different types of situations: how they frame and anchor, whom and what they emulate, how they choose, and what they learn.
Let us now think about some examples of this phenomenon. The apparel industry has a high level of product differentiation, both because of functional and fashion diversity, as well as its steep quality ladders. Market structures, organization, and geography are correspondingly diverse within this industry, which exhibits everything from geographically-fragmented chains of mass production within large firms to highly-localized, specialized clothing “clusters” or districts. The processes of emulation, learning and innovation are very different in the many contexts defined by such a variety of organizational forms. Skills do not transfer readily between them. There is “spillover” from one to another, in that mass production tries generally to knock-off and copy what comes from fast-fashion or high-fashion, which in turn learns from an alternative circuit of cities, the arts, and the “street.” So in this case, there is probably some kind of complementary relationship between two very different organizational and geographical contexts.

Another example comes from the literature on organizational diversity within capitalism (Sabel and Zeitlin, 1985; Hall and Soskice, 2001; Scranton, 1991). German, Japanese and American carmakers have significant differences in the organization and geography of their core activities; so do companies that make consumer electronics. Unlike the clothing industry, the products here are roughly comparable, so we are confronted with alternative organizational and geographical arrangements for substitutable (even if imperfectly) products (Krugman, 1991).\textsuperscript{14} This might be due to the increasing market interpenetration of these industries and the international technology spillovers it facilitates and requires.

\textsuperscript{14} This is not the same idea as imperfect substitution in quality ladders, but competition within the rungs of long quality ladders.
This might tempt some to conclude that context, in the way that we are defining it, is unimportant, since apparently it leads to the same ending point (substitutable products). But upon closer examination, these industries struggle with dynamic change in different ways and with varying degrees of success. No close observer would claim that Detroit, with its far-flung system of suppliers and arms-length relations, does as well as its German and Japanese counterparts in getting out new models with constant engineering improvements (Goldberg, 1995). Thirty years of competition between Detroit and Tokyo has left them with distinctively different products, and Detroit has failed to imitate Tokyo successfully even when it has announced its intention to do so. Even in a world of enormous technology sharing and information circulation, the actors in each system appear to be keyed into differences in information, different emulation dynamics, possibly discounting (time horizons), and choice behaviors. They seem to frame problems differently (Whitley, 2004). If not, universally-available information, world market structures, and international technological spillovers would push all three toward more organizational and geographical and performance convergence than in fact we observe; this recalls our restaurant example.

A third type of case is that of winner-take-all systems in the world, such as the City of London, Silicon Valley, or Hollywood, or industrial districts built around a combination of distinctive products and locally-constructed techniques rooted in the local system as a whole (Becattini, 1993). These are systems that are not only at the technological frontier in their respective activities, but where that frontier is occupied by a very small number of systems. These highly-successful clusters serve the global market and are caught up in all sorts of long-distance, formal organizational procedures,
professional and regulatory norms, and so on. Some kind of highly contextual process occurs in their core regions, extended through the local and world-wide informal and semi-formal networks at whose center they lie.

5. Distributed context: a great transformation?

Major changes are occurring in the feasibility and costs of managing long-distance relationships in the economy, and hence in the possibilities for fragmenting production, organizationally and geographically. Are we at the brink of a “great transformation” in local versus long-distance relations (Polanyi, 1944)? Such a transformation toward what may be termed “distributed contexts” would consist of four major dimensions. First, previously-existing borders between using intra-organizational hierarchies and external markets to coordinate activity would be redefined by the ways new technologies facilitate coordination and monitoring. In production, the big vertically-integrated producers who were once necessary to coordinate complex transactions and eliminate bilateral hold-up risks in production, appear increasingly to be giving way to networked “heterogeneous” production. The notion of a great transformation assumes the “completion” of this process. Second, the role of geographical distance would have a lesser role, even for the most complex intra- and inter-organizational relations. In the past, vertical integration carried with it very big distance costs. The same forces that reduce the need for integration also reduce the need for proximity: much more efficient communication technologies that permit complex monitoring and contracting at great distance (Baldwin, 2006), and permit the fragmentation of production systems at the fine-grained level of tasks, rather than subsystems (Grossman and Rossi-Hansberg, 2006).
Third, and underlying the first two, the boundary between formal and informal processes of coordination, contracting and monitoring would be profoundly modified. It would be possible to “quasi-formalize,” thus combining ease of deal-making with long distance and flexibility – the “best of both worlds” – and greatly reducing the relevance of existing trade-offs between formal organization and long-distance, and flexibility.

Fourth, information would be obtained via “global pipelines” (Bathelt, Malmberg and Maskell, 2004). Such distributed contexts would be more organizational than geographical, and the organizational basis would be radically different from the past, owing to the possibility of managing complex interactions via the radically increased intermediation of information technologies. The five dimensions of the actor’s situation would be radically altered for the actors in distributed contexts.

What we have just laid out is speculative. Much of the literature counters this vision, or at least parts of it, by asserting the ongoing (even heightened) complexity of deal-making, increasing and more volatile output environments, thus requiring face-to-face contact, and social networks to carry out production (Storper and Venables, 2004; Olson and Olson, 2000). Communities of practice remain central in both visions of things, but whereas in the first they tend to be core “nodes” of geographically-extended production systems, in this alternative view that is not necessarily the case (Amin and Roberts, 2008).

5.1 Some economic aspects of distributed contexts: welfare effects again

Trade theory teaches us to think of the possible advent of distributed contexts as a natural and unproblematic dimension of economic development, with benefits for
outlying regions in particular, and for the economy as a whole through technological maturity and Ricardian gains to trade. But in the case of such radical restructuring and change in context, how sure can we be about these presumed welfare effects? The Romer growth theory was offered earlier as the strong case for such positive welfare effects, through economy-wide increasing returns. But, viewed in the light of a theory of context as the basis for discernably different economic behaviors, a different type of question is asked: how does the change in context alter the framing, emulation, choice, and learning behavior of actors, and hence the developmental dynamic of the sector in question? If Hollywood were a global production network with no local core, would it be the same industry as with its present organization and geography? A producer of different, but better outputs? A producer of worse, less welfare-enhancing outputs? A fair guess is that the situations of its actors would have changed, and with them, the processes of framing, emulation, cognition, learning and choice. Globalization through de-localization is not just a matter of choosing alternative techniques, then, but that such a change in techniques may have unplanned effects on the productive environment and hence on the industry’s dynamics – what its people know, what they want to do, and what they ultimately do (Martin and Sunley, 2006; Boschma and Lambooy, 1999; Jaffe et al, 1993).

As noted, in mainstream trade theory, any such questions disappear axiomatically. If context is more than just a residual effect of optimal choices of organization and location, then even an efficient process of fragmenting and geographically redistributing a production chain may have unaccounted future effects. These decisions to change a production system can alter future patterns of emulation, cognition, learning and choice. As suggested in our restaurant example, the resulting products are imperfect substitutes,
so that there is not just a change in production technique, but in the envelope of outputs. In other words, the decision to restructure a production chain involves not just discrete choice processes today, but “roads not taken” (Arthur, 1989, 1994; Hodgson, 1998; Essletzbichler and Rigby, 2004; Dosi, 1998). We can compare the road taken to the existing road, but what we cannot measure is the possible qualities and quantities of future outputs of contexts that are being eliminated or transformed by today’s actions. The creation or elimination of contexts is largely an unintended outcome of decisions made about the organization of production and location, with these dynamic effects not figured into the decision-making process or priced in any way. They are developmental externalities of the globalization process.

5.2 Globalization, the Diversity of Contexts, and Development

The emergence of very large-scale distributed contexts might be reducing certain dimensions of competition among territorially-anchored contexts, in that the major feature of the new context for economic action becomes the spatially-extended network. In this case, there might be enhanced “within technique” competition among places, but within a world economy generating significant institutional convergence between production systems at a global level (single best practices), such competition is more about how places deliver up resources to the global network context, hence reducing the importance of local or national extra-organizational context in generating a variety of techniques and products (Gertler, 2001). The information used by actors would be “stovepiped” via the global division of labor. This information would increasingly take precedence over information coming from outside such a network (e.g. from a local or
national context of information exchange) in the key processes of emulation, cognition and learning. Certain formal models of trade with heterogeneous firms predict precisely this reduction in variety (Baldwin and Forslid, 2006).

This brings us back to issues of long-term economic development. One key perspective on why the modern economic take-off occurred in Western Europe argues that it was the combination of a balance between inter-relationship – the dense trading and political relationships that emerged during the mercantile period – with separation, difference, and specificity – the proliferation of different technological systems and their products – that promoted competition and overcame progress-killing local monopolies (North, 2005). Standard theories of globalization argue that a similar process is occurring today, at the global rather than continental scale. But if global distributed contexts are emerging and reducing such local and national environments to a subservient role of delivering up resources to global systems, with a strong degree of organizational convergence, then this assumption would be questionable. Though there are clearly many circumstances where eliminating or changing contexts is justified in efficiency terms, there is the problem of imperfectly substitutable products, on one hand, and the externality of eliminating certain developmental dynamics on the other. Mainstream economic geography and international economics has had virtually nothing to say about these issues.

6. Consequences for research

The argument of this paper has consisted of four main points. First, a proper conceptualization of the relationship between the regional and the global requires a
dynamical theory. This theory should be based on growth theory, but this in turn requires ending the prevailing confusion of local versus economy-wide sources of increasing returns. Second, the local or national differences in the ways production systems, labor markets, supply chains and relationships come together are sources of M-A externalities. Though there is a rich literature on sources of such differences, we argued that it pays insufficient attention to the sources of action at the level of the actor: the “dark matter” of regional difference, especially in regard to innovation pathways.

A third point of this paper is to argue that the answer can be found in developing the notion of “context.” Most economists reject out of hand the possibility of significant differences in expectations and choices from one place to another, while geographers and sociologists tend to assume the importance of such differences without giving them precise analytical definition. Situationalism, based on strong results from observation and experimental evidence, shows that all economic action is contextual. In some cases, the context is principally structured by formal, long-distance relationships, while at the other extreme, it reflects highly localized, informal and unplanned interactions (and then there is everything in-between). This leads us to the fourth point, which rejects the common opposition of “local contexts” to globally-abstract systems of production; instead, we have applied the analytical concept of context both to local economies and to new forms of geographically-extended and fragmented production, and ask how each actor-environment shapes economic learning and development.

Taken together, these arguments suggest a research agenda for economic geography, in the four main areas. The first concerns the welfare impacts of localization and globalization. In concrete terms, this would entail developing analytical ways of
shedding light on “paths taken and paths not taken” in development. A second area for research is to understand the micro-economics of situations – what they consist of and how they affect preferences, emulation, and resulting choices in ways that influence pathways taken for regions at all positions relative to technological frontiers. A third area is the detailed geographical economics of how the creation of new technologies and knowledge and their geography unfold over time and space and lead to economic growth as a whole, the “MA + R” externalities process. Finally, the pressing contemporary issue of how geographical and organizational changes in production systems are altering contexts, behaviors and choice/development pathways, should be a major element in globalization and regional development research. To shed light on the relative welfare properties of traditional contexts and the geographically-distributed ones that may be emerging, we would need to be able examine several of their dimensions. The technological dimension concerns the “bandwidth” of communication and interaction at different geographical scales, a subject already pregnant in the literature on face-to-face contact, but which needs a great deal of further development (cf. Storper and Venables, 2004; Bathelt, Malmberg and Maskell, 2004). The institutional dimension concerns who sets the rules of interaction in contexts and influences the type of information developed and used. The market dimension concerns the evolution of costs of supplying different types of information and interaction. Understanding these features would in turn allow us better to understand the evolution of the situations of actors and to consider the resulting cognitive processes in production contexts.

These are dauntingly complex issues that would require major theoretical and empirical investment, but pursuing them would undoubtedly place economic geography
at the center of efforts in the social sciences to understand the dynamics of twenty-first century global capitalism.

REFERENCES


Rules of thumb: [www.rulesofthumb.com](http://www.rulesofthumb.com)


